RESEARCH PAPER

Residual Densely Connected Progressive Attention Based Transnet for Crop Recommendation Using Soil Conditions

Pani Bhathe Viswanath[1], Dr. Santhosh Kumar G[2*]
[1] Research Scholar, Department of Computer Science and Engineering (Data Science), Dayananda Sagar University, Bangalore - 562112, India
[2*] Associate Professor, Department of Computer Science and Engineering (Data Science), Dayananda Sagar University, Bangalore - 562112, India (Corresponding Author)

ABSTRACT
Agriculture is the major contributor to a country's economy and the employment sector. Selecting the wrong crops that are not appropriate for given soil and environmental circumstances is one of the main challenges for farmers, causing low productivity and soil fertility problems over a long period. Building on the fusion of cutting-edge machine learning approaches and metaheuristic optimization algorithms, this study presents a fresh crop recommendation model meant to solve such major problems. The model starts with gathering input data from public datasets, including significant agricultural factors like nitrogen (N), phosphorus (P), potassium (K), pH value, temperature, humidity, and rain circumstances dramatically impacting crop suitability. In the pre-processing stage, z-score normalization is used to improve consistency and model accuracy; input data will be normalized, and unwanted data will be removed. The Dung Sparrow Optimization (DSO) algorithm, which simulates adaptive dung beetle and sparrow behaviors, is used for feature selection to maximize exploration and exploitation. The procedure improves the choice of the most insightful features for improving learning speed, performance, and convergence rate. Using Residual Densely Connected Progressive Attention Based Transnet (RDPAT), an advanced transformer based deep learning approach developed for precise crop prediction and soil nutrient analysis, the selected features are processed. The RDPAT system employs residual connections, dense layers, and progressive attention mechanisms to extract and rank informative patterns. Experimental results show the high performance of the model, with 99.26% accuracy, 96.25% recall, 98.93% precision, and 18.76s of execution time, as well as its efficacy and reliability for practical agricultural use.
Keywords: Crop Recommendation, Soil Conditions, z-score normalization, Sparrow Optimization, Transnet, Progressive attention.
How to cite this article: Viswanath PB, Kumar GS. Residual Densely Connected Progressive Attention Based Transnet for Crop Recommendation Using Soil Conditions. Int J Drug Deliv Technol. 2026;16(24s): 1032-1046. DOI: 10.25258/ijddt.16.24s.124
Source of support: Nil.
Conflict of interest: None

1. Introduction

Agriculture continues to be a pillar of India's increasing socioeconomic infrastructure, supporting livelihoods and fueling national progress. One of the greatest challenges that farmers have is choosing the best crops for their farms, as they tend to use non-scientific and traditional methods.This problem is significant in a nation where almost 58 percent of the population relies on agriculture for survival [1]. The cultivation of healthy and sufficient crops and food is vital for all, particularly as the world's population keeps growing. Crop production plays a crucial role in a nation's economy [2]. In periods of climate change, even farmers who depend on traditional knowledge find it hard to make sound decisions regarding crop health. This tends to result in crop failure, and consequently, it leads to a reduction in total crop output [3].

In agriculture, crop prediction is vital and highly dependent on environmental factors like rainfall, humidity, and temperature. In the past, farmers were able to select which crop to grow, monitor their cultivation, and determine the best time to harvest [4]. Early and precise estimation of crop yield is critical for quantitative and financial evaluation at the field level, assisting in determining agricultural commodity strategic plans, guiding import-export policies, and increasing farmers' incomes [5]. Understanding how water requirements and crop yield will alter in response to future local climate conditions is essential in developing more precise and effective adaptation planning [6]. There is a desperate call for a cropping system that guarantees food security through high yields and, at the same time, protects the environment [7].

A recently created technology with growing momentum in the agricultural sector is the crop suggestion system. It gives farmers insightful advice about crop growth techniques, fertilizer use, and choosing the most lucrative crops to cultivate depending on the current demand on the market [8]. A fresh, simple, and inexpensive approach to effectively monitoring the dynamic physical properties of crops in the field [9] is offered by unmanned aerial vehicles (UAVs). For accurate agriculture and environmental monitoring, several applications in the form of field and imaging spectroscopy have been created [10]. Machine vision forms a key part of machine learning (ML), which is heavily used in applications where complex control and automation are needed. The emphasis has now shifted from conventional image processing and machine learning to modern artificial intelligence (AI) and deep learning (DL) techniques [11].

Big data technologies such as ML and DL are the fundamental elements of information and communication technology (ICT). Their sophisticated data analysis ability enables farmers to obtain valuable insights, enhance their knowledge of farming practices, and make improved decisions [12]. Utilizing farm Big Data has the potential to generate useful knowledge. The implementation and design of a big farm data warehouse are, however, quite challenging [13]. Agro advisory systems playa vital role in increasing agricultural productivity and assuring food security, particularly in developing nations where agriculture is the main part of economic development [14]. Through the incorporation of ML, farmers are able to leverage data to address decision-making issues, water management, soil management, crop management, and livestock management [15].

1.1 Motivation and problem statement

Agriculture is a pillar of most developing nations and the basis of their economic development. Maximizing productivity necessitates that one chooses the best crop to suit a particular region. A prevalent challenge that inhibits farmers and agricultural players from embracing scientifically validated crop recommendation methods is their constant refusal or inability to practice these best practices. However, the current models faced several limitations, such as complexity in the process, which is among the most significant challenges, with great computational power required to handle massive and diverse datasets in some instances. The systems also overfit, where the model fits perfectly with training data but does not generalize correctly to new unseen inputs. Another crucial weakness is the overdependence on data availability and quality, which can restrict the validity of the models in regions with incomplete agricultural data. In addition, the vast majority of crop recommendation models are not good at coping with limited generalization. Therefore, their performance degrades when they are used in different environmental or geographic conditions. Lastly, scalability is an issue as changing the system to enable extensive use for different farm environments proves to be technically demanding as well as resource-intensive. This limitation in this research paper presentsa novelDL model,a combination of a residual dense layer and an attention mechanism. This will reduce overfitting problems and generalizability issues. To address these limitations, the current research paper introduces a new deep learning model that integrates residual dense layers with an attention mechanism. The main contributions are given below.

- To implement a z-score normalization to normalize the data and remove unwanted information from the input data.
- To construct a Dung Sparrow optimization algorithm (DSO) to select the optimal features.
- To present a Residual Densely connected progressive attention based Transnet (RDPAT) to enhance the recommendation system's performance.

The paper is formulated from Section 1,which consists of an introduction to the crop recommendation system, and Section 2, which consists of related works on autonomous vehicles. Section 3 consists of the proposed methodology. Section 4 consists of results and discussion.Finally,the conclusion and future work are given in Section 5.

2. Related Work

This section reviews some recent works related to the proposed approach of the Crop Recommendation system.

Wankhade et al. [16] developed the Efficient DL with Attention Mechanism Model (EDLAM), an Advanced Synthetic Minority Over-sampling Support Vector Machine (PSMOTE_SVM) model to remove noise in the data. A Boosted Sequential Forward Feature Selection (BSFFS) method was then used to select the most important features. The model then made used Bidirectional Long Short Term Memory

(Bi-LSTM) with a Global Attention framework for crop growth prediction. The outcome of the model achieved an accuracy level of 95.1% and a precision level of 95.74%. The model possesses complexity with high computational requirements.

