Engineered Science
DOI: https://dx.doi.org/10.30919/es1481

Deep Convolutional Sequential Network for Digital Soil Mapping and Crop Recommendations Based on Soil Data

Sheela Asare* and Phani Kumar S.

Abstract

Digital Soil Mapping involves using digital techniques to create detailed maps of soil properties across landscapes, utilizing data from remote sensing, soil sampling, and environmental variables. It is essential for effective land management, precision agriculture, and environmental monitoring, providing insights into soil characteristics such as texture, moisture, and nutrient content. However, challenges faced by existing models include inconsistent data quality and availability, the spatial variability of soil properties, difficulties in integrating multisource data, temporal dynamics of soil characteristics, computational complexity, and issues with user accessibility, which can hinder effective implementation and adoption. Hence, to address the existing challenges, this research introduces the explainable selective group enhanced attention-distributed deep convolutional sequential network (X-SGEA-D2CSN) for digital soil mapping (DSM) and crop recommendation. The proposed approach effectively addresses the DSM issues by meticulously capturing the subtle soil features and environmental patterns while filtering out irrelevant noise, ensuring better crop recommendations. Specifically, the proposed model effectively captures the complex patterns and long-term temporal dependencies associated with the soil properties crucial in understanding the dynamic agricultural environments. The experimental results demonstrate that the X-SGEA-D2CSN model achieves better performance, reporting a high accuracy of 94.17%, precision of 95.24%, and recall of 93.1% for 90% training due to the multi-scale feature learning with attention-based refinement, and optimizing the selection of relevant features.

Keywords: Crop recommendation; Bidirectional gated recurrent unit; Long short-term memory; Convolutional neural network; Attention mechanism.
Received: 19 December 2024; Revised: 12 March 2025; Accepted: 03 April 2025.
Article type: Research article.

1. Introduction

Digital soil mapping (DSM) is a geospatial approach that utilizes advanced technologies and data-driven models to create detailed soil property maps, providing insights into soil characteristics such as texture, moisture, nutrient levels, and pH across different landscapes. By integrating various data sources, including satellite imagery, soil samples, and environmental variables, DSM allows for the generation of high-resolution maps that inform agricultural practices and land management decisions. Crop recommendation, on the other hand, leverages the data obtained from DSM to suggest the most suitable crops for specific soil types and conditions, optimizing agricultural productivity and sustainability. This procedure takes into account elements like soil fertility, climatic conditions, and water availability, allowing farmers to make educated choices regarding crop selection that correspond with the distinct characteristics of their land, ultimately improving yields and lessening the environmental effects of agriculture. Specifically, the irrigation systems need to be modernized and regulated to meet the water requirements for agricultural production.[1]

DSM is an innovative approach that leverages geographic information systems (GIS) and advanced modeling techniques, such as geostatistics and machine learning (ML), to create detailed, high-resolution soil maps.[2,3] This method addresses the limitations of traditional soil surveys, which are often costly, time-consuming, and reliant on expert knowledge, by utilizing readily available environmental data to estimate soil properties like organic carbon stock, nutrient availability, and water holding capacity. DSM improves the efficiency of soil mapping by examining the connections between soil properties and different environmental elements, especially topography, which plays a crucial role in soil distribution. By

Department of Computer Science Engineering, GITAM (Deemed to be University), Hyderabad, Rudraram, Telangana, 502329, India
**Email:* sheela.asare@gmail.com (S. Asare)

accurately mapping the spatial variability of soil traits, DSM aids a range of stakeholders, such as soil scientists, land use planners, and environmental managers, in making well-informed choices related to land management and agricultural practices.[4-8] Nonetheless, weak relationships between the input variables and particular soil qualities are frequently observed in the majority of lowland agricultural areas.[9,10] Lowland locations do poorly because of the low-gradient relief of the landscape, which makes it difficult to estimate soil parameters with any degree of accuracy. Various modeling techniques are typically evaluated to select a single "best" model or an "optimal" group of models to address this difficulty and increase prediction accuracy by lowering predicted value uncertainty. The benefit of ML methods lies in their capacity to measure the complex and multidimensional interactions that exist between environmental variables and soil parameters across a range of soil landscapes.[11] The integration of ML techniques in DSM addresses several limitations of conventional soil mapping methods, thereby improving the accuracy of soil attribute predictions.[12,13] Additionally, ML is suitable for DSM even in cases where the data is sparse.[14] To predict the spatial distribution of soil attributes and types, several studies have utilized advanced ML techniques within DSM.[11,15,16] Some of the widely used ML models in DSM include support vector machines, multivariate regression, regression trees, Cubist, random forests, and gradient boosting machines.[17–19]

The development of several ML models has prompted research on model comparisons, wherein, even with identical input data, various models may produce very different digital soil maps.[11,13,15] As a result, it is recommended to evaluate and compare different modeling approaches to identify the most effective one for best practices in DSM.[11,15,20] However, selecting the most successful model can be difficult, as each model has its strengths and weaknesses based on specific contexts. Consequently, a particular model may excel in certain situations and environments compared to others.[21,22] Consequently, ensemble modeling is an additional strategy that aids in combining the data and understanding obtained from individual models.[23,24] Compared to predictions made with single ML models, ensemble models have the potential to produce predictions that are both more stable and better. Additionally, they lessen the possibility of selecting the "wrong" model.[25,26] Common ensemble learning ML algorithms used in DSM are gradient boosting and random forest models, which use a bagging method.[27] However, in DSM studies, there has been less emphasis on developing strategies that utilize various types of ML models as base learners (heterogeneous ensemble learning), since the ensemble models have typically been built using only one type of predictive learner (homogeneous ensemble learning). In addition, most of the existing models account for the linear relationship between soil properties, fail to consider the covariates and overlook the correlations, ignoring the covariates' non-linear association with the soil properties, resulting in subpar performance.[26]

The explainable selective group enhance attention-distributed deep convolutional sequential network (X-SGEA-D2CSN) model effectively captures critical agricultural features through a series of convolutional layers, enhancing interpretability and predictive accuracy. The model employs a selective group attention mechanism to focus on the most relevant data aspects, filtering out noise and improving feature significance. By maintaining long-term dependencies, the X-SGEA-D2CSN allows for a comprehensive understanding of how features evolve, which is crucial for accurate predictions in dynamic agricultural environments.

Additionally, the integration of the selective group enhances attention (SGEA) and enhances multi-scale feature learning, optimizing the selection of meaningful features. The SGEA mechanism contributes significantly to the performance of the X-SGEA-D2CSN model by enhancing multi-scale feature learning and refining extracted features through an attention-based mechanism. It adeptly captures intricate patterns and dependencies at varying levels of granularity, integrating both local and global contextual information to ensure that the model identifies the most relevant features from the data. This approach allows for selective filtering and emphasis on critical aspects, reducing dimensionality and improving the decision-making process. By distilling the input data into a highly informative representation, the SGEA significantly boosts the model's predictive accuracy, making it an invaluable component for crop recommendation based on soil properties derived from DSM.

The X-SGEA-D2CSN model significantly enhances crop recommendation by leveraging DSM to improve predictive accuracy. It uses a dual processing approach to capture sequential dependencies and long-term relationships in soil and environmental data, enabling a deeper understanding of how soil properties evolve. The model excels at extracting intricate patterns, focusing on the most relevant data, improving interpretability, and reducing noise. By processing diverse data sources in a distributed manner, the proposed approach ensures efficient and accurate crop recommendations tailored to specific soil conditions, optimizing decision-making in agriculture.

The paper is structured to provide a clear and coherent flow of information. Section 2 offers an overview of recent developments in the field, highlighting key methodologies and the challenges encountered. In Section 3, the methodology for the DSM and crop recommendation model is detailed, laying the groundwork for subsequent analysis. Section 4 presents the mathematical framework underlying the X-SGEA-D2CSN model. The findings from the implementation of the proposed DSM and crop recommendation model are discussed in Section 5. Finally, Section 6 concludes the manuscript, summarizing the main insights and implications of the research.

2. Literature review

An ML-based DSM approach utilizing quantile regression (QR) was introduced by Babak Kasraei and Brandon Heung,[28] providing a methodological framework for evaluating uncertainty. Their analysis revealed that the predictive interval coverage probability was stable and reliable across various studies, ML techniques, and projected soil variables; however, the model exhibited significant computational complexity.

To address measurement errors in DSM, Stephan van der Westhuizen *et al*.[29] proposed a two-stage maximum likelihood method, which enhanced prediction accuracy, particularly when the measurement error variance was substantial and heterogeneous, although it required extensive training time. Odunayo David Adeniyi *et al*.[2] developed an ensemble learning model based on a stacking strategy to predict the geographical variation of soil parameters such as pH and topsoil depth, which improved prediction accuracy by reducing uncertainty, yet maintained considerable computational complexity. Virginia Estevez *et al*.[3] introduced three distinct supervised ML techniques, support vector machine (SVM), gradient boosting (GB), and random forest (RF) for mapping acid sulfate soils, finding that GB and RF produced highly precise probability maps, while the SVM struggled with overestimating non-acid sulfate soil occurrences. Hassan Fathizad*et al*.[4] employed an RF model to forecast the spatial distribution of soil quality indices, achieving a coefficient of determination of 0.69, which could inform environmental studies and land management, though the model faced significant overfitting issues. Solmaz Fathololoumi *et al*.[5] aimed to predict soil moisture (SM) using multi-temporal satellite images alongside static and dynamic environmental variables, achieving accurate predictions but also grappling with high overfitting. S. Dharumarajan*et al*.[6] tested the quantile regression forest (QRF) and regression kriging (RK) methods for forecasting soil depth in Karnataka, enhancing model accuracy through data augmentation and the removal of erroneous profile points, though computational complexity remained high.

Moreover, Kamal Nabiollahi *et al*.[7] analyzed the spatial variability of soil salinity and sodicity in 436 km² of agricultural land in Kurdistan Province, Iran, using a combination of RF and covariate data, and their findings indicated elevated pH, EC, and SAR values in the central region, highlighting the challenge of determining optimal hyperparameter values for ML models, which significantly affects their ability to accurately predict complex soil issues. Isaac Kofi Nti*et al*.[30] presented the tree-based ensemble learning model for predicting crop suitability and productivity. Specifically, the tree-based approach achieved better performance and assisted in making informed decisions concerning crop selection. However, the tree-based approach resulted in reduced performance on the training data but offered better generalization over unseen data.

Further, the base learners shared the similarities or features, resulting in inadequate information and subpar performance. Shubhadip Dasgupta *et al*.[31] introduced the hybrid ensemble ML-based approach for soil samples collected from four different agro-climatic regions. In addition, the hybrid approach carried out the spatial prediction and mapping of the four available. However, the model required recalibration with the introduction of new crop varieties for a significant percentage of the area under consideration. Meanwhile, the correlation associated with soil-grain micronutrient content varies for other agroclimatic regions, which affects the generalization performance.

Therefore, the conventional methods have several challenges, which include inconsistent data quality and availability, the spatial variability of soil properties, difficulties in integrating multisource data, temporal dynamics of soil characteristics, computational complexity, and issues with user accessibility, which can hinder effective implementation and adoption. Thus, to overcome these challenges, the current research proposed an X-SGEA-D2CSN model for DSM and crop recommendation. The model employs a selective group attention mechanism to focus on the most relevant data aspects, filtering out noise and improving feature significance. Additionally, the integration of the SGEA enhances multi-scale feature learning, optimizing the selection of significant features.

3. Crop recommendations framework

3.1 Study area

In this research, the soil map of Maharashtra serves as a pivotal tool in overcoming the challenges faced in crop recommendation systems, particularly by offering valuable insights into the region's diverse soil types, climate conditions, and available resources. Inspired by the agro-climatic diversity of regions like the indo-gangetic plain (IGP), where distinct zones are categorized based on their soil characteristics; our approach aims to replicate this method for Maharashtra. The soil map of Maharashtra uses a color-coded system to classify regions, such as banana mania (yellow) areas for loamy soils that are ideal for fertility and ideal for growing crops like rice, wheat, sugarcane, and paddy. Lavender blue areas for light clay soil that supports a wide variety of crops like fruits and vegetables. Lavender-pink areas for sandy loam soil that support a wide variety of crops like wheat and rice. Pale green for glacier or water bodies soils that support a wide variety of crops like sugarcane. This strategic soil mapping approach will enable us to enhance crop recommendation systems by ensuring that they are region-specific and aligned with soil characteristics, ultimately improving crop yield and resource management. The primary motivation behind using this soil map-based method is to bridge existing gaps in current crop recommendation models, such as poor data fusion, low prediction accuracy, and lack of transferability between regions. Moreover, by ensuring that the system is interpretable and user-friendly, farmers will be able to easily understand and trust the recommendations, allowing them to make more informed decisions regarding

crop selection and land management. The application of this soil map-driven approach will lead to better land use planning, more sustainable farming practices, and increased agricultural productivity across Maharashtra. In essence, the soil map becomes a strategic tool for optimizing crop choices and also enhancing the overall effectiveness of agricultural systems, contributing to long-term food security and environmental sustainability. The soil map of Maharashtra is shown in Fig. 1 and Table 1 depicts the different colors that represent the various types of soil and their pH value, Potassium, Phosphorus content, Humidity level, and the recommended crops.

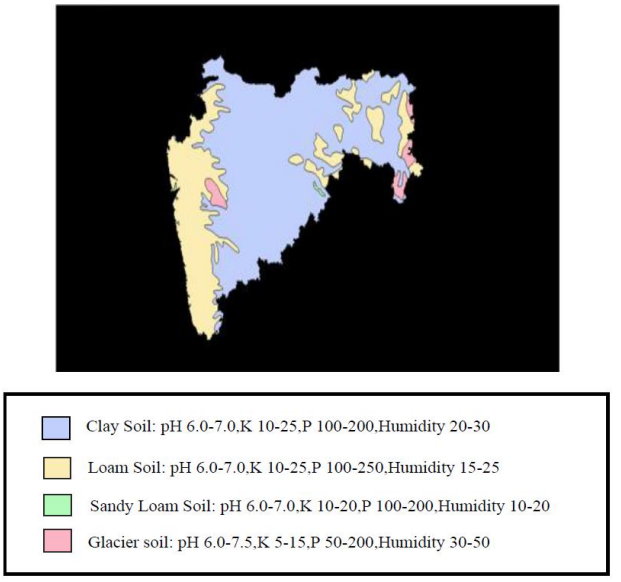


Fig. 1: Soil map of Maharashtra.

3.2 Problem statement

The primary problem addressed by this research is the challenge of providing accurate and reliable crop recommendations based on DSM due to the complex and dynamic nature of agricultural environments. The variations in soil properties, climate conditions, and other environmental factors contribute to the uncertainty in predicting suitable crops for specific locations. The existing models often struggle to adapt to these changing conditions, leading to inadequate recommendations that do not account for the intricate interactions between soil characteristics and agricultural practices. Therefore, there is a pressing need for advanced methodologies that can effectively analyze and interpret DSM data to generate precise and context-sensitive crop recommendations, ultimately supporting sustainable agricultural practices and optimizing crop yields. The current research proposes to design the X-SGEA-D2CSN model to analyze and process various data inputs and provide increased explainability of the results to guarantee enhanced and credible crop recommendations. The model for DSM and crop recommendation is based on data acquired from the Soil Grid Dataset and can be mathematically expressed as Eq. (1).

$$G_d = \sum_{n=1}^{t} G_n \tag{1}$$

where G_d refers to the entire Soil grid database, while G_n indicates the count of images found within the dataset, which varies from 1 to t. The image input undergoes a preprocessing stage that employs image sharpening techniques using the OpenCV (cv2) filter,[32] enhancing the quality and clarity of the images. The sharpened version of the image is denoted as G_n^* in Eq. (2).

$$G_n^* = G_n + \lambda.\left(G_n - CV2.filter(G_n, ker\,n\,el)\right) \tag{2}$$

where the sharpening factor is denoted as $\lambda, CV2.filter(G_n, ker\,n\,el)$ applies a convolution filter to G_n. Following the pre-processing, the refined image data is forwarded to a feature extraction stage, where various vegetation indices, such as 23 features, are computed to derive significant features indicative of crop health and growth potential. The soil organic carbon model utilizes data collected from crop recommendation data and is represented in mathematical form as described in Eq. (3).

$$H_e = \sum_{n=1}^{u} H_n \tag{3}$$

where H_e denotes the crop recommendation database, and H_n denotes each sample of the dataset. The total samples in the dataset are denoted as u, each row H_n contains 10 features (soil properties).

The crop recommendation data is pre-processed using k-nearest neighbors (KNN) imputation to handle any missing values effectively. Meanwhile, the pre-processed crop recommendation data undergoes statistical feature extraction to capture relevant agronomic attributes. Finally, both sets of extracted features, image features, and statistical features from the crop recommendation data, are sent separately to the X-SGEA-D2CSN model. This model leverages these comprehensive inputs to generate precise and reliable crop recommendations, facilitating informed agricultural decision-making. The model employs categorical cross-entropy loss in Eq. (4), which is ideal for scenarios with multiple output categories, to recommend 22 crop types. This loss function measures the difference between the actual class distribution and the model's predicted probabilities, penalizing incorrect predictions more heavily. It ensures higher predicted probabilities for the correct class by increasing the loss when predictions are inaccurate. The objective is to minimize this loss during training, enabling the model to adjust its parameters and enhance its ability to accurately recommend crops over time.

$$Loss = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij}\, log\left(p_{ij}\right) \tag{4}$$

where the total number of samples in the batch is denoted as N, the total number of classes is denoted as$c$, y_{ij} is the binary indicator (0 or 1) that indicates whether class j is the correct label for observation$i$, p_{ij} denotes the predicted probability that observation i belongs to class j.

3.3 System model

The model depicted in the image outlines a crop

Table 1: Types of soil in Maharashtra

Sl. No.	Colors	Soil type	pH value	Potassium content (ppm)	Phosphorous content (ppm)	Humidity (%)	Recommended crops
1.	Lavender (blue)	Clay (light) soil	6.0 to 7.0	10 – 25	100 - 200	20 - 30	Vegetables (kidney beans)
2.	Banana Mania (yellow)	Loam soil	6.0 to 7.0	10 – 25	100 - 250	15- 25	Fruits (apples)
3.	Lavender (pink)	Sandy loam Soil	6.0 to 7.0	10 – 20	100 - 200	10 - 20	Rice
4.	Pale (green)	Glacier / Water body's boil	6.0 to 7.5	5 – 15	50 - 200	30 - 50	Coconut

recommendation framework using satellite imagery and deep learning (DL) for precise agricultural decisions. Satellite images are captured and processed to identify specific soil regions. Key soil parameters such as temperature and humidity are measured within these regions. These parameters are then fed into a feature extraction process, where critical features are identified for further analysis. The extracted features are input into a deep learning model, which processes the data to recommend suitable crops for the given soil conditions. The model evaluates multiple conditions and suggests either acceptance or rejection of a specific crop recommendation based on the analysis, optimizing agricultural productivity. Fig. 2 illustrates the system model for the crop recommendation framework.

4. Proposed X-SGEA-D2CSN model

The main purpose of this research is to propose a reliable crop recommendation system using the X-SGEA-D2CSN model. The process begins with the collection of two types of inputs, the soil grid database and crop recommendation data. This image input then passes through pre-processing, where image sharpening through an OpenCV (cv2) filter is applied to the input image to produce better image quality with great contrast and clarity. At the same time, the crop recommendation data is also prepared for analysis using KNN imputation in case any missing value is encountered during the analysis process.

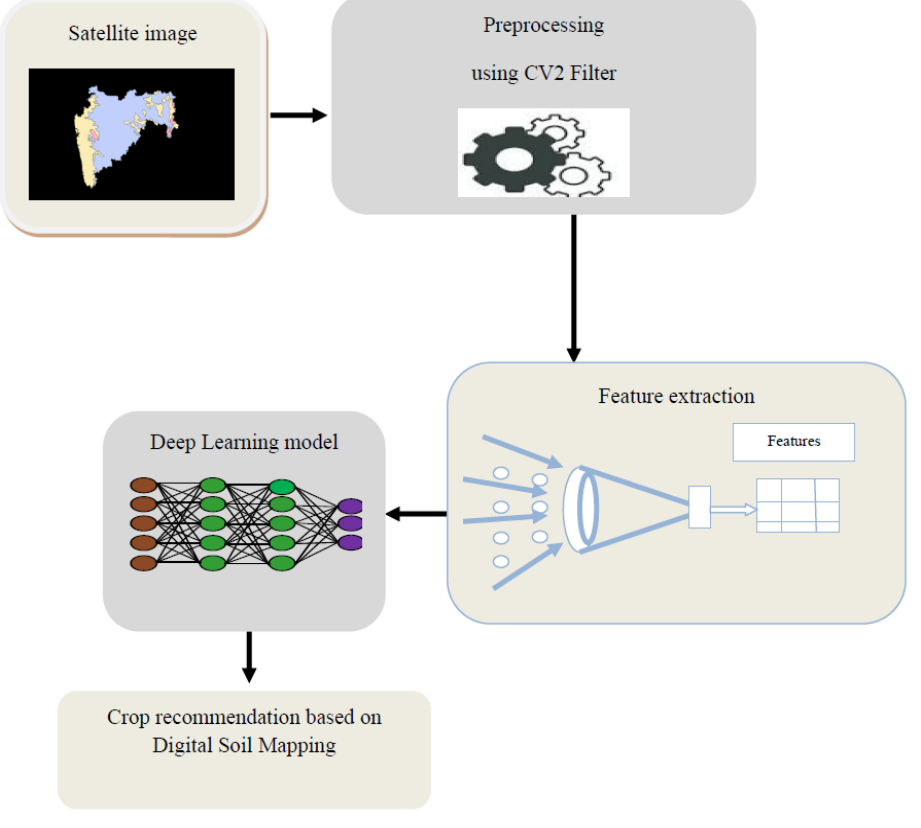


Fig. 2: System model for the crop recommendation model.

After pre-processing, the refined image is advanced to a feature extraction stage, where a variety of vegetation indices are calculated. These indices include the normalized green red difference index (NGRDI), advanced ratio vegetation index 12 (ARV12), chlorophyll vegetation index (CVI), normalized difference vegetation index (NDVI), green normalized difference vegetation index (GNDVI), adjusted soil-adjusted vegetation index (ATSAVI), chlorophyll index green (CI_GREEN), color transformation vegetation index (CTVI), green difference vegetation index (GDVI), global environment monitoring index (GEMI), green optimized soil-adjusted vegetation index (GOSAVI), green soil-adjusted vegetation index (GSAVI), infrared percentage vegetation index (IPVI), ratio vegetation index (RVI), modified ratio vegetation index (MRVI), modified soil-adjusted vegetation index (M_SAVI), Normalized Green Reflectance (NORM_G), Normalized Near-Infrared Reflectance (NORM_NIR), normalized red reflectance (NORM_R), rescaled index (RI), difference vegetation index (DVI), and transformed vegetation index (TVI). These features provide valuable insights into the soil's health and micronutrient levels, which are crucial for effective agricultural management. On the other hand, statistical feature extraction is applied to the pre-processed crop recommendation data to obtain agronomic features. Lastly, the outcome obtained from both the vegetation topographic feature extraction and statistical features is sent separately to the X-SGEA-D2CSN model, which takes these comprehensive inputs and provides accurate crop recommendations to help the agricultural field in decision-making. The proposed architecture for the DSM and crop recommendation model is depicted in Fig. 3.

4.1 Soil mapping

Initially, the input soil images obtained from the Soil grid dataset are pre-processed and the vegetation index-based features are extracted to analyze the nutrient content, soil condition, change of chlorophyll concentration, biomass, leaf area index, moisture level, and so on. Further, the information obtained from feature extraction is fed to the X-SGEA-D2CSN model that effectively carries out the digital soil mapping.

4.1.1 Input satellite image for DSM

For DSM, the input soil images are acquired from the Soil grid dataset and each image present in the dataset $G_d = \{G_1, G_2 \ldots . G_k, \ldots . G_n\}$ has a dimension of $(1129,585,3)$, where G_d indicates the database, G_n indicates input images, and n indicates the total number of images present in the database.[33]