Sudhamathi et al. [17] constructed Extra Tree Regressor (ETR). Initially, the dataset was downloaded from GitHub, and the StandardScaler technique was utilized for preprocessing and data normalization. The data was then separated into preprocessed sets, with 70% utilized for training and 30% for testing. Kernel Principal Component Analysis (KPCA) was utilized to determine useful characteristics. The Least Absolute Shrinkage and Selection Operator (LASSO) Regression was used for feature selection. The model achieved an accuracy rate of 95%. The drawback was overfitting issues.

Subramaniam et al. [18] designed a weight-tuned deep convolutional neural network (WTDCNN) for crop prediction. First, agricultural data were gathered fromthe southern Indian region. Data cleaning and normalization were performed in the preprocessing stage to prepare the dataset for analysis. Next, dimensionality reduction was done using Squared Exponential Kernel-based Principal Component Analysis (SEKPCA) to minimize data complexity while keeping important characteristics. The WTDCNN model attained a higher accuracy of 98.96%. The model faced issues with data dependency.

Kiruthika et al. [19] suggested Improved Distribution-based Chicken Swarm Optimization with Weight-based Long Short-Term Memory (DCSO-WLSTM) for crop prediction and recommendation. The technique was designed to improve the precision and effectiveness of crop selection according to soil conditions, soil types, and crop yield information to solve the common problems encountered by Indian farmers. The IDCSO-WLSTM method had an accuracy rate of 92.68%. The model has drawbacks in thatit does not generalize to other regions or crops if it has been trained on a given dataset.

Ayesha Barvin et al. [20] implemented Graph-based Crop Recommendation Algorithms. The research centered on determining the best crop for a season through the study of essential factors like nitrogen, potassium, and phosphorus c N, P, K, pH, temperature, humidity, and soil pH. The researchers employed supervised learning to organize the input data as nodes on a graph with edges indicating possible relationships among the features. The experimental outcome showed that the GCN model had a greater accuracy rate of 98%. A disadvantage was scalability issues,which were potential difficulties in handling large-scale agricultural data. Table 1 presentsan overview of existing models andtheir performance.

Table 1: Overview of existing models andtheir performance.

Author Name and Reference	Method	Performance	Disadvantages
Wankhade et al. [16]	EDLAM	Accuracy level of 95.1%, precision level of 95.74%	It possesses complexity in high computational requirements.
Sudhamathi et al. [17]	ER-ETR	Accuracy rate of 95%	Overfitting issues
Subramaniam et al. [18]	WTDCNN	Accuracy of 98.96%	Data dependency
Kiruthika et al. [19]	DCSO-WLSTM	Accuracy rate of 92.68%	Limited generalization
Ayesha Barvin et al. [20]	Graph-based Crop Recommendation Algorithms	Accuracy of 98%	Scalability Issues

The current crop recommendation system has a number of complexitiesin the process, which is one of the key challenges, with extensive computational power needed to process and varied datasets. The systems also tend to overfit, whereby the model predicts well with the training data. Another essential weakness is the heavy reliance on data quality and availability, which can limit the model's validity in areas with poor agricultural data. Additionally, most crop recommendation models do not handle limited generalization well, so their effectiveness diminishes when applied to other environmental or geographical conditions. Finally,scalability remains a challenge since modifying the system for widespread application across various agricultural environments can be technically challenging and demanding in terms of resources. To overcome these limitations and improve system performance, this research paper proposes the RDPAT model.

3. Proposed Methodology

Weakening soil fertility has decreased crop yields, and it has become challenging for farmers to decide onthe most appropriate crops for their soil conditions. Correct prediction of crops can improve crop yields significantly. ML provides tremendous help in making a crop forecast. When forecasting crops, we need to consider soil geographic and environmental features. Therefore, feature selection is the key aspect of identifying suitable crops. The decision will use a publicly available dataset as its initial data collection inputs. There are different attributes to consider: N, P, K, pH, Temperature, Humidity, Rainfall. These attributes actually decide which crop to recommend. The overall workflow diagram for the proposed model is illustrated in Figure 1.

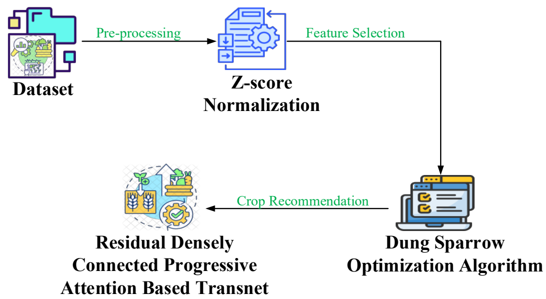

Figure 1: The overall workflow diagram for the proposed model

The diagram shows the overall process of the suggested model, starting from gathering farm data from open sources. Collected data undergoes a pre-processing stage; here, input data will be normalized using z-score normalization to make data for further analysis. Optimal features from the dataset are collected using the Dung Sparrow optimization algorithm (DSO), which selects the best feature set using the fitness function. This algorithm combines the strength of both dung beetle optimizer and sparrow optimizer to improve convergence. Selected features are trained on "Residual Densely connected progressive attention based Transnet (RDPAT)" for soil nutrition analysis, and based on this crop, it will be recommended. This transformer model is made with dense layers to provide optimal results using the selective features of the metaheuristic algorithm.

3.1 Z-Score Normalization

One of the most important pre-processing techniques, data normalization, significantly raises the general performance of the proposed models. The qualities of features in the input data set often have vastly different magnitudes and distributions. High-magnitude characteristics can disproportionately impact the training process, hence lowering the learning accuracy and convergence rate of the model [22]. Solving this calls for the normalization of every characteristic in the database. Chosen as the chosen standardizing approach in the current study is Z-score normalization. This procedure guarantees a uniform scale for all characteristics and standardizes data to have a standard normal distribution. Particularly, Z-score normalization standardizes data by normalizing data using the formula of dividing data by the standard deviation and subtracting the mean, as shown in equation (1). This transformation generates traits with a mean of zero and a standard deviation (SD) of one that can be used to normalize their influence in the model's training.

$$K = \frac{(I - \mu)}{\sigma} \tag{1}$$

In this equation, K is the normalized value, I denotes the original data point, μ denotes the mean of the data, and σ is represents the SD.