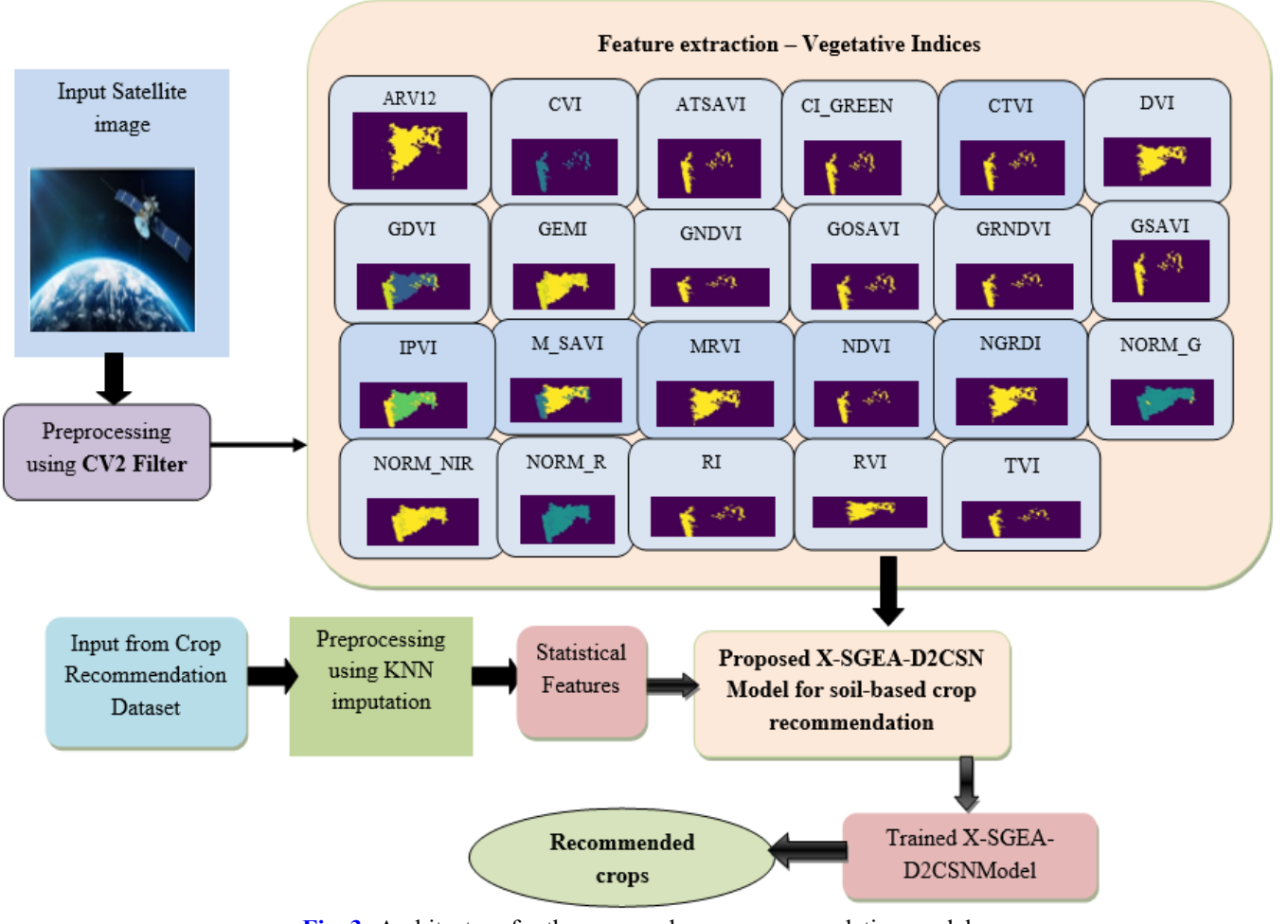


Fig. 3: Architecture for the proposed crop recommendation model.

4.1.2 Preprocessing using the CV2 filter

In the preprocessing stage, each image from the selected dataset undergoes sharpening using an image sharpening technique, which is essential for enhancing the quality of the images. The CV2 filter improves the edge definition, allowing for clearer delineation of changes in soil texture and vegetation cover boundaries, which is critical for analysis. In particular, fine details significantly contribute to the classification of soil organic carbon and feature extraction in satellite images. These distinctions are enhanced through image sharpening, as it sharpens the edges of objects and eliminates blurring effects. Hence, the pre-processed images (as shown in Eq. (2)) appear clearer and provide improved input for the subsequent model.

4.1.3 Feature extraction from region-wise soil images

For the DSM, information regarding the soil properties and vegetation health is essential for which the vegetative indices are used. In this research, feature extraction focuses on identifying and quantifying vegetation indices and metrics that indicate soil conditions and water extent. These indices are derived from spectral reflectance data obtained from satellite imagery, enhancing the assessment and monitoring of land cover types, particularly in wetland-upland interfaces, and evaluating vegetation responses to abiotic stress factors.

The NGRDI is a normalized ratio that quantifies the difference between the green and red spectral bands, specifically designed to mitigate the effects of varying irradiance on the spectral properties of vegetation. This index enhances the reliability of vegetation analysis by providing a more accurate reflection of plant health and vigor. The formula for calculating the NGRDI is expressed in Eq. (5).[34]

$$NGRDI = \frac{G(G_n^*) - R(G_n^*)}{G(G_n^*) + R(G_n^*)} \tag{5}$$

where the green band of the image is denoted as G and the red band of the preprocessed dataset is denoted as R.

ARVI2 is a vegetation index based on the amplification of vegetation spectral features regarding the differences in specially selected bands, changes in the chlorophyll index, or vegetation stress.[35] This index assists us in differentiating between one vegetation type and its conditions given in Eq. (6).

$$ARVI2 = \frac{NIR(G_n^*) - R(G_n^*)}{NIR(G_n^*) + R(G_n^*)} \tag{6}$$

where the preprocessed data for the NIR and the input image G_n^* is denoted as CVI, which is the other index that is used in the calculation of chlorophyll content in vegetation. It is possible to notice that higher chlorophyll content usually points to the fact that the vegetation is healthier. The CVI is useful in the regular assessment of vegetation vigor, stress, and overall condition, which is explained in Eq. (7).

$$CVI = \frac{NIR(G_n^*) \times R(G_n^*)}{G(G_n^*)^2} \tag{7}$$

where the preprocessed reflectance values for the NIR, red, and green bands necessary for computing the CVI are denoted as G_n^*. The GRNDVI is a vegetation index that focuses on green vegetation using the green and red spectral regions of the electromagnetic spectrum.[36] It assists in making the distinction between healthy vegetation and other cover on land and can be very useful in distinguishing between vegetation from soil, and so on, also expressed as *GRNDVI* in Eq. (8).

$$GRNDVI = \frac{NIR(G_n^*) - (G(G_n^*) + R(G_n^*))}{NIR(G_n^*) + (G(G_n^*) + R(G_n^*))} \tag{8}$$

where reflectance values from the green and red bands are denoted as $G(G_n^*)$ and $R(G_n^*)$. ATSAVI is designed to reduce the impact of the soil on the vegetation indices. Hence, it is useful in evaluating the health of vegetation. This index is also corrected for soil brightness and is particularly relevant in heterogeneous settings,[37] which is given as Eq. (9).

$$ATSAVI = \frac{\alpha(NIR(G_n^*) - aR(G_n^*) - L)}{R(G_n^*) + aNIR(G_n^*) - aL + 0.08(1 + a^2)} \tag{9}$$

where α indicates the weighting coefficient, a indicates the slope of the soil line, and L indicates the intercept of the soil line, are chosen based on empirical studies. CI_GREEN is calculated as an index concerning the amount of chlorophyll present in vegetation to assess plant health and stress, given in Eq. (10). This is particularly important for the evaluation of green vegetation cover and may be used in precision agriculture and environmental management.[38]

$$CI\_GREEN = \frac{NIR(G_n^*)}{G(G_n^*)} - 1 \tag{10}$$

where $NIR(G_n^*)$ refers to the reflectance in the near-infrared range derived from the processed input, while $G(G_n^*)$ indicates the reflectance within the green spectral band from the processed input.

The GSAVI is an improved version of the SAVI that minimizes the impact of soil influence in areas with low vegetation density. Unlike the earlier SAVI, which incorporates the red band, GSAVI integrates both the green and near-infrared (NIR) bands. This enhancement increases the sensitivity of NDVI to vegetation greenness, particularly in sparse vegetation areas where ground exposure is significant. Additionally, GSAVI calculates a soil adjustment factor, reducing the influence of soil reflectance on vegetation condition assessments.

The NDVI is a key method for evaluating crop health in agriculture, acknowledging that spatial variability in soil properties necessitates different nitrogen applications within a field to optimize yield. After collecting NDVI point data, geostatistical methods are used to create a spatial continuity surface, aiding in precision agriculture. NDVI relies on the spatial and spectral resolutions of multispectral remote sensing images, distinguishing vegetation based on its absorption in the red and blue segments of the visible spectrum and strong reflectance in the near infrared. The IPVI also assesses

vegetation quantity through near-infrared and red-light reflectance, with values ranging from 0 to +1, and features such as iso-vegetation lines converging at the origin and a soil line with a slope of 1 explained in Eq (11).[37]

$$IPVI = \frac{NIR(G_n^*)}{NIR(G_n^*)+R(G_n^*)} \quad (11)$$

The NORM_G index is most useful in distinguishing between vegetation and other types of land cover, especially where vegetation cover is likely to be low or where vegetation may overlay other land cover types such as soil or water. By dividing the difference of the green reflectance by other effective spectral bands, NORM_G minimizes the impact of soil background and atmospheric factors to properly evaluate the vegetation activity. The RVI is a good remote sensing tool for vegetation conditions, which is an index of the ratio of NIR reflectance to red reflectance. Comparing the NIR and red lights, moderately healthy vegetation has a high reflectance of NIR and a high absorbance of red light. Hence, the RVI is a good measure of vegetation health. It can be easily calculated from satellite imagery, which helps on a vast scale and provides an assessment of the dynamics of vegetation. The RVI is extremely useful for assessing the growth of crops, monitoring changes in the land cover, and guiding the environmental practices for agricultural and ecological development, expressed in Eq. (12).[39]

$$RVI = \frac{R(G_n^*)}{NIR(G_n^*)} \quad (12)$$

where NIR refers to the reflectance of the near-infrared band, while R represents the reflectance of the red band. Additionally, NORM_NIR is a metric that normalizes the reflectance of the NIR band using a standard scale ranging from 0 to 1

This normalization operation makes it possible to compare NIR reflectance values obtained on different images and under various conditions, which will, in turn, enable a more accurate evaluation of vegetation state and land cover. NORM_NIR increases the accuracy for further analyses using remote sensing data by normalizing the NIR standardized variable and enables tracking the vegetation change substantially expressed in Eq. (13) as follows.[40]

$$NORM\_NIR = \frac{NIR(G_n^*)}{NIR(G_n^*)+R(G_n^*)+G(G_n^*)} \quad (13)$$

Like NORM_NIR, NORM_R derives from the normalization of the band reflectance containing the red color to a scale of 0 to 1 or any other standard scale of reflectance. Normalized red reflectance facilitates comparisons of this color in different images and situations, providing more accurate assessments of the state of vegetation and other types of coverage. NORM_R generalizes red band values and improves the analyses of the remote sensing, assisting the monitoring of vegetation and temporal changes in the land cover, which is explained in Eq. (14).

$$NORM\_R = \frac{R(G_n^*)-R_{min}}{Rmin_{max}} \quad (14)$$

where the minimum and maximum reflectance values in the red band are denoted as R_{min} and R_{max}. The RI is a relatively simple index that estimates vegetation cover by proportioning the reflectance of NIR to the red region. This ratio gives information on vegetation density and health because healthy vegetation returns more NIR and absorbs more red light. By quantifying this relationship, the RI enables the assessment of vegetation conditions and the changes in plant cover over time, expressed in Eq. (15) as follows.

$$RI = \frac{NIR(G_n^*)}{R(G_n^*)} \quad (15)$$

The DVI is a simple index derived from the reflectance of the red band subtracted by the reflectance of the NIR band. This is an easy way of evaluating the status of the vegetation because healthy vegetation will give a higher NIR than red reflectance and explained in Eq. (16). DVI, therefore, holds the advantage of offering a measure of comparison on vegetation density and health with which to assess plant status and trends in land cover.[41]

$$DVI = NIR(G_n^*) - R(G_n^*) \quad (16)$$

The TVI is a metric that combines both NIR and red reflectance to create a transformed measure that is particularly sensitive to changes in vegetation. By effectively capturing the relationship between these two reflectance bands, TVI enhances the detection of variations in vegetation health and density. This sensitivity makes TVI a valuable tool for monitoring plant conditions, enabling more accurate assessments of land cover and ecological changes over time. NIR and red reflectance are added together to form the TVI, which is a quantification of a kind of transformed measure that is especially responsive to vegetation fluctuations. Through successful quantification of this relation between two reflectance bands, TVI improves the ability to identify fluctuations in vegetation health and cover. This sensitivity of the TVI expressed as Eq. (17), makes it a more useful tool for enhancing the measurement of the status of plant conditions hence facilitating improved assessment of the coverages and other ecological changes on the land.[39]

$$TVI = \sqrt{\frac{(NIR(G_n^*)-R(G_n^*))}{(NIR(G_n^*)+R(G_n^*))}+0.5} \quad (17)$$

The GEMI is a vegetation index that is derived to reduce the impact of soil and atmospheric on remotely sensed vegetation data. The current research reveals that GEMI is very useful in enhancing the accuracy of vegetation monitoring while considering differences in the background of the soil and the atmosphere. This is derived from the reflectance from the red and near-infrared bands of the satellite imagery with certain equations to normalize the effect of the two to ensure a more accurate vegetation health index is obtained. The GEMI index also aids in differentiating among

the different vegetation and performance of changes on vegetation at some point in time to give a better evaluation of the environment and the vegetation cover status, given in Eq. (18) as follows.

$$GEMI = \frac{h.(2.h-120.(rS+G_n^*))-0.25(rS+G_n^*)}{(rS+G_n^*)-0.125} \quad (18)$$

where h denotes the GEMI equation combining squared and linear terms of reflectance, and reflectance in the red band is denoted as rS.

Remotely sensed vegetation reflectance serves as an indirect indicator of soil salinity, as soil salinity significantly affects vegetation growth. Salt is a critical factor that limits plant development due to its impact on soil chemistry. As the salt content in the surface soil rises, both vegetation cover and vegetation indices tend to decrease, while soil electrical conductivity correspondingly increases. The modified soil-adjusted vegetation index (MSAVI) addresses this challenge by accounting for the bare soil line, effectively minimizing the background influence of soil on the vegetation canopy.[37] This makes MSAVI a valuable tool for more accurately assessing vegetation health in the context of soil salinity. The formula for MSAVI is given in Eq. (19):

$$MSAVI = \frac{2.(\rho_{nir}+G_n^*)+1-\sqrt{[2.(\rho_{nir}+G_n^*)+1]^2-8.(\rho_{nir}+G_n^*-\rho_{red})}}{2} \quad (19)$$

where near-infrared band reflectance is denoted as ρ_{nir}, and red band reflectance is denoted as ρ_{red}.

The MRVI is a vegetation index focused specifically on increasing the monitoring efficiency of vegetation cover with reduced susceptibility to soil background and atmospheric conditions affecting the results. MRVI is based on the modification of the traditional red-based vegetation indices and the variation of the weighting factors of the red spectral band for more appropriate and better identification of vegetation in the heterogeneous large areas. This index is especially good where the red reflectance is affected by such factors as the brightness of the soil or shadowing, which complicates vegetation estimation. The GOSAVI is a vegetation index intended to improve the vegetation assessment by minimizing the impact of the brightness of soil, which in turn affects the data captured through remote sensing. GOSAVI is an improvement of SAVI, which takes a unique soil adjustment factor in such a way that enhances the Vegetation Index for different environments.[37] It is most useful in areas where soil reflectance interferes with the vegetation signal, which improves the accuracy of vegetation classification and biomass estimation.

The GNDVI is a vital vegetation index used to assess plant health and greenness by utilizing the green band of the electromagnetic spectrum, as opposed to the red band employed in other indices like the NDVI. The GNDVI is calculated using the following formula given in Eq. (20).[36]

$$GNDVI = \frac{NIR(G_n^*)-G(G_n^*)}{NIR(G_n^*)+G(G_n^*)} \quad (20)$$

where the green band reflectance is denoted as G.

The CTVI is an improved vegetation index that incorporates thermal and spectral data for a better evaluation of the health status of vegetation and the level of stress it undergoes. Compared with the existing indices, which are calculated based only on reflection coefficients, the application of temperature data in the CTVI index allows the identification of plant moisture and thermal stress conditions. Integrating thermal data helps CTVI differentiate healthy and stressed vegetation better during water availability changes. CTVI uses thermal information together with spectral bands like NIR and green (G) bands to provide a rich view of vegetation status. In this research, output G_n^* that goes through a pre-processing process is incorporated into CTVI computation to enhance the thermal and spectral inputs, which enhances the reliability of the index. The following enhanced CTVI improves the observation of agricultural yield, drought stress, and vegetation health, making it helpful in assessing the environment. The output dimension obtained from the vegetation and topographic feature (VTF) Extraction outcome is $(N, 22)$. The final vegetation and topographic features from the image are denoted in Eq. (21):

$$I_{image} = [NGRDI//ARV12//CVI//NDVI//GNDVI//ATSAVI//CI\_GREEN//CTVI//GDVI//GEMI//GOSAVI//GSAVI//IPVI//RVI//MRVI//MSAVI//NORMG//NORMNIR//NORMR//RI//DVI//TVI] \quad (21)$$

4.2 Crop data for recommendation

The crop informationis required for the recommendation prediction and the crop data is represented in Eq. (22) as follows.[42]

$$H_e = \sum_{n=1}^{u} H_n \quad (22)$$

where H_e denotes the crop recommendation database, H_n denotes each sample of the dataset. The total number of samples or rows in the dataset is denoted as u, each row H_n contains 10 features (soil properties), some of which may have missing values. The input dimension obtained for the crop recommendation dataset is (N, 10). The data is preprocessed to represent the clean input for the feature extraction. During the data cleaning, missing values in each row H_n are imputed using KNN Imputation, which involves identifying the k-nearest neighbors based on the available features. This method estimates the average value of the missing features from these neighbors, effectively providing a functional approach to handle incomplete data. Here, the whole preprocessed dataset is indicated as H_e^*, H_n^* refers to each row after applying the KNN Imputation addressing the missing features with an imputation function $imputation(H_n)$. For a missing feature y_i in row H_n, the imputed value $y_i^{'}$ is computed in Eq. (23) as follows.

$$y_i' = \frac{1}{k}\sum_{j=1}^{k} y_j \quad (23)$$

where the imputed value for the feature i in row H_n is denoted as y_i', KNN values for feature i are denoted as y_j, and the number of nearest neighbors used for imputation is denoted as k. After imputing all missing values, each row H_n becomes a fully imputed version, which is denoted as H_n^* expressed in Eq. (24).

$$H_e^* = \sum_{n=1}^{u} H_n^* = \sum_{n=1}^{u}\left(H_n + imputation(H_n)\right) \quad (24)$$

4.3 Feature extraction using statistical features

The statistical features for feature extraction entail obtaining key metrics of the dataset that quantify the fundamental attributes. The statistical features, including mean, standard deviation, median, variance, skewness, and kurtosis, are extracted from the data. From these features, insights related to soil and crop recommendations are derived from the extracted features to improve decisions in the management of agriculture.

The mean feature describes the central tendency of the data, which is obtained through averaging the data attributes. In this research, the mean is utilized as the reference point, facilitating the assessment of individual data points, pattern analysis, and improved decision-making in crop recommendations, as explained in Eq. (25).[43]

$$M = \frac{1}{N}\sum_{j=1}^{N} H_{e,i}^* \quad (25)$$

where total samples are denoted as N, individual data points from the preprocessed dataset are denoted as $H_{e,i}^*$.

Standard deviation is a statistical metric that quantifies the extent of variation or dispersion in a dataset relative to its mean. A low standard deviation indicates that the data points are clustered closely around the mean, suggesting consistency within the dataset. In contrast, a high standard deviation reflects a broader spread of values, indicating greater variability and less predictability in the data attributes. In this research, standard deviation is utilized to assess variability in crop features, offering insights into the consistency and reliability of the collected data given in Eq. (26) as follows.[43]

$$SD = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(H_{e,i}^* - M\right)^2} \quad (26)$$

Variance is an essential statistical measure that quantifies the dispersion of data attributes within a dataset. A higher variance indicates a wider spread among data points, reflecting greater variability and less consistency within the data. Conversely, a lower variance signifies that the data attributes are more closely clustered around the mean, suggesting greater uniformity. In this research, variance is essential for understanding soil variability and crop characteristics, aiding in informed agricultural practices and decision-making, which is expressed in Eq. (27).

$$V = \frac{1}{N}\sum_{j=1}^{N}\left(H_{e,i}^* - M\right)^2 \quad (27)$$

Unlike the mean, which can be affected by extreme values or outliers, the median is more robust and provides a more accurate reflection in assessing the crop features, especially when there are significant outliers that could skew the results in Eq. (28).

$$Median = median(H_e^*) \quad (28)$$

Skewness measures the asymmetry of a dataset's distribution, indicating whether it leans left or right. A negative skewness indicates a leftward skew, while a positive value signifies a rightward skew. In a perfectly symmetrical distribution, skewness equals zero.[43] A skewness value greater than zero suggests left skewness, meaning the frequency distribution has more values towards the left of the mean with a long tail to the right. Understanding skewness is crucial in this research for analyzing the distribution of soil and crop data, helping to identify any biases or outliers.

$$SK = \frac{\frac{1}{N}\sum_{j=1}^{N}\left(H_{e,i}^* - mean\right)^3}{\left(\frac{1}{N}\sum_{j=1}^{N}\left(H_{e,i}^* - mean\right)^2\right)^{3/2}} \quad (29)$$

Eq. (29) explains Kurtosis measures how a distribution differs from a normal curve, providing insights into the level of variation in data values compared to a normal distribution. High kurtosis indicates a peaked distribution with potential outliers, while low kurtosis suggests lighter tails and a more even distribution.[43] In this research, analyzing kurtosis aids in identifying extreme information on the crop data, which can significantly influence the recommendations is expressed in Eqs. (30)-(33).

$$K = \frac{\frac{1}{N}\sum_{j=1}^{N}\left(H_{e,i}^* - mean\right)^4}{\left(\frac{1}{N}\sum_{j=1}^{N}\left(H_{e,i}^* - mean\right)^2\right)^2} - 3 \quad (30)$$

The maximum value is useful for assessing the potential optimum outcomes in crop yield, allowing for effective decisions in agricultural sectors.

$$Maximum = max(H_e^*) \quad (31)$$

The minimum value is the lowest data value that can be found in the dataset, which makes us the smallest value range to be expected. These minimum values assist in identifying the negative factors impacting crop yields and the soil properties, thereby providing measures for the improvement of agricultural practices for sustainability.

$$Minimum = min(H_e^*) \quad (32)$$

The final statistical feature outcome is denoted as:

$$I_{CSV} = [M//SD//V//Median//SK//K//Maximum//Minimum] \quad (33)$$

The output dimension obtained from the statistical feature extraction outcome is $(N, 8)$.

4.4 X-SGEA-D2CSN in DSM and crop recommendation

The output features obtained from the vegetation and topographic feature extraction $(N, 22)$ and the statistical feature extraction $(N, 8)$ form the input for the X-SGEA-

D2CSN model in DSM and crop recommendation. Initially, the model processes two sets of features, the VTF and SF, through multiple convolutional layers, producing progressively refined feature maps that capture spatial and contextual characteristics essential for crop classification. The output from these feature maps is concatenated and fed into a dual-directional processing mechanism, allowing the model to learn from both historical and future data sequences. This dual processing is crucial for understanding temporal dependencies in soil and crop data. The SGEA mechanism further refines the features by focusing on the most relevant information, effectively filtering out noise and enhancing the model's ability to detect intricate patterns across various scales. This attention-based refinement ensures that critical features significantly influence the model's predictions, leading to more accurate crop recommendations. After training, explainability techniques are utilized to interpret the model's decisions, providing visual insights into which input features contribute most to the output predictions.

For the VTF features, the initial processing step applies a convolution operation with a specific kernel size and bias term, producing an output tensor of dimensions (N, 12, 8). This process is repeated in subsequent steps, further refining the feature representation and resulting in progressively smaller and more complex feature maps with dimensions (N, 6, 16) and (N, 3, 32), respectively. Similarly, the SF undergoes three processing steps, beginning from the input and gradually extracting deeper features, leading to outputs of dimensions (N, 5, 8), (N, 5, 16), and ultimately (N, 2, 32). The kernels utilized in these steps are critical as they define the spatial and contextual characteristics learned by the model from the data, while the biases adjust the output of the operations, enhancing the model's capability to accurately identify patterns crucial for crop type prediction, as explained in Eqs. (34)-(40) as given below.[44]

Data processing for SF involves:

$$P_1 = Conv(I_{image}, L_1) + b_1 \tag{34}$$

$$P_2 = Conv(P_1, L_2) + b_2 \tag{35}$$

$$P_3 = Conv(P_2, L_3) + b_3 \tag{36}$$

$$P_4 = Conv(I_{data}, L_4) + b_4 \tag{37}$$

$$P_5 = Conv(P_4, L_5) + b_5 \tag{38}$$

$$P_6 = Conv(P_5, L_6) + b_6 \tag{39}$$

where the convolutional kernel for i^{th} layer is L_i and the corresponding bias is denoted as b_i. The outputs from the convolutional layers for both VTF features and SF are concatenated:

$$P_{concat} = Concat(P_3, P_6) \tag{40}$$

In the X-SGEA-D2CSN model, the output from the concatenated features is processed to enhance the model's ability to capture temporal dependencies in the data. This approach operates by simultaneously processing the input sequence in two directions, allowing the model to learn from both past and future information. By leveraging this dual processing, the X-SGEA-D2CSN incorporates context from both directions, leading to a more comprehensive understanding of the input sequence. The combined output forms a richer feature representation, resulting in an output tensor that encapsulates the sequential dependencies and context for each input sequence is given in Eqs. (41)-(43) as follows.

$$h_t^{forward} = GRU\left(P_{concat}, h_{t-1}^{forward}\right) \tag{41}$$

$$h_t^{backward} = GRU\left(P_{concat}, h_{t+1}^{fbackward}\right) \tag{42}$$

The combined output is given by:

$$h_t = \left[h_t^{forward}, h_t^{backward}\right] \tag{43}$$

The Selective group-enhanced attention mechanism plays a crucial role in enhancing the performance of the model by combining multi-scale feature learning with attention-based refinement. After processing the output from the previous output, this mechanism applies a series of operations to further refine the extracted features. By leveraging multi-scale learning, the model captures intricate patterns and dependencies at various levels of granularity, effectively integrating both local and global contextual information. This helps guarantee that the model identifies the most significant characteristics from the data, resulting in more precise predictions. Additionally, the mechanism selectively highlights and prioritizes the key information, enabling the model to concentrate on essential elements of the input while minimizing the influence of less critical features. This selective focus not only improves the model's ability to identify essential patterns but also reduces dimensionality, making the decision-making process more efficient and precise. The final output of the SGEA mechanism provides a distilled and highly informative representation of the data, significantly boosting the overall effectiveness of the model in tasks such as crop classification in Eqs. (44) and (45).