3.2 Dung Sparrow Optimization Algorithm

From the above discussion, we can see that dung beetles make use of celestial information in order to roll and move their dung balls, as they want to move in a straight line. To replicate rolling action, you must move in a continuous, straight manner across the whole search region. In this work, it is assumed that not only celestial information influences the movement of the dung beetle, but also the light source's intensity determines the path of the dung beetle. While the dung beetle is rolling the ball, it can either reposition itself or make a new adjustment, and it can be expressed as follows

$$i_x(u+1) = i_x(u) + \alpha \times z \times i_x(u-1) + c \times \Delta i$$
$$\Delta i = \left| i_x(u) - I^a \right| \tag{2}$$

In that case u is the current iteration time and $i_x(u)$ is the position of the xth dung beetle at the uth iteration. c is a constant value within the interval (0, 1). The parameter α is a characteristic coefficient whose value can range between -1 and 1. Also I^a denotes the worst position of global and Δi used to represent illumination intensity. If a dung beetle comes upon an obstruction impeding its onward movement, it will need to reorient utilizing a dance in order to locate a fresh direction. For ball-rolling dung beetles, the authors sought to replicate this dance behaviourin order to compute a fresh, forward rolling direction using a tangent function [23]. Once the dung beetle has effectively set a new orientation, it will roll the ball opposite the distance of the rolling. Thus, the position of the ball-rolling dung beetles is reviewed and formulated.

$$i_x(u+1) = i_x(u) + \tan(\theta)\left|i_x(u) - i_x(u) - i_x(u-1)\right| \tag{3}$$

here θ representsthe angle of deflection. Sparrows foraging can be divided into followers and explorers. The followers rely on the explorer to have food access; explorers are more likely to have more energy reserves, find food resources, interact, and lead the whole population of sparrows to their foraging site and directional choices, so saving energy is their main priority. In this context, the individual sparrow's fitness value reflects the measure of energy reserves. The formula for the explorer position update at each iteration is described as follows.

$$Y_{j,k}^{v+1} = \begin{cases} Y_{j,k}^{v}.\exp\left(-\dfrac{j}{\alpha.iter_{\max}}\right), & if\, WV_2 < SV \\ Y_{j,k}^{v} + R.M, & if\, WV_2 \geq SV \end{cases} \tag{4}$$

In the equation v is the iteration number at the moment, and $iter_{\max}$ represents a constant that signifies the maximum number of iterations. $Y_{j,k}$ is the location of the jth sparrow in the kth dimension. The parameter $\alpha(\alpha \in (0,1))$ is random variables. $WV_2(WV_2 \in [0,1])$, $SV(SV \in [0.5,1])$ represents the warning and safety value. R represents the random variable with a distribution of normal. $M \times e$ Matrix, where every entry in the matrix is 1. When $WV_2 < SV$ it means that no predators are available in the search nature at that time. On the other hand, when $WV_2 < K'SV$ it means that some of the sparrows in the population have sensed a predator and have warned of the break. All the sparrows have to find safer locations right away in search of sustenance in this case. Some followers could always keep an eye on the explorers for food supplies to increase their rate of predation. Their energy state directly influences the foraging rank of the followers in the population. Following is the process for changing a follower's position.

$$Y_{j,k}^{v+1} = \begin{cases} R.\exp\left(\dfrac{Y_{worst} - Y_{j,k}^{v}}{j^2}\right), & if\ j > o/2 \\ Y_Q^{v+1} + \left|Y_{j,k}^{v} - Y_Q^{v+1}\right|.B^{+}.M, & if\ otherwise \end{cases} \tag{5}$$

The equation (5), Y_Q represents the best place so far employed by the explorers and Y_{worst} represents the position of worst. B is a $1 \times e$ matrix where each entry is randomly allocated the value of 1 or -1. When $i > o/2$ it means that the jth follower, with a lower fitness value, is starved and feeling severe hunger, it needs to find food in another destination to restore its energy. The position of explorers and followers is also dynamic and can shift over time. A better food source can turn any sparrow into an explorer, but the total number of explorers to followers in the population stays the same. Put another way, if one sparrow switches to being an explorer, another must become a follower. While searching for food, followers always go looking for the explorers to gain access to the top food sources and hence can obtain food directly from these optimal areas or search around the explorers.

Assume that 10% to 20% of the population of sparrows is randomly assigned as danger-aware. When these danger-aware sparrows detect danger, the population shows antipredation behaviors. Sparrows near the boundaries of the population will quickly move towards safer positions to take a more advantageous position. At the same time, those in the interior of the population will move randomly to approach their neighboring sparrows. The mathematical formula for this phenomenon is the following.

$$Y_{j,k}^{v+1} = \begin{cases} Y_{best}^{v} + \beta.\left|Y_{j,k}^{v} - Y_{best}^{v}\right|, & if\ g_j > g_h \\ Y_{j,k}^{v} + N.\left(\dfrac{\left|Y_{j,k}^{v} - Y_{worst}^{v}\right|}{(g_j - g_a) + \varepsilon}\right), & if\ g_j = g_h \end{cases} \tag{6}$$

In the equation (6), Y_{best} represents the global optimal of the current place. Parameter β denotes a step control parameter and is an instance of a random number that obeys the distribution of normal data with a mean of 0 and a variance of 1. $N(N \in [-1,1])$ is likewise a random variable but g_j is an indicator of the fitness value of the individual sparrow. g_h and g_a are the current global best and worst values, respectively. The small constant ε is used to avoid division by zero in the denominator. If $g_j > g_h$ then the sparrow is located at the boundary of the population and is most susceptible to predators. The parameter N controls the direction in which the sparrow will move and controls the foot length.The algorithm for Dung Sparrow optimization is shown in Table 2.

Table 2: Algorithm for Dung Sparrow optimization.

Initialized population, parameters
While $(u \le U_{\max})$ do
  For $x \leftarrow 1\, to\, N$ do
    If $x ==$ ball rolling dung then
$\delta = rand(1)$
      If $\delta < 0.9$ then
Renovate ball rolling dung battle position by equation (2).
      Else
Renovate ball rolling dung battle position by equation (3).
      End if
    End if
  End for
  For iteration $v = 1$ do
    For $j = 1\, to\, O$ do
    If $WV_2 < SV$ then
Renovate position of explorer equation (4).
    Else
Renovateposition of explorer equation (5).
    End if
    If $j > o/2$ then
      Update position of follower using
$R.\exp\left(\frac{Y_{worst} - Y_{j,k}^{v}}{j^2}\right)$
      Else
      Update position of follower using
$Y_Q^{v+1} + \left|Y_{j,k}^{v} - Y_Q^{v+1}\right|.B^{+}.M$
      End if
      If $(g_j > g_h)$ then
      Update position for antipredation behavior
$Y_{best}^{v} + \beta.\left|Y_{j,k}^{v} - Y_{best}^{v}\right|$
      Else
      Update position for antipredation behavior
$Y_{j,k}^{v} + N.\left(\frac{\left|Y_{j,k}^{v} - Y_{worst}^{v}\right|}{(g_j - g_a) + \varepsilon}\right)$
    End if
      Update global best and worst values.
    End For
  End for
End while
Return best solution

The DSOA presents multiple benefits in terms of powerful optimization, high convergence speeds, and flexibility across different problem fields. The swarm intelligence mechanism improves exploration and exploitation, which is powerful for sophisticated optimization problems. Moreover, DSOA has been shown to be resilient to local optima, thus enabling the efficient escape of suboptimal solutions.The algorithm's capacity to balance exploration and exploitation leads to higher-quality solutions in a variety of settings since the search for new solutions is balanced by iterating on existing ones.

3.3 Residual Densely connected progressive attention based Transnet

This work proposes the Progressive Attention Network (PAN), an extension of the RDPAT model with the addition of an Efficient Multi-Scale Attention Module (EMAM). The EMAM is formulated to give high importance to the significant feature representations in data, allowing for coarse-grained feature extraction. Moreover, a Spatial and Channel Attention Module (SCAM) is proposed to further improve the feature extraction process for coarse-to-fine feature representation. For the input data, $I \in \mathbf{R}^{D \times L \times A}$ it is first processed by an encoder using the Xception architecture, which collects general features. The result of the process, having passed through the reconstruction network, is referred to as $\hat{I}$. Introducing noise in this process increases the coding region of the image, in effect covering up the spoiled blank coding points. The formula for common feature extraction is the following.

$$\hat{I} = G_{xcep}(\widetilde{I}) \tag{7}$$

here $\widetilde{I}$ represents the result of adding white noise during the training period. Efficient Multi-Scale Attention Module (EMAM) consists of three essential elements: feature grouping, parallel sub networks, and cross-spatial learning. Feature grouping divides the input feature map into sub-features for more targeted processing, whereas parallel sub networks derive both local and global information from these group feature maps [24]. The architecture for Residual Densely connected progressive attention based Transnet is illustrated in Figure 2.