$$SK = SKNET(h_t) \tag{44}$$

$$h_tSG = SG(SK) \tag{45}$$

where SG denotes the spatial group-wise model output. Thus, the combined output of the SGEA is represented by Eq. (46):

$$SGEA = h_tSG \tag{46}$$

After the output from the SGEA is generated, it is passed to the X-SGEA-D2CSN model for further temporal analysis. The X-SGEA-D2CSN is particularly adept at capturing sequential dependencies and long-term relationships in data, which is crucial for tasks involving temporal or time-series patterns. By processing the refined features from the selective group enhanced attention, the X-SGEA-D2CSN can maintain information across time steps, allowing it to understand how

features evolve and how past information influences current predictions in Eq. (47).[44]

$$h_t LSTM = LSTM(SGEA, h_{t-1}^{LSTM}) \quad (47)$$

After training the model, incorporating explainability techniques is vital for effectively interpreting its predictions. One such powerful tool is gradient-weighted class activation mapping (Grad-CAM), which offers valuable visual insights into the model's decision-making process. The initial step in Grad-CAM involves computing the gradients of the output class score with respect to the feature maps generated by the last convolutional layer.[44] These gradients reveal the sensitivity of the model's predictions to variations in the feature maps, pinpointing the specific areas of the input image that significantly impact the class decision. In addition, the important weights are computed by averaging these gradients across all spatial locations in the feature maps, allowing for a summary of their contribution to the prediction. Finally, the Grad-CAM heatmap is generated by summing the weighted feature maps, followed by applying a ReLU activation to retain only the positive influences. This heatmap visually indicates the regions of the input image that the model considered most important for making its prediction, thereby enhancing interpretability and trust in the model's outputs in Eq. (48).

$$GradCAM = Re\,L\,U(\sum_k \alpha_k A_k) \quad (48)$$

where α_k denotes the importance weight calculated for Grad-CAM, A_k are the feature maps of the last convolutional layer. Integrated Gradients is a technique for attributing a model's output to individual input features, enhancing interpretability. It starts with a baseline input, representing a neutral state, and calculates the difference between the actual input and this baseline for each feature. The approach combines the gradients of the model's output about the input along a trajectory from the baseline to the actual input, illustrating how the output varies as features are gradually added. The resulting attribution values highlight the importance of each feature, providing insights into which aspects of the input drive the model's decisions, thereby improving transparency in deep learning models in Eq. (49).

$$IG_i(x) = (x_i - x_{baseline}).\int_q^1 \frac{\partial F(x_{baseline}+\alpha(x-x_{baseline}))}{\partial x_i} d\alpha \quad (49)$$

where x_i denotes the i^{th} feature in the input, baseline input feature value is denoted as $x_{baseline,i}$, F denotes the function representing the model's output.

The X-SGEA-D2CSN is designed to understand long-term dependencies and identify sequential patterns within the data, enhancing the accuracy of its final predictions. It uses categorical cross-entropy loss as its loss function, which is especially well-suited for multi-class classification problems that involve 22 distinct crop types. This loss function evaluates the discrepancy between the actual class distribution and the probabilities predicted by the model. Further, this operation is carried out by computing a weighted sum of the log probabilities of the predicted classes, imposing a greater penalty for incorrect predictions.

Specifically, if the predicted probability for the correct class is low, the loss increases. The objective is to minimize this loss during training, enabling the model to refine its predictions over time. Hence, the model learns to adjust its parameters to improve accuracy in recommending suitable crops effectively.

5. Results and discussion

The research introduced an innovative approach to a DSM and crop recommendation system utilizing the X-SGEA-D2CSN model. The effectiveness of this model was assessed through comparisons with various leading state-of-the-art models.

5.1 Dataset description

In this research, the SoilGrid dataset is utilized,[33] which provides spatial predictions of soil properties like organic carbon content, pH, and soil texture at various depths. For our specific application, we take a single high-resolution image representing a region, such as Maharashtra, and split it into smaller pixel-wise segments. Each pixel corresponds to a specific geographical area (a few square kilometers), and within each pixel, the soil properties are predicted based on the underlying environmental covariates. These pixel-wise segments allow us to associate soil attribute details with distinct regions of the image, effectively creating a detailed, spatially-resolved dataset.

The total data is derived from the single Maharashtra image, which, when divided pixel-wise, results in a larger number of smaller "images" or regions. The dataset is then split into training and testing sets, where a portion (typically 70-80%) of the pixel-wise regions are used for training the model, and the remaining 20-30% are reserved for testing. This approach enables high-resolution soil predictions by training on the fine-grained spatial details of the soil attributes across the region, ensuring the model can generalize effectively to predict soil properties for other areas as well.

The crop recommendation dataset is constructed by augmenting existing rainfall, climate, and fertilizer datasets specific to India.[42] It encompasses several key data fields, including N, which represents the ratio of nitrogen content in the soil; P, denoting the ratio of phosphorus content; K, for the ratio of potassium content; temperature, measured in degrees Celsius; humidity, expressed as relative humidity in percentage; pH, indicating the soil's acidity or alkalinity; and rainfall, recorded in millimeters. This comprehensive compilation enables a thorough analysis of soil and environmental factors influencing agricultural practices.

5.2 Experimental setup

The experiment involving the DSM and crop recommendation system is carried out using a Python script on a Windows 10

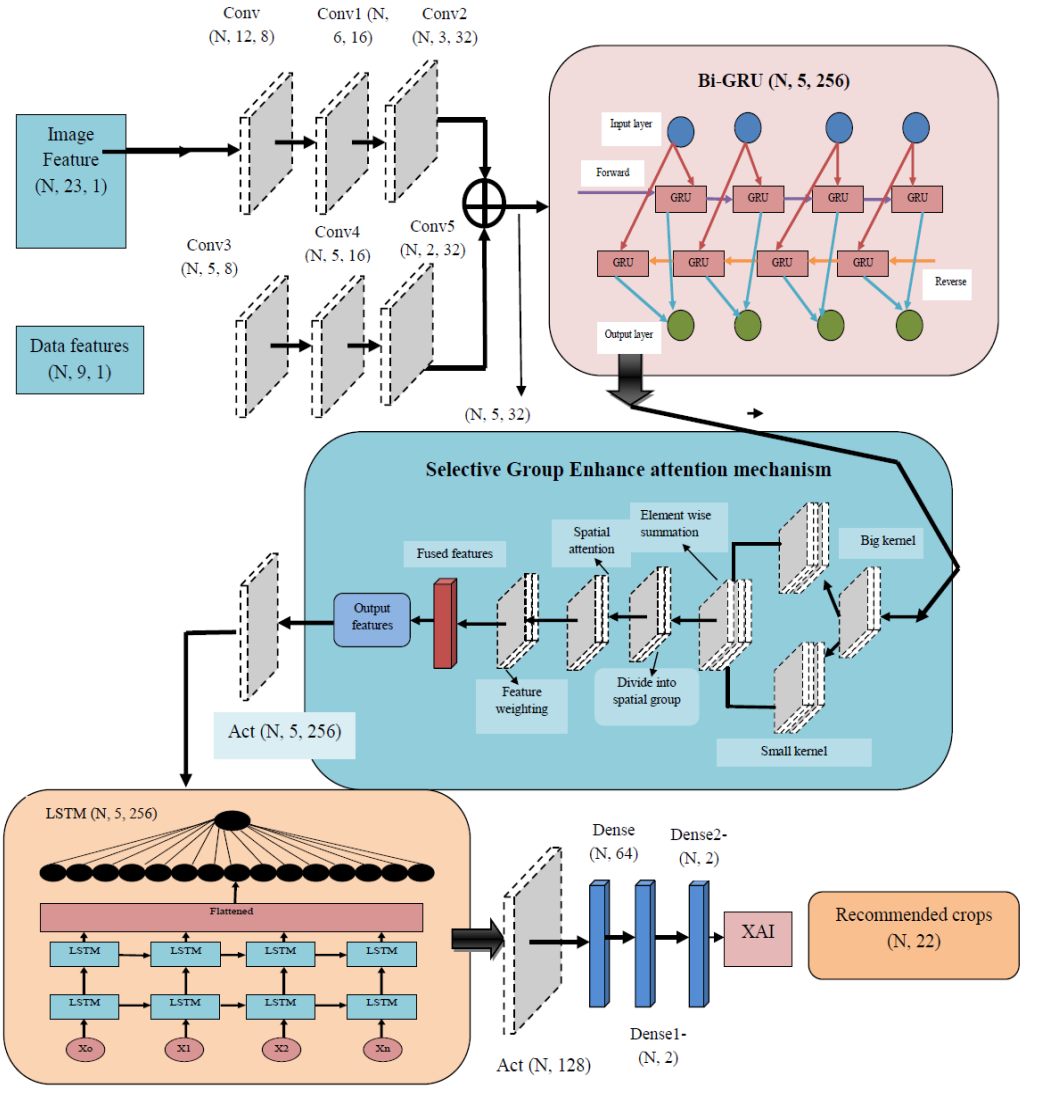


Fig. 4: Soil mapping and crop recommendation using the proposed X-SGEA-D2CSN model.

PC equipped with 8GB of RAM. This configuration facilitates effective data processing and analysis, permitting the model to handle the required computations and algorithms essential for the recommendation system. The Windows 10 operating system offers a user-friendly interface, while the 8GB of RAM provides adequate memory capacity for handling multiple tasks and datasets, ensuring smooth execution of the experiment.

5.3 Experimental results

The experimental results obtained using the X-SGEA-D2CSN model are illustrated in Figs. 4 and 5, which showcases a comprehensive analysis of various vegetation indices, each designated by a specific label from Figs. 5a to 5w. Specifically, a) ARV12 assesses the reflectance characteristics of vegetation, while b) ATSAVI offers insights into soil-adjusted vegetation metrics. c) CI_GREEN focuses on chlorophyll content to indicate plant health, and d) CTVI evaluates plant temperature and vigor. The index e) CVI illustrates crop vigor, whereas f) DVI represents vegetation density. Further, g) GDVI and h) GEMI enhance the understanding of vegetation health and environmental conditions. Notably, i) GNDVI utilizes green wavelengths for improved health assessment, while j) GOSAVI helps in gauging soil and vegetation interactions. Indices such as k) GRNDVI and l) GSAVI provide crucial insights into vegetation density and soil adjustments, respectively. The index m) IPVI quantifies vegetation quantity, and n) M_SAVI refines soil noise impacts in vegetation analysis. Additionally, o) MRVI and p) NDVI, a widely recognized metric, offer valuable insights into overall vegetation health. Indices q) NGRDI, r) NORM_G, s) NORM_NIR, t) NORM_R, u) RI, v) RVI, and w) TVI further enhance the model's analytical capabilities, showcasing a robust framework for extracting meaningful information. The experimental results of the explainable artificial intelligence (XAI) model are presented in Fig. 6 for different inputs. For input 1, Fig. 6a shows the Soil Grid Dataset, while Fig. 6b details the crop recommendation dataset. Fig. 6c illustrates

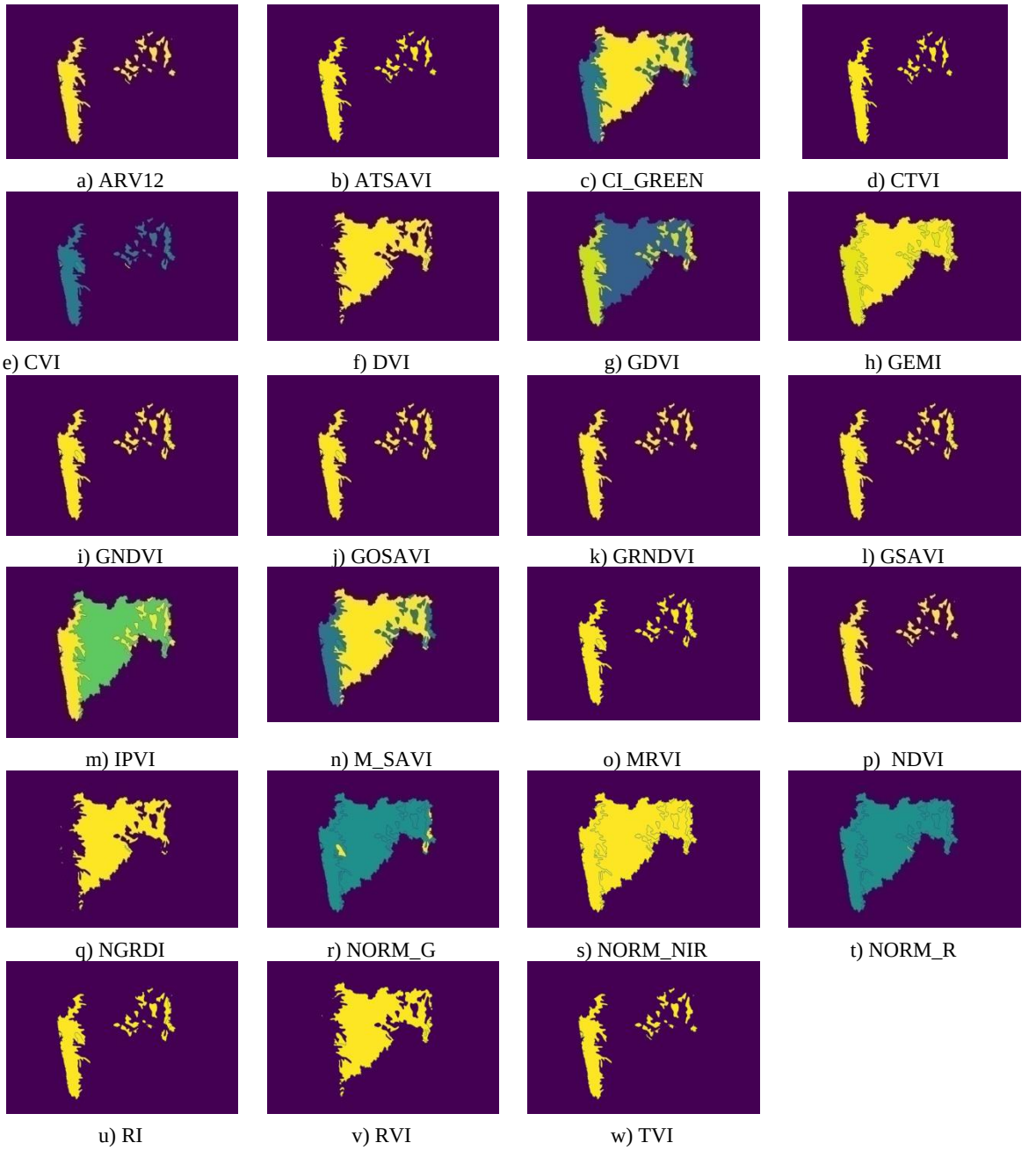
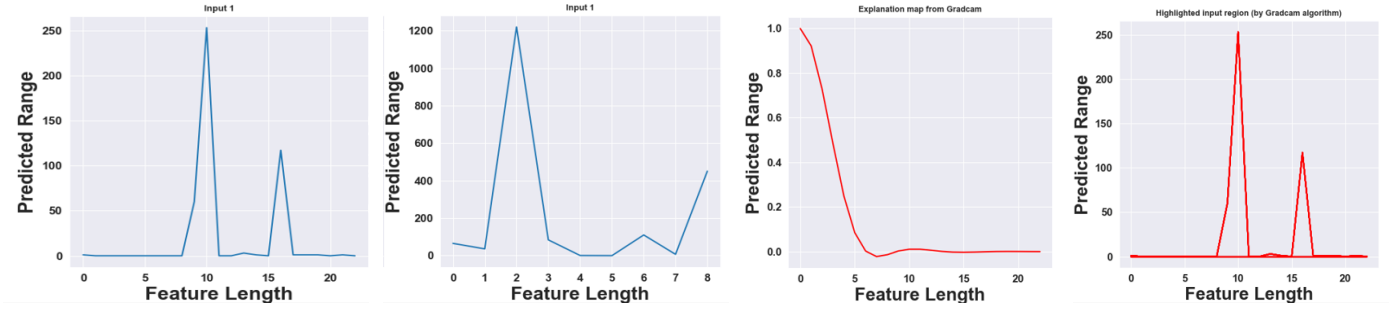


Fig. 5: Experimental results and analysis of various vegetation indices using the X-SGEA-D2CSN model.

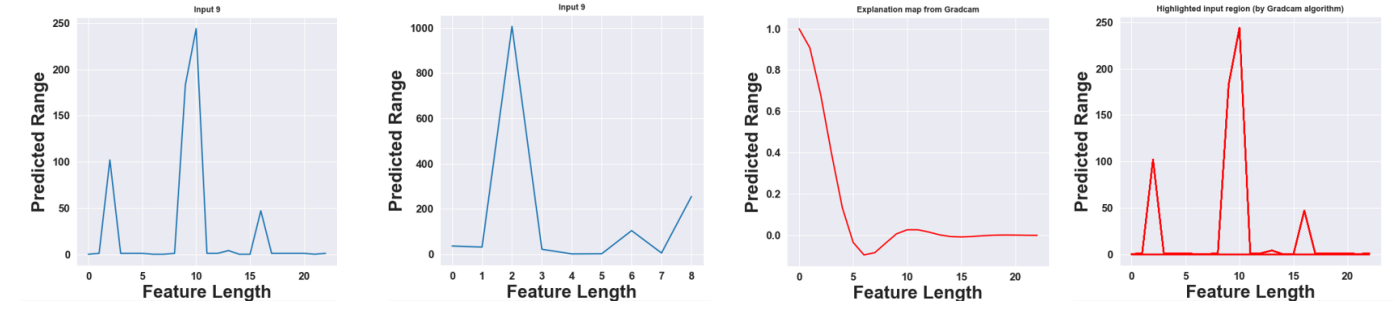
the feature representation output with the explainable MAP from Grad-CAM, which highlights important features influencing predictions. Fig. 6d identifies input regions crucial for decision-making, as determined by the Grad-CAM algorithm. Similarly, for input 9, the Figure follows the same structure, displaying the Soil Grid Dataset in Fig. 6e, the Crop Recommendation Dataset in Fig. 6f, and the feature representation output in Fig. 6g, with Fig. 6h showing the results of highlighting key regions identified by Grad-CAM.

5.4 Performance analysis based on TP

The results of the DSM and crop recommendation system developed using the X-SGEA-D2CSN model are illustrated in Fig. 7, which emphasizes critical performance metrics across various training epochs. In Fig. 7a, the accuracy percentages demonstrate a significant improvement at epochs 100, 200, 300, 400, and 500, with respective values of 88.78%, 88.92%, 90.54%, 90.98%, and a remarkable 95.01% achieved at a TP of 90. This continuous improvement highlights the model's capacity to sharpen its forecasting skills as time progresses. Fig. 7b further supports this finding by showcasing the precision metrics for the same epochs, reflecting a positive trend with values rising from 85.97% to 95.00% at a TP of 90, highlighting the model's reliability in correctly identifying crop types. Additionally, Fig. 7c provides insight into the recall percentages, which range from 86.06% to 95.85%, indicating a robust capacity for the model to detect relevant instances accurately.

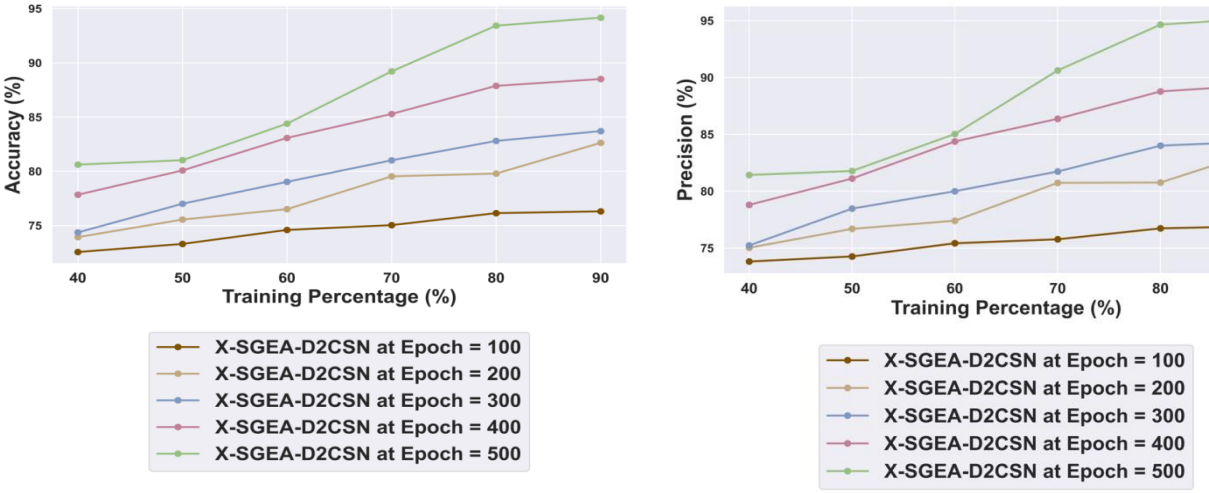


a) Input from Soil Grid Dataset b) Crop Recommendation Dataset c) Feature representation with explanation map d) Highlighted input region

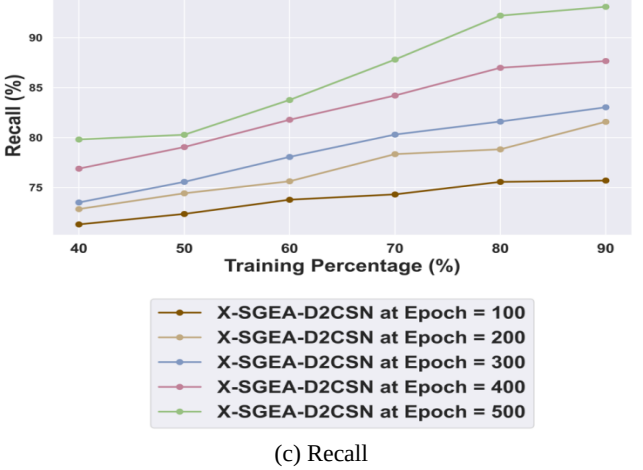


e) Input from Soil Grid Dataset f) Crop Recommendation Dataset g) Feature representation with explanation map h) Highlighted input region

Fig. 6: Experimental results obtained using the XAI.

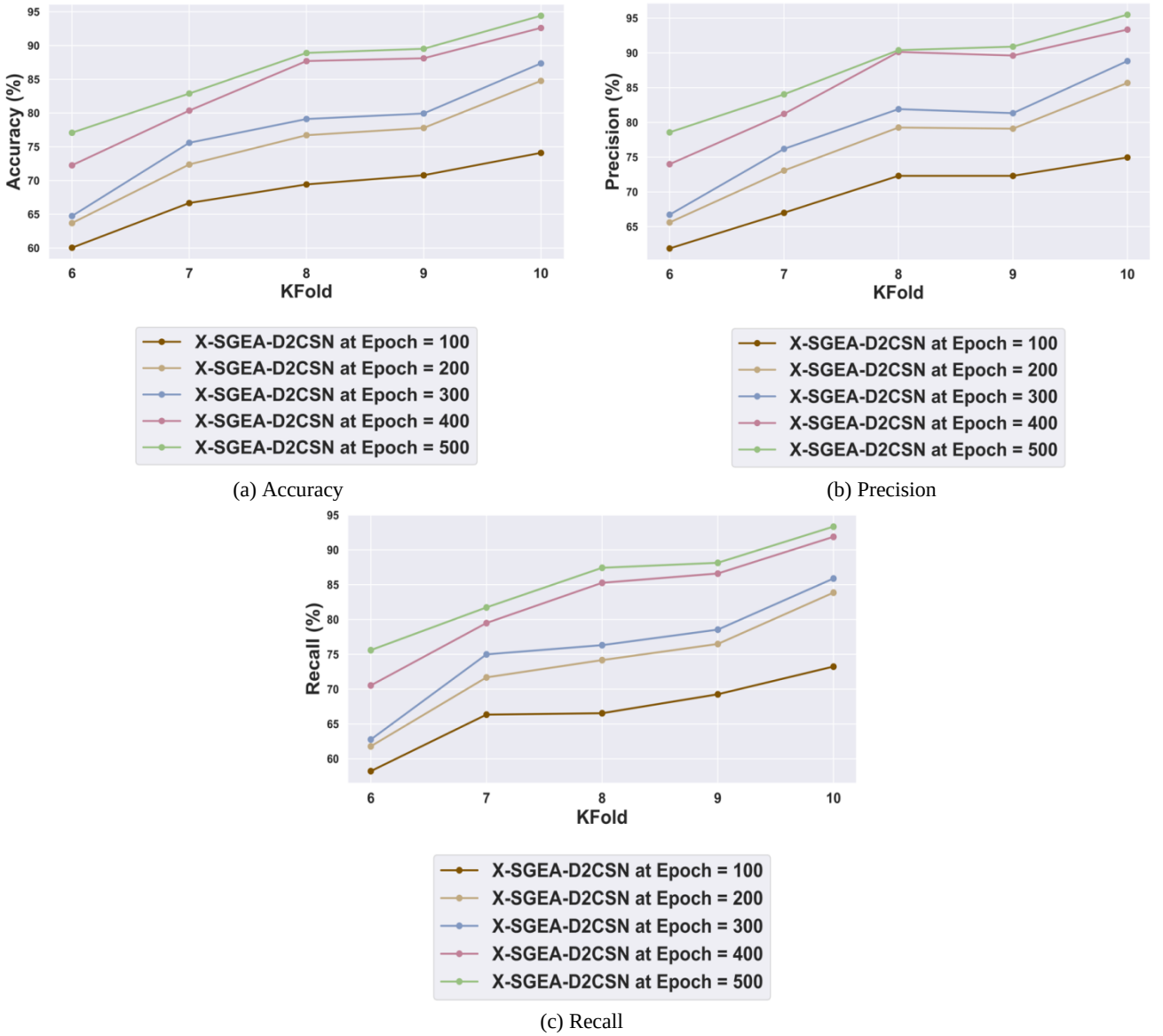


(a) Accuracy (b) Precision



(c) Recall

Fig. 7: Performance analysis based on TP a) accuracy, b) precision, and c) recall.