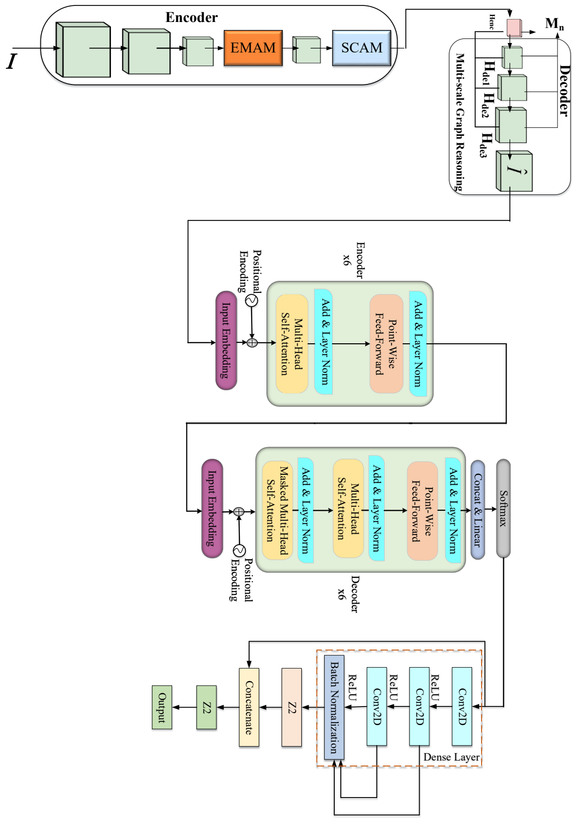

Figure 2: The architecture for Residual Densely connected progressive attention based Transnet

Cross-spatial learning summarizes attention-weight descriptors from the parallel sub networks, recognizing pairwise relations between pixels in the input image. In order to obtain various semantic information from the input image feature map $I \in \mathbf{R}^{D \times L \times A}$, EMAM divides the feature maps into H groups across the channel. The expression for the divided feature map is given below.

$$I = [I_0, I_x, ..., I_{H-1}], I_i \in \mathbf{R}^{D//H \times L \times A} \tag{8}$$

here $D // H$ represents D divided by H. Multi-scale spatial information is obtained by the large receptive local fields of the neurons. The EMAM achieves this with three parallel pathways to extract attention-weighted descriptors. Two pathways make use of 1×1 convolution to attend to smaller receptive fields, and the third pathway uses 3×3 convolution to attend to a larger receptive field. Cross-channel interactions are treated in the channel direction to model dependencies between all channels effectively while managing computing resources. Channel information is encoded using two 1D global average pooling processes in two spatial directions within the branch employing a 1×1 convolution. Express this as.

$$I_{avg} = avgpool_i(i) \in \mathbf{R}^{D//G \times 1 \times A} \tag{9}$$

$$J_{avg} = avgpool_y(I) \in \mathbf{R}^{D//G \times 1 \times A} \tag{10}$$

Moreover, a single kernel 3×3 is used in the third branch to learn multi-scale feature representations. Based on the batch dimension, the EMAM reorders and transposes G groups, so the input tensor is transformed into the form. The two encoded features on the image height dimension are stacked by the EMAM using a method comparable to Coordinate Attention (CA). Without downsampling, the two branches share the same 1×1 convolution. To model the 2D Binomial distribution of the 1×1 convolution outputs, the 1×1 convolution results are factored into two vectors, and two non-linear sigmoid functions are employed for fitting. The channel-wise attention graphs in each group are multiplied together to create independent cross-channel interactive elements between the two parallel branches. Simultaneously, the third branch gets local cross channel interactions via a 3×3 convolution to expand the feature space. Thus, the EMAM efficiently uses inter-channel information to control the relative significance of different channels without losing precise spatial structure data inside each channel. The EMAM helps to collect characteristics thoroughly by gathering data throughout many spatial dimensions. For the efficient capturing and retention of knowledge, the design employs distinct tensors derived from the output of the 1×1 and 3×3 convolutional branches. The output of the 1×1 branch includes 2D global average pooling representation of worldwide spatial information. The output from the little branch is then reshaped to fit the corresponding dimensions before the joint activation mechanism for channel features. The 2D global pooling algorithm formula is given below.

$$j_d = \frac{1}{L \times A} \sum_{y}^{L} \sum_{x}^{W} i_d(x, y) \tag{11}$$

here i_d denotes the input features at the dth channel, j_d represents the output associated with the dth channel. Spatial attention (SA) and channel attention (CA) modules improve feature extraction by concentrating on useful spatial locations and channel wise features. The structure generally consists of two broad aspects: spatial attention, which accentuates salient areas in the feature map, and CA, which weighs salient feature channels. The computation process includes computing convolutional operations, pooling, and activation functions to process the feature representation, eventually enhancing the capability of the model to perceive fine-grained details.

$$H_d = CA(G) \otimes H$$
$$H_d^{'} = SA(H_d) \otimes H_d \quad (12)$$

Let the input image be denoted as H, and $\otimes$ symbolize multiplication of element-wise. Channel attention mechanism generates weighted output H_d, and the lastimproved feature map $H_d^{'}$ is captured. The particular equations for CA and spatial attention SA mechanisms are outlined below.

$$CA = \sigma(h_1(h_2(h)) + h_1(h_3(H)) = \sigma(A_1(A_0(H_{avg})) + A_1(A_0(H_{\max}))) \quad (13)$$

$$SA(H) = \sigma(h(h_d(H_{avj}, H_{\max}))) \quad (14)$$

In the current scenario h_1, it represents the Multi-Layer Perceptron (MLP), and $h_2$and h_3 stands for the average pooling and max pooling operations, respectively. A_0And $A_1$the weights corresponding to the two linear layers. h_dStands for feature concatenation, and f stands for convolution operation. $\sigma(.)$Stands for the sigmoid function. Furthermore, H_{avg} and $H_{\max}$ refers to representations based on average pooling and max pooling, respectively. The RDPAT model is an advanced crop recommendation framework that combines different elements to make the best agricultural decisions. The model has three main components:input embedding, encoder-decoder structure, andsoftmax output layer.

Network traffic naturally contains sequential feature data, namely IP address quadrate source and destination port numbers. To make use of the spatial data of these input characteristics, our model has additional positional encoding at the base of the encoder stack. The two may be easily added together by making the positional embeddings the same dimensions as the input embeddings. This additive builds the ultimate input into the self-attention layer. For carrying out positional encoding (POE), sine and cosine functions with varying frequencies should be used. Odd places within the sequence are most specifically represented by a cosine function, as seen in equations (15, 16), while even locations are represented by a sine function.

$$POE(pos, 2m) = \sin(pos / 1000^{2m/e_{model}}) \quad (15)$$

$$POE(pos, 2m+1) = \cos(pos / 1000^{2m/e_{model}}) \quad (16)$$

Each of the six encoders in the encoder stack has a multi-head self-attention network as well as a point-wise feed-forward network (FFN). The size of the FFN layers is a hyperparameter that can be tuned throughout training. To get the best results, the FFN layers hold 1024 neurons, and the padding size for embeddings is 32. Furthermore, the output of each sub-layer is computed using residual connections and layer normalization, which are subsequently transferred to the next encoder in the stack. This architecture improves the general efficacy of the model by helping the information flow. Introduce a multi-headed masked self-attention sublayer to strengthen every decoder; it enhances the stability of the intrusion detection model by randomly masking some features and estimating them from unmasked ones.

To maintain the hierarchical architecture, add six decoders to the decoder stack. Use a softmax layer to categorize after building the encoder and decoder stacks. With layer normalization in between to steady and improve training, the self-attention layer has a residual link to the point-wise feed-forward network (FFN). These approaches help to solve issues like shifting covariates and vanishing gradients. Both decoder and encoder also feature a point-wise feed-forward network made of two linear transforms activated by the ReLU function. Moreover, the same weights are applied to all rows of the attention matrices, essentially making the embeddings more information.

$$FFN(x) = \sigma(\max(0, pV_1 + c_1)V_1 + c_2) \quad (17)$$

The masked scaled dot-product attention of the head l in the decoder. It attends over a sequence of queries T, keys K, and values U. The masked scaled dot-product attention computation is given by equation (19), where the scaling factor is e_k and the masking matrix is W.

$$Att(T,K,U) = soft\max\left(W + \frac{TK^N}{\sqrt{e_k}}\right)U \quad (18)$$

In equation (19), division $\sqrt{e_k}$ stabilizes gradients during training. Matrix $W \in \mathbf{R}^{k\times k}$ avoids attention to future positions so that predictions for position m only depend on previously known outputs, reducing overfitting. Masking is used by adding the softmax function -∞, normalizing every row into a probability distribution. A new representation of input is then constructed through the dot product of this matrix and U. improves performance by enabling each attention head to have its own T, K and U weight matrices. This process is identical to single-head attention, with parameter dimensions being $V_m^K \in R^{e_{\text{model}} \times e_k}, V_m^U \in R^{e_{\text{model}} \times e_U}, V_m^Q \in R^{e_{\text{model}} \times le_U}$. The outputs from each of the heads are concatenated, allowing the model to attend various positions and giving a multi-head attention (MHA) layer.

$$MHA(T,K,U) = Concat(head_1, \ldots, head_l)V^Q \quad (19)$$

$$head_m = Attention(TV_m^T, KV_m^K, UV_m^U) \quad (20)$$

For faster convergence and stable training, layer normalization is used for every sample $m \in R^e$ as outlined in equation (22).

$$LayerNorm(m) = \frac{m-\mu}{\delta}.\alpha + \beta \quad (21)$$

here $\mu \in R, \delta \in R$ denotes mean and standard deviation of the input features. The operation $(.)$ represents the dot-wise element operator, whereas $\alpha \in R^e, \beta \in R^e$ are the learnable parameters for the affine transformation. The Residual Dense Network (RDN) was first constructed for denoising and super-resolution of data. The RDB is employed to extract strong local characteristics from densely linked convolutional layers. Its design allows for efficient feature reuse and enhances gradient flow when training. It also implements a continuous memory (CM) method by allowing connections from the output of the previous RDB to reach every layer of the current RDB. In this work, we advocate using RRDN to detect crop leaf sickness. The first model, designed for picture super-resolution, lacks dimensionality reduction, making it ineffective for processing large datasets in image classification. To fix this, first convolve the Res-dense block (RDB) input picture and then perform batch normalization. In this case, $Z2$ is activated by LeakyReLU and $Z1$ is the residual concatenated tensor obtained by concatenating $Z2$ the input layers.

$$Z = NM(CV(IL)) \quad (22)$$

The above equation NM represents the normalization operation, CV represents convolution operation, and IL represents the input layer. Z Is the tensor that has been normalized in RDB. In the initial RDB block, $Z1$ is the resultant tensor is passed to subsequent RDB blocks to be used throughout the model lifetime, which is desirable for image super-resolution. However, in classification problems, there is a heavy demand for weight, both reducing recommendation efficiency and accuracy. In order to rectify this situation, remove the tensor $Z2$, which has no residual concatenation within the RDB, and redirect the focus toward employing it in residual concatenation.

$$Z1 = concat(Z, IL) \quad (23)$$

$$Z2 = O(Z) \quad (24)$$

In equation (23), Concat indicates the residual concatenation operation. $Z1$ is the tensor of the concatenation of Z and IL, with O representing the LeakyReLU operation. $Z2$ denotes the tensor after the LeakyReLU operation with alpha = 0.3. Here, the input tensor size of 196*196*64 is next fed through a 3-layer RDB to provide feature extraction. At 98*98*64, the size following this surgery is suitable for leftover addition. The output from the 3-layer RDB and the original input tensor are then combined via a residual connection action to produce a fresh tensor $Z3$.

$$Z3 = TR^3(CV(CV(IL))) \quad (25)$$

In equation (26), the 3-layer RDB operation is represented by an operator TR^3. In order to increase classification performance, the input layer can be reloaded for residual connection after three times pooling operation, having an output image size of 1*1*128. Subsequently, there is a residual connection operation involving the tensor $Z3$, resulting in the output tensor $Z4$ is ready for classification, as shown in equation (27).

$$Z4 = concat(Z3, PQ^3(CV(IL))) \quad (26)$$

In this case, $PQ^3(.)$represents the pooling operations three times. A dense layer is used to perform classification. An $Z2$regularizer is used in the dense layer to prevent overfitting. The optimizer used is the Adadelta function, while the loss function used is cross entropy.

4. Results and Discussion

The research has been implemented using the Python programming language, configured on a system with an Intel(R) Core(TM) i5-4670S CPU running at 3.10 GHz, along with 16.00 GB of RAM and a 64-bit operating system-based architecture.

4.1 Dataset Description

The data used in this experimental study comes from a public database called the Kaggle database. Seven factors make up the data: pH, P, Temperature, Humidity, N, K, and Phosphorus. Various crops have different combinations of these factors connected with them. The lookup table presents the suggested crops together with the numerical values for seven parameters. As indicated in the lookup table, the Kaggle dataset has 2,200 samples for 22 crops. With supervised learning methods combined with chosen functions and rule-based classifiers [21], it is arranged in a comma-delimited manner, which is employed for training the best model.

4.2Performance Evaluation

The performance of the RDPAT model will be examined using metrics such as F1-score, precision, and recall. Different state-of-the-art models will be compared with the proposed approach in order to prove the efficiency of the RDPAT technique. Initially, accuracy is evaluated for the dataset crop recommendation, asillustrated in Figure 3.