(a) Accuracy

(b) Precision

(c) Recall

Fig. 8: Performance analysis based on K-fold a) accuracy, b) precision, and c) recall.

5.5 Performance analysis based on K-fold

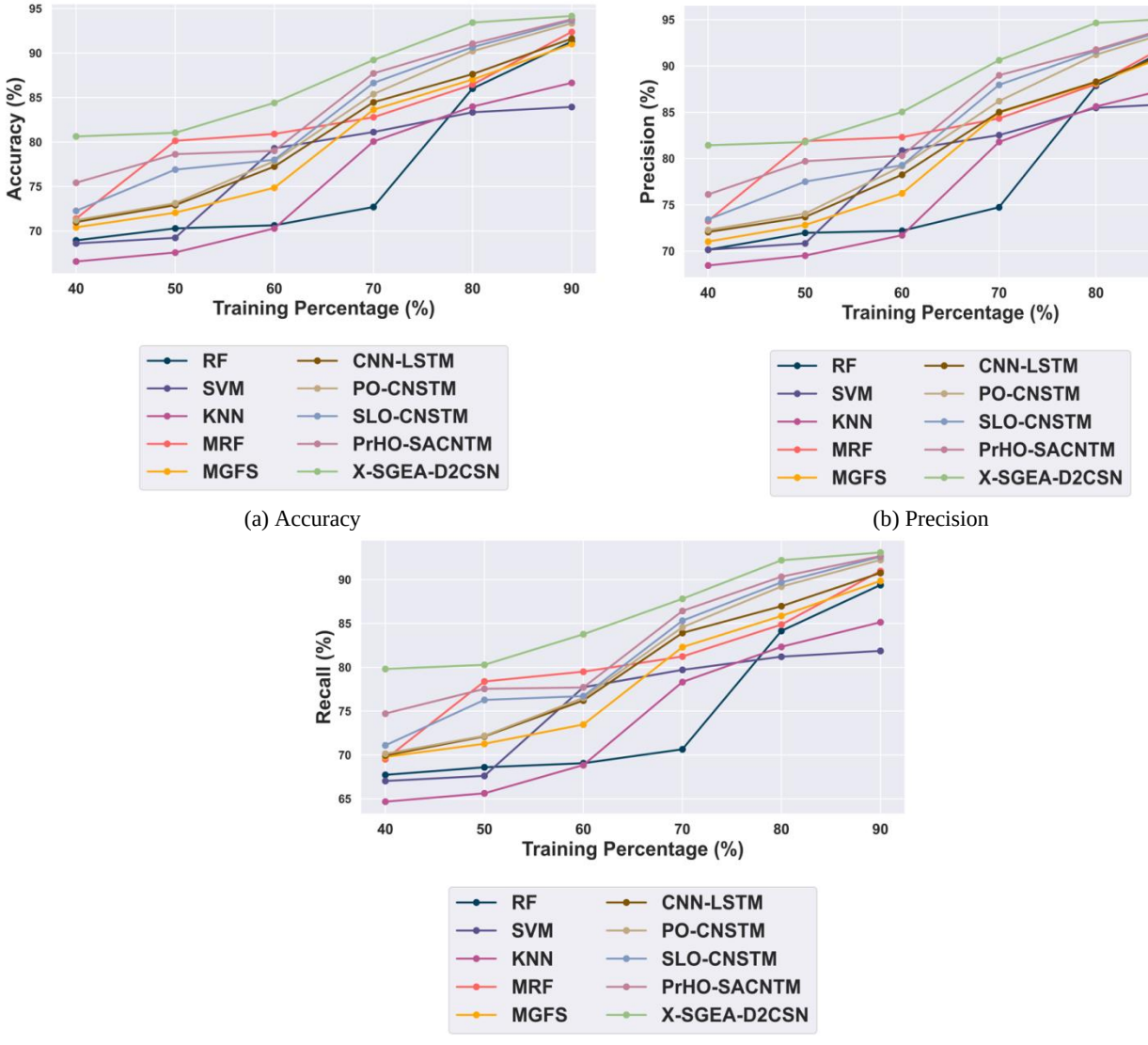
The outcomes of the DSM and crop recommendation system, which employs the X-SGEA-D2CSN model, are presented in Fig. 8, highlighting the performance metrics of the model throughout different training epochs. Fig. 8a presents accuracy percentages, revealing a significant improvement at a k-fold of 10, with accuracy increasing from 88.78% at epoch 100 to an impressive 95.01% by epoch 500. This upward trend underscores the model's enhanced ability to make accurate predictions as training progresses. Similarly, Fig. 8b highlights precision metrics, which also show substantial growth, starting at 85.97% and reaching 95.00% by epoch 500, further demonstrating the model's effectiveness in identifying relevant crop recommendations. Finally, Fig. 8c illustrates the recall percentages, which exhibit consistent improvement, rising from 86.06% to 95.85% over the epochs. Together, these metrics reflect the robust performance of the X-SGEA-D2C model and its potential for impactful application in precision agriculture.

5.6 Comparative analysis

To showcase the capabilities of the X-SGEA-D2CSN model, a comparative analysis is performed against various established methodologies. This study incorporates several techniques, including random forest (RF),[4] support vector machine (SVM), [3] k-nearest neighbors (KNN),[24] Markov random field (MRF),[25] multi-granularity feature selection (MGFS),[26] CNN-LSTM,[27] PO-CNSTM,[45] SLO-CNSTM,[46] and PrHO-SACNTM.[47] By evaluating the performance of these models alongside X-SGEA-D2CSN, the analysis aims to highlight the strengths and advancements of the proposed approach in addressing the specific challenges of the research.

5.6.1 Comparative analysis based on TP

The X-SGEA-D2CSN model exhibited significant superiority over the PrHO-SACNTM model within the DSM and crop recommendation system, as demonstrated in Fig. 9. With an impressive accuracy of 95.01%, the X-SGEA-D2CSN model improved by 1.94%, highlighting its superior predictive



(a) Accuracy (b) Precision

(c) Recall

Fig. 9: Comparative analysis based on TP a) accuracy, b) precision, and c) recall.

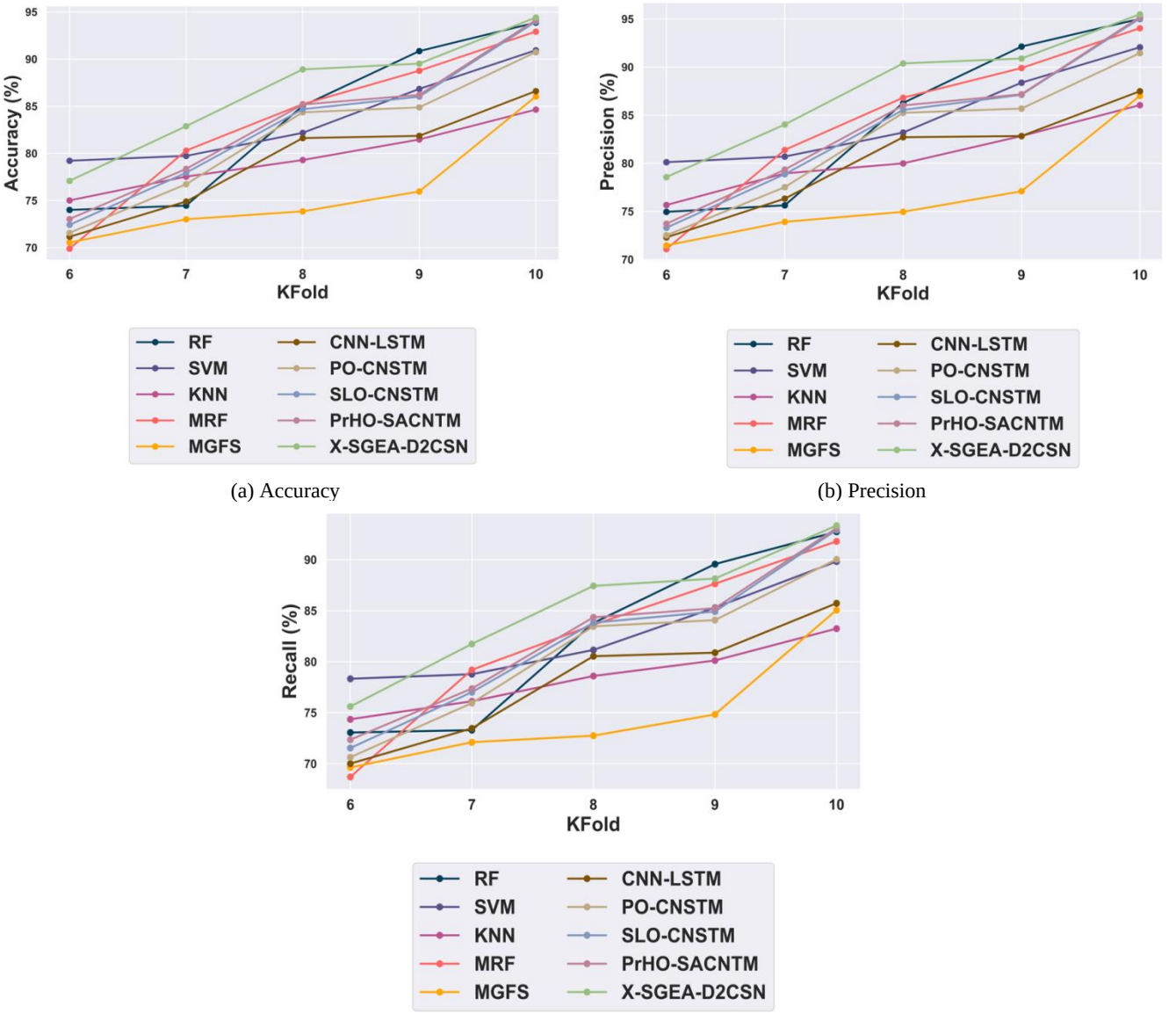
capabilities, as shown in Fig. 9a. In terms of precision, the X-SGEA-D2CSN model surpassed the PrHO-SACNTM by 0.41%, achieving a precision of 95.00% at a TP of 90, as illustrated in Fig. 9b. Additionally, the model showcased outstanding recall performance, with the EK-HDED-CN model exceeding the PrHO-SACNTM by 0.36%, reaching a recall of 95.85% at the same TP, as seen in Fig. 9c. Together, these metrics highlight the X-SGEA-D2CSN model's overall enhanced effectiveness in DSM and crop recommendation tasks, demonstrating its potential for delivering higher accuracy, precision, and recall, positioning it as a more reliable and efficient tool for precision agriculture.

5.6.2 Comparative analysis based on K-fold

The X-SGEA-D2CSN model showcased exceptional performance in the DSM and crop recommendation system, significantly outperforming the PrHO-SACNTM model in various key metrics. As depicted in Fig. 10a, the X-SGEA-D2CSN achieved a remarkable accuracy of 95.01%, representing a notable 1.94% improvement over the PrHO-SACNTM model. This enhancement highlights the model's superior capability to deliver accurate predictions. Moreover, the X-SGEA-D2CSN model outperformed the PrHO-SACNTM in terms of precision, exceeding it by 0.41% and attaining an impressive precision score of 95.00% at a k-fold of 10, as depicted in Fig. 10b. In addition, the EK-HDED-CN model showed a greater recall than the PrHO-SACNTM model, surpassing it by 0.36% and achieving a recall of 95.85% at the same k-fold of 10, as shown in Fig. 10c.

5.7 Comparative analysis based on receiver operating characteristics

The trade-off between the false positive rate (FPR) and true positive rate (TPR) is effectively illustrated by the receiver operating characteristic (ROC) curve, which serves as a vital tool for evaluating model performance across a range of decision thresholds. This curve enables researchers and practitioners to visualize how changes in the threshold affect



(a) Accuracy

(b) Precision

(c) Recall

Fig. 10: Comparative analysis based on K-fold a) accuracy, b) precision, and c) recall.

the model's ability to correctly identify positive instances while minimizing incorrect classifications. The area under the curve (AUC) provides a clear measure of the model's overall effectiveness, allowing for straightforward comparisons between different models. As shown in Fig. 11, the ROC curve demonstrates a positive correlation between TPR and FPR, where TPR represents the model's sensitivity in accurately identifying true positives. A higher TPR indicates better performance in detecting actual positive cases, while a lower FPR signifies fewer false alarms. This comprehensive view provided by the ROC curve and AUC is instrumental in selecting the most appropriate model for specific applications, balancing sensitivity and specificity according to the requirements of the task at hand.Notably, the X-SGEA-D2CSN model achieves an impressive TPR of 0.95, significantly surpassing competing models such as RF, SVM, KNN, MRF, MGFS, CNN-LSTM, PO-CNSTM, SLO-CNSTM, and PrHO-SACNTM, whose TPRs ranged from 0.89 to 0.94. This marked improvement in sensitivity underscores the X-SGEA-D2CSN model's superior performance in the DSM and crop recommendation system, demonstrating its enhanced capacity for accurate classification and its robustness across varying thresholds, making it an ideal choice for precision agriculture applications.