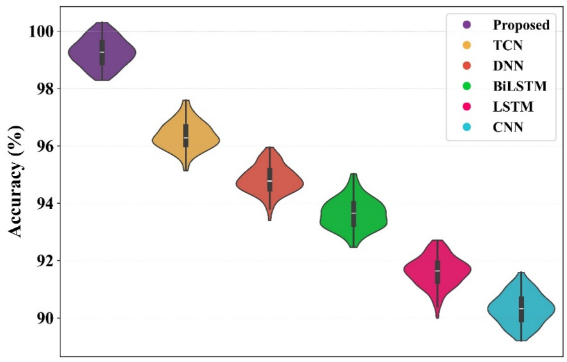

Figure 3: Accuracy is evaluated for the dataset crop recommendation

The RDPAT model presents an appreciably higher accuracy rate when compared with various other models, with an accuracy rate of 99.26%, while TCN [25], DNN [26], BiLSTM [27], LSTM [28], and CNN [29] had accuracy rates of 96.33%, 94.83%, 93.63%, 91.57%, and 90.34%, respectively. Such high performance highlights the strength of the RDPAT model in predicting appropriate crops based on soil properties, rendering it a useful tool for farmers looking to maximize their crop yields. Analysis of Precision for the RDPAT model and other models in the crop recommendation dataset is given in Figure 4.

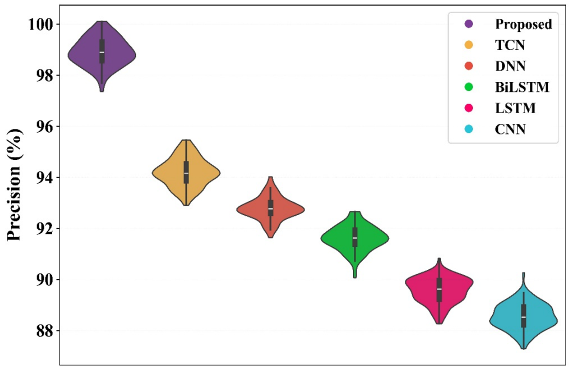

Figure 4: Analysis of Precision for the RDPAT model and other models in the crop recommendation dataset

The RDPAT model showed higher accuracy than other models tested in this study. The compared models, such as TCN, DNN, BiLSTM, LSTM, and CNN, recorded precision levels of 94.15%, 92.28%, 91.52%, 89.61%, and 88.31%, respectively. By comparison, the proposed model recorded a much higher precision value of 98.93%, highlighting its performance in the effective prediction of appropriate crops depending on soil conditions. This high performance indicates the capability of advanced ML methods in agricultural applications. Analysis of the F1 score for the RDPAT and the existing model in the crop recommendation dataset is illustrated in Figure 5.

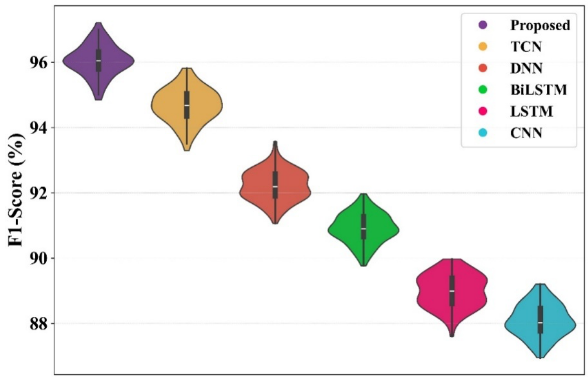

Figure 5: Analysis of the F1 score for the RDPAT model and the existing model in the crop recommendation dataset

RDPAT performs better than a number of other models as far as F1 score is concerned, registering better results than TCN, DNN, BiLSTM, LSTM, and CNN, which had F1 scores of 94.73%, 92.28%, 90.82%, 89.02%, and 88.10%, respectively. The new RDPAT model, however, has an even better F1 score of 96.05%. This improvement is substantial and reflects the model's capability to balance recall and precision equally well, rendering it a stable option for crop prediction tasks. Compassion of Recall for the RDPAT model and other models in the crop recommendation dataset is given in Figure 6.

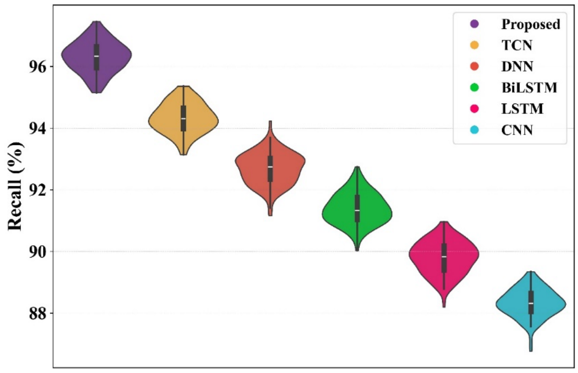

Figure 6: The compassion of Recall for the RDPAT model and other models in the crop recommendation dataset

The RDPAT model scored a higher recall than the other models. The existing models, namely TCN, DNN, BiLSTM, LSTM, and CNN, had a recall rate of 94.73%, 92.69%, 91.28%, 89.83%, and 88.31%, respectively. Comparatively, the proposed RDPAT model had a remarkable recall of 96.25%. This increased recall is a clear reflection of the efficiency of the model in accurately labelling true negatives, making the model even more useful in precision agriculture. An analysis of the specificity ofthe RDPAT model and other models in the crop recommendation dataset is given in Figure 7.

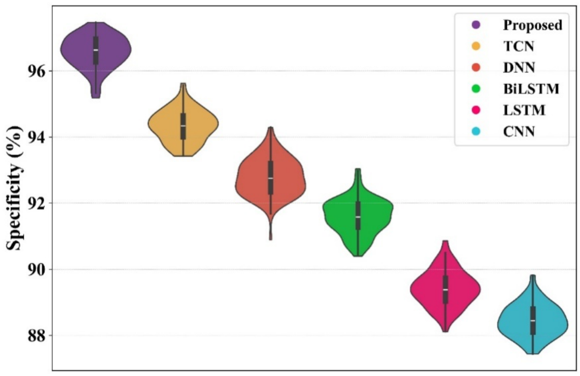

Figure 7: Analysis of Specificity for RDPAT model and other models in crop recommendation dataset

The RDPAT model showed better accuracy than other models. The other models, such as TCN, DNN, BiLSTM, LSTM, and CNN, hadspecificity values of 94.39%, 92.73%, 91.59%, 89.38%, and 88.48%, respectively. The proposed RDPAT model had a remarkable specificity of 96.58%. This remarkable improvement in specificity will highlight the model's ability to reduce false positives efficiently, and itis also a useful tool for agriculture. Compassion of Execution time for the RDPAT model and other models in the crop recommendation dataset is given in Figure 8.

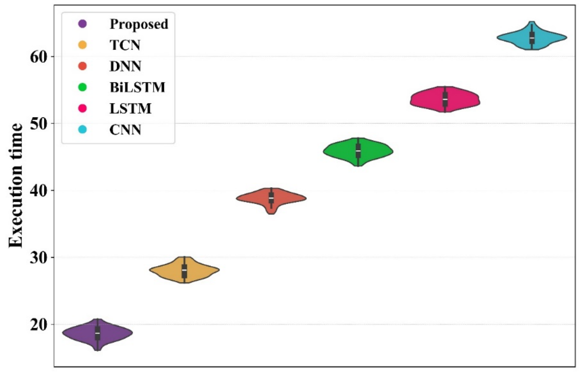

Figure 8: Compassion of Execution time for RDPAT model and other models in the crop recommendation dataset

The analysis focuses on the execution time of the RDPAT as compared to four other models, namely TCN, DNN, BiLSTM, LSTM, and CNN, whose execution times were 27.88s, 38.87s, 45.89s, 53.71s, and 62.77s, respectively. Surprisingly, the proposed model posted the lowest execution time of 18.76s. This efficiency not only improves the model's usability for real-time applications but also indicates its scalability potential for agricultural decision-making processes. Analysis of the ROC curve for the RDPAT model and other models in the crop recommendation dataset is given in Figure 9.

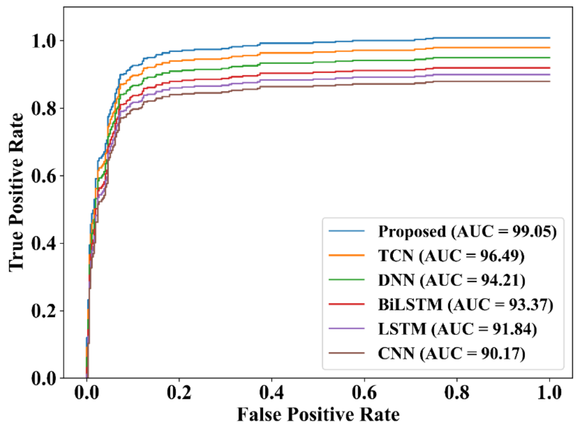

Figure 9: Analysis of the ROC curve for the RDPAT model and other models in the crop recommendation dataset

The RDPAT model performs better compared to other models, such as TCN, DNN, BiLSTM, LSTM, and CNN, which obtained AUCs of 96.49%, 94.21%, 93.37%, 91.84%, and 90.17%, respectively. The RDPAT model has a much greater AUC of 99.05%.

This high performance indicates the efficacy of the RDPAT model in correctly predicting crop suitability and is an important tool for maximizing agricultural productivity. The Training and testing loss and the training and testing accuracy of the RDPATmodel are represented in Figure 10.

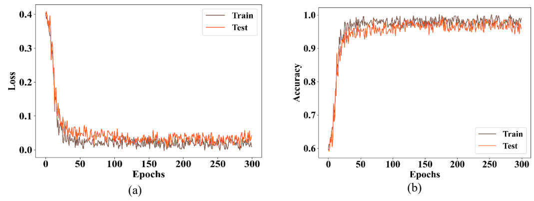

Figure 10: The Training and testing loss, Training and testing accuracy of the RDPAT model

Figure 10(a) depicts the RDPAT model's training and test losses throughout several epochs. Our suggested model has a remarkably low training and test loss, showing that it is highly efficient in decreasing errors during training and testing. The consistent pattern indicates that the model is optimized and can effectively generalize to new inputs. Figure 10(b) depicts the accuracy of training and testing the RDPAT model over various epochs. The model shows tremendous training and testing accuracy, mirroring its capability to perform very well in data learning. The high accuracy level further confirms the reliability and success of the model in providing precise predictions in farm applications. The confusion matrix for the proposed model in the crop recommendation dataset is illustrated in Figure 11.

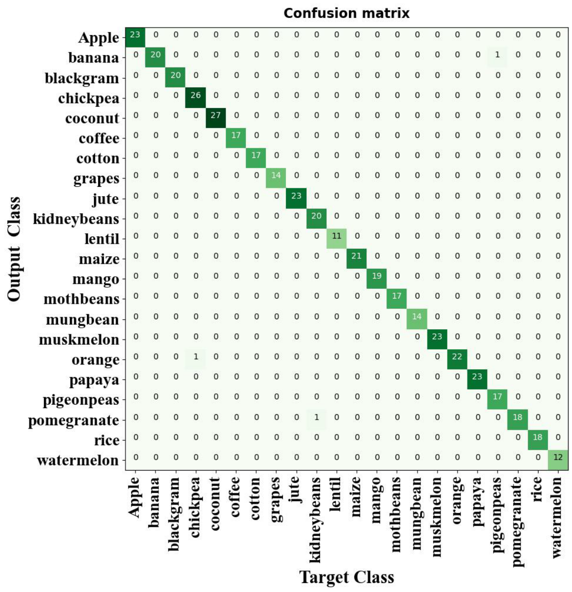

Figure 11: Analysis ofconfusion matrix for RDPAT model in crop recommendation dataset

The confusion matrix displays the classification results for different classes of fruits and crops with the target classes and their output counts. The rows are target classes, and the columns are the correct and incorrect numbers of classifications. For instance, when the target class is Apple, there are 23 correct and 0 for every other class, reflecting good performance in the recommendation. The same trend is seen for Banana (20 correct), Blackgram (20 correct), and so on, with high accuracy in most categories. Some classes, like Pomegranate and Rice, have low values in the respective output classes, indicating scope for improvement. Overall, the matrix reflects good performance in separating the different classes with a few exceptions.

4.3 Discussion

The existing models face several limitations, such as the Wankhade et al. [16] model, which is complex and has high computational requirements. The proposed model utilizes z-score normalization to preprocess and normalize the input data. This is because the normalized data helps reduce computational requirements. Additionally,Sudhamathi et al. [17] have faced complexity in overfitting issues. The RDPAT model usesa combination of dense connections and attention mechanisms that can help to reduce overfitting. Moreover, Subramaniam et al. [18] faced a drawback in the data dependency problem. The proposed approach uses normalization techniques to remove extraneous data, potentially resolving data dependency issues. Furthermore,Kiruthika et al. [19] had limited generalization issues. The RDPAT model usesa residual dense layer mechanism,which can better generalize new data; it will solve the generalization issues. Finally, Ayesha Barvin et al. [20] discovered that the model had scalability limitations. The proposed model used the DSO algorithm, making it flexible and reducingscalability issues.A comparison study of the performance of the existing models and the proposed model is shown in Table 3.

Table 3: A comparison study of the performance of the existing models and the proposed model

Author Name and Reference	Method	Accuracy	Precision
Wankhade et al. [16]	EDLAM	95.1%	95.74%
Sudhamathi et al. [17]	ER-ETR	95%	-
Subramaniam et al. [18]	WTDCNN	98.96%	-
Kiruthika et al. [19]	DCSO-WLSTM	92.68%	-

Ayesha Barvin et al. [20]	Graph-based Crop Recommendation Algorithms	98%	-
Proposed	RDPAT	99.26%	98.93%

5. Conclusion

In conclusion,Changes in soil fertility have led to a decline in crop production, making it challenging for farmers to select the best crops based on soil conditions. Accurate crop prediction can significantly enhance crop productivity, and machine learning plays a crucial role in this process. Crop forecasting is influenced by soil, geographic, and environmental characteristics, with feature selection being essential for identifying suitable crops. The process starts by gathering input data from publicly accessible datasets that include key agricultural parameters like pH level, K, P, N, rainfall, humidity, and temperature. These parameters have a crucial impact on deciding whether a crop will be suitable in a particular area or not. To make the data consistent and improve its quality, z-score normalization is used for the pre-processing step to normalize the input values. The essential part of the method involves using the DSO algorithm to select the optimal features. This leads to the selection of the most informative features, maximizing the learning ability and speed of convergence of the model. These features are then fed to the RDPAT, anexisting model transformer-based DL model optimized for soil nutrition analysis and accurate crop prediction. RDPAT uses residual connections, dense layers, and progressive attention mechanisms to get and rank informative patterns from the optimized input data to produce highly precise recommendations. The model attains excellent prediction performance with 99.26% accuracy and 98.93% precision,proving efficient in real-world agricultural applications.Future work will include crop recommendations based on climate conditions, as well as the development of a user-friendly web and mobile application for farmers.