5.8 Comparison of various models

The comparative analysis reveals multiple challenges encountered by current models including RF, SVM, KNN, MRF, MGFS, CNN-LSTM, PO-CNSTM, SLO-CNSTM, and PrHO-SACNTM within the framework of DSM and crop recommendation systems.These challenges include limited accuracy in predicting optimal crop choices due to inadequate handling of complex data relationships, susceptibility to

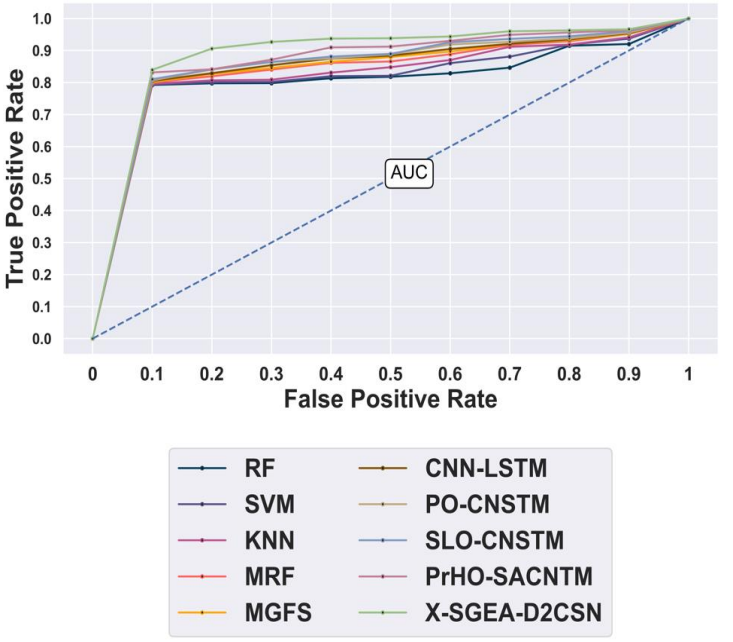


Fig. 11: Analysis based on ROC.

Table 2: Comparative discussion table during TP 90 and k-fold 10.

	90% TP			10 k-fold		
Models	Accuracy	Precision	Recall	Accuracy	Precision	Recall
RF[4]	91.3%	93.21%	89.39%	93.87%	95.01%	92.73%
SVM[3]	83.94%	86.01%	81.87%	90.94%	92.06%	89.82%
KNN[24]	86.65%	88.16%	85.14%	84.64%	86.05%	83.24%
MRF[25]	92.37%	93.77%	90.98%	92.92%	94.04%	91.79%
MGFS[26]	91.01%	92.17%	89.85%	86.03%	87.01%	85.06%
CNN-LSTM[27]	91.62%	92.49%	90.76%	86.60%	87.49%	85.72%
PO-CNSTM[45]	93.34%	94.44%	92.24%	90.74%	91.45%	90.03%
SLO-CNSTM[46]	93.67%	94.73%	92.6%	93.98%	95.05%	92.90%
PrHO-SACNTM[47]	93.8%	94.91%	92.69%	94.15%	95.20%	93.09%
X-SGEA-D2CSN	94.17%	95.24%	93.1%	94.41%	95.49%	93.34%

overfitting, and difficulties in generalizing across diverse agricultural scenarios. Many of these models also struggle with high computational costs and inefficient feature selection, which can hinder performance in real-world applications. The proposed X-SGEA-D2CSN model successfully tackles these issues by incorporating sophisticated optimization methods that boost both feature extraction and selection, ultimately enhancing the model's reliability and precision. By utilizing its distinctive architecture, the X-SGEA-D2CSN model promotes deeper insight into the complex interactions found in agricultural data, resulting in more precise predictions and recommendations. Additionally, its ability to adapt to various agricultural environments and its lower computational overhead position it as a more efficient and reliable solution compared to its predecessors. Table 2 depicts the comparative discussion of the X-SGEA-D2CSN model.

6. Conclusion

In conclusion, the X-SGEA-D2CSN model represents a transformative advancement in DSM and crop recommendation systems, combining sophisticated methodologies to enhance predictive accuracy and decision-making in agriculture. By effectively capturing sequential dependencies and long-term relationships in data, the model provides a comprehensive understanding of crop dynamics and environmental influences. Its innovative architecture enables the integration of diverse data sources and ensures efficient processing and filtering of critical information. Hence, the X-SGEA-D2CSN model offers actionable insights for farmers and agricultural stakeholders, ultimately promoting sustainable practices and optimizing yield potential. The proposed X-SGEA-D2CSN model achieves impressive performance metrics, attaining an accuracy of 94.17%, precision of 95.24%, and recall of 93.1% during 90% training. Future work for the X-SGEA-D2CSN model includes enhancing adaptability by integrating diverse data sources such as climatic variables and soil health indicators, optimizing performance through real-time data integration and efficient processing algorithms, and exploring scalability across various geographic regions to support decision-making

in dynamic farming environments.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

References

[1] A. Tiruye, P. Ditthakit, Q. B. Pham, W. Wipulanusat, U. Weesakul, S. Thongkao, N. L. Kushwaha, Satellite and model-based estimation of crop water requirement of major irrigated crops in the koga irrigation scheme, Ethiopia, *Engineered Science*, 2024, **30**, 1155doi: 10.30919/es1155

[2] O. D. Adeniyi, A. Brenning, A. Bernini, S. Brenna, M. Maerker, Digital mapping of soil properties using ensemble machine learning approaches in an agricultural lowland area of Lombardy, Italy, *Land*, 2023, **12**, 494, doi: 10.3390/land12020494.

[3] V. Estévez, A. Beucher, S. Mattbäck, A. Boman, J. Auri, K.-M. Björk, Ö. Peter, Machine learning techniques for acid sulfate soil mapping in southeastern Finland, *Geoderma*, 2022, **406**, 115446, doi: 10.1016/j.geoderma.2021.115446.

[4] H. Fathizad, M. Ali Hakimzadeh Ardakani, B. Heung, H. Sodaiezadeh, A. Rahmani, A. Fathabadi, T. Scholten, R. Taghizadeh-Mehrjardi, Spatio-temporal dynamic of soil quality in the central Iranian desert modeled with machine learning and digital soil assessment techniques, *Ecological Indicators*, 2020, **118**, 106736, doi: 10.1016/j.ecolind.2020.106736.

[5] S. Fathololoumi, A. R. Vaezi, S. K. Alavipanah, A. Ghorbani, D. Saurette, A. Biswas, Effect of multi-temporal satellite images on soil moisture prediction using a digital soil mapping approach, *Geoderma*, 2021, **385**, 114901, doi: 10.1016/j.geoderma.2020.114901.

[6] S. Dharumarajan, R. Vasundhara, A. Suputhra, M. Lalitha, R. Hegde, Prediction of soil depth in Karnataka using digital soil mapping approach, *Journal of the Indian Society of Remote Sensing*, 2020, **48**, 1593-1600, doi: 10.1007/s12524-020-01184-7.

[7] K. Nabiollahi, R. Taghizadeh-Mehrjardi, A. Shahabi, B. Heung, A. Amirian-Chakan, M. Davari, T. Scholten, Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests, *Geoderma*, 2021, **385**, 114858, doi: 10.1016/j.geoderma.2020.114858.

[8] K. Adhikari, A. E. Hartemink, Linking soils to ecosystem services: a global review, *Geoderma*, 2016, **262**, 101-111, doi: 10.1016/j.geoderma.2015.08.009.

[9] A. X. Zhu, B. Hudson, J. Burt, K. Lubich, D. Simonson, Soil mapping using GIS, expert knowledge, and fuzzy logic, *Soil Science Society of America Journal*, 2001, **65**, 1463-1472, doi: 10.2136/sssaj2001.6551463x.

[10] A. B. McBratney, M. L. Mendonça Santos, B. Minasny, On digital soil mapping, *Geoderma*, 2003, **117**, 3-52, doi: 10.1016/S0016-7061(03)00223-4.

[11] B. Minasny, A. B. McBratney, B. P. Malone, I. Wheeler, Digital mapping of soil carbon, *Advances in Agronomy*, 2013, **118**, 1-47, doi: 10.1016/b978-0-12-405942-9.00001-3.

[12] I. V. Florinsky, R. G. Eilers, G. R. Manning, L. G. Fuller, Prediction of soil properties by digital terrain modelling, *Environmental Modelling & Software*, 2002, **17**, 295-311, doi: 10.1016/s1364-8152(01)00067-6.

[13] R. Grimm, T. Behrens, M. Märker, H. Elsenbeer, Soil organic carbon concentrations and stocks on Barro Colorado Island: Digital soil mapping using random forests analysis, *Geoderma*, 2008, **146**, 102-113, doi: 10.1016/j.geoderma.2008.05.008.

[14] J. Seibert, J. Stendahl, R. Sørensen, Topographical influences on soil properties in boreal forests, *Geoderma*, 2007, **141**, 139-148, doi: 10.1016/j.geoderma.2007.05.013.

[15] C. Tu, T. He, X. Lu, Y. Luo, P. Smith, Extent to which pH and topographic factors control soil organic carbon level in dry farming cropland soils of the mountainous region of Southwest China, *CATENA*, 2018, **163**, 204-209, doi: 10.1016/j.catena.2017.12.028.

[16] X. Song, F. Liu, G. Zhang, D. Li, Y. Zhao, J. Yang, Mapping soil organic carbon using local terrain attributes: a comparison of different polynomial models, *Pedosphere*, 2017, **27**, 681-693, doi: 10.1016/S1002-0160(17)60445-4.

[17] A.-X. Zhu, F. Liu, B. Li, T. Pei, C. Qin, G. Liu, Y. Wang, Y. Chen, X. Ma, F. Qi, C. Zhou, Differentiation of soil conditions over low relief areas using feedback dynamic patterns, *Soil Science Society of America Journal*, 2010, **74**, 861-869, doi: 10.2136/sssaj2008.0411.

[18] M. R. Pahlavan-Rad, A. Akbarimoghaddam, Spatial variability of soil texture fractions and pH in a flood plain (case study from Eastern Iran), *CATENA*, 2018, **160**, 275-281, doi: 10.1016/j.catena.2017.10.002.

[19] B. Heung, H. C. Ho, J. Zhang, A. Knudby, C. E. Bulmer, M. G. Schmidt, An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping, *Geoderma*, 2016, **265**, 62-77, doi: 10.1016/j.geoderma.2015.11.014.

[20] E. Uğurlu, Renewale energy sources and climate change mitigation, Energy Policy Advancement. Springer, Cham, 2022, ISBN: 978-3-030-84992-4.

[21] K. Adhikari, A. E. Hartemink, Digital mapping of topsoil carbon content and changes in the driftless area of Wisconsin, USA, *Soil Science Society of America Journal*, 2015, **79**, 155-164, doi: 10.2136/sssaj2014.09.0392.

[22] R. Lal, Soil carbon sequestration to mitigate climate change, *Geoderma*, 2004, **123**, 1-22, doi: 10.1016/j.geoderma.2004.01.032.

[23] H. Keskin, S. Grunwald, W. G. Harris, Digital mapping of soil carbon fractions with machine learning, *Geoderma*, 2019, **339**, 40-58, doi: 10.1016/j.geoderma.2018.12.037.

[24] M. R. Sundari, G. S. R. Krishna, V. S. Naveen, G. Bharathi, Crop recommendation system using k-nearest neighbors algorithm, Proceedings of 6th International Conference on Recent Trends in Computing: ICRTC 2020, Springer Singapore, 2021, 581-589, doi: 10.1007/978-981-33-4501-0_54.

[25] Y. Li, G. Cai, K. Tan, R. Zeng, X. Chen, X. Wang, Emergy-based efficiency and sustainability assessments of diversified multi-cropping systems in South China, *Journal of Cleaner Production*, 2013, **414**, 137660, doi: 10.1016/j.jclepro.2023.137660.
[26] X. Zhang, S. Chen, J. Xue, N. Wang, Y. Xiao, Q. Chen, Y. Hong, Y. Zhou, H. Teng, B. Hu, Z. Zhuo, W. Ji, Y. Huang, Y. Gou, A. C. Richer-de-Forges, D. Arrouays, Z. Shi, Improving model parsimony and accuracy by modified greedy feature selection in digital soil mapping, *Geoderma*, 2023, **432**, 116383, doi: 10.1016/j.geoderma.2023.116383.
[27] F. Liao, X. Feng, Z. Li, D. Wang, C. Xu, G. Chu, H. Ma, Q. Yao, S. Chen, A hybrid CNN-LSTM model for diagnosing rice nutrient levels at the rice panicle initiation stage, *Journal