Reference

1. Priyadharshini, A., Swapneel Chakraborty, Aayush Kumar, and Omen Rajendra Pooniwala. "Intelligent crop recommendation system using machine learning." In 2021, 5th international conference on computing methodologies and communication (ICCMC), pp. 843-848. IEEE, 2021.
2. Senapaty, Murali Krishna, Abhishek Ray, and NeelamadhabPadhy. "IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture." Computers 12, no. 3 (2023): 61.
3. Mishra, Dipali, and D. Deepa. "Automation and integration of growth monitoring in plants (with disease prediction) and crop prediction." Materials Today: Proceedings 43 (2021): 3922-3927.
4. Raja, S. P., Barbara Sawicka, Zoran Stamenkovic, and G. Mariammal. "Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers." IEEe Access 10 (2022): 23625-23641.
5. Rashid, Mamunur, Bifta Sama Bari, YusriYusup, Mohamad Anuar Kamaruddin, and Nuzhat Khan. "A comprehensive review of crop yield prediction using machine learning approaches with special emphasis on palm oil yield prediction." IEEE Access 9 (2021): 63406-63439.
6. Arunrat, Noppol, Sukanya Sereenonchai, WinaiChaowiwat, and Can Wang. "Climate change impact on major crop yield and water footprint under CMIP6 climate projections in repeated drought and flood areas in Thailand." Science of the Total Environment 807 (2022): 150741.
7. Pergner, Isabell, and Christian Lippert. "On the effects that motivate pesticide use fromthe perspective of designing a cropping system without pesticides but with mineral fertilizer—a review." Agronomy for Sustainable Development 43, no. 2 (2023): 24.
8. Sharma, Vikrant, Satvik Vats, Priyanshu Rawat, and Madhvan Bajaj. "Crop recommendation system: A review." Automation and Computation (2023): 384-396.
9. Xie, Tianjin, Jijun Li, Chenghai Yang, Zhao Jiang, Yahui Chen, Liang Guo, and Jian Zhang. "Crop height estimation based on UAV images: Methods, errors, and strategies." Computers and Electronics in Agriculture 185 (2021): 106155.
10. Lassalle, Guillaume. "Monitoring natural and anthropogenic plant stressors by hyperspectral remote sensing: Recommendations and guidelines based on a

meta-review." Science of the Total Environment 788 (2021): 147758.
11. Rakhmatulin, Ildar, Andreas Kamilaris, and Christian Andreasen. "Deep neural networks to detect weeds from crops in agricultural environments in real-time: A review." Remote Sensing 13, no. 21 (2021): 4486.
12. Bhat, Showkat Ahmad, and Nen-Fu Huang. "Big data and AI revolution in precision agriculture: Survey and challenges." IEEE Access 9 (2021): 110209-110222.
13. Ngo, Vuong M., Thuy-Van T. Duong, Tat-Bao-Thien Nguyen, Cach N. Dang, and Owen Conlan. "A big data smart agricultural system: recommending optimum fertilisers for crops." International Journal of Information Technology 15, no. 1 (2023): 249-265.
14. Cravero, Ania, Sebastian Pardo, Samuel Sepúlveda, and Lilia Muñoz. "Challenges to use machine learning in agricultural big data: a systematic literature review." Agronomy 12, no. 3 (2022): 748.
15. Gupta, Rishi, Akhilesh Kumar Sharma, Oorja Garg, Krishna Modi, Shahreen Kasim, ZirawaniBaharum, HairulnizamMahdin, and Salama A. Mostafa. "WB-CPI: Weather based crop prediction in India using big data analytics." IEEE Access 9 (2021): 137869-137885.
16. Wankhade, Seema Rameshwarrao, and Anjali B. Raut. "Development of a model for estimation of soil parameters using deep learning." International Journal of Information Technology (2024): 1-17.
17. Sudhamathi, T., and K. Perumal. "Ensemble regression based Extra Tree Regressor for hybrid crop yield prediction system." Measurement: Sensors 35 (2024): 101277.
18. Subramaniam, Leelavathi Kandasamy, and RajasenathipathiMarimuthu. "Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops." e-Prime-Advances in Electrical Engineering, Electronics and Energy 8 (2024): 100611.
19. Kiruthika, S., and D. Karthika. "IOT-BASED professional crop recommendation system using a weight-based long-term memory approach." Measurement: Sensors 27 (2023): 100722.
20. Ayesha Barvin, P., and T. Sampradeepraj. "Crop recommendation systems based on soil and environmental factors using graph convolutional neural network: A systematic literature review." Engineering Proceedings 58, no. 1 (2023): 97.
21. Musanase, Christine, Anthony Vodacek, Damien Hanyurwimfura, Alfred Uwitonze, fertilizer recommendation system for revolutionizing farming practices." Agriculture 13, no. 11 (2023): 2141
22. Herwanto, Heru Wahyu, Anik Nur Handayani, Aji PrasetyaWibawa, Katya Lindi Chandrika, and Kohei Arai. "Comparison of min-max, z-score, and decimal scaling normalization for zoning feature extraction on javanese character recognition." In 2021,7th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), pp. 1-3. IEEE, 2021.
23. Cen, Jian, Linzhe Zeng, Xi Liu, Fuyu Wang, Shijun Deng, Zongwei Yu, Guomin Zhang, and Weiyue Wang. "Research on energy-saving optimization method for central air conditioning system based on multi-strategy improved sparrow search algorithm." International Journal of Refrigeration 160 (2024): 263-274.
24. Zhang, Dinglei, Ying Li, and Shihui Ying. "Trans-Net: A transferable pretrained neural network based on temporal domain decomposition for solving partial differential equations." Computer Physics Communications 299 (2024): 109130.
25. Ghosh, Asit, Sunil Kumar Mohapatra, Pritam Pattanaik, Prasant Kumar Dash, and Sujata Chakravarty. "A comprehensive crop recommendation system integrating machine learning and deep learning models." In 2024, 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), pp. 1-6. IEEE, 2024.
26. Avinash, Y. B., and Harikrishna Kamatham. "Optimal Crop Recommendation by Soil Extraction and Classification Techniques Using Machine Learning." In International Conference on Soft Computing and Signal Processing, pp. 189-198. Singapore: Springer Nature Singapore, 2023.
27. SSL, Durai Arumugam, and Praveen Kumar. "Serial cascaded deep feature extraction-

based adaptive attention dilated model for crop recommendation framework." Applied Soft Computing 162 (2024): 111790.

28. Gill, Harsuminder Kaur, Vivek Kumar Sehgal, and Anil Kumar Verma. "A context aware recommender system for predicting crop factors using LSTM." In 2021 Asian Conference on Innovation in Technology (ASIANCON), pp. 1-4. IEEE, 2021.

29. Motwani, Aditya, Param Patil, Vatsa Nagaria, Shobhit Verma, and Sunil Ghane. "Soil analysis and crop recommendation using machine learning." In 2022 International Conference for Advancement in Technology (ICONAT), pp. 1-7. IEEE, 2022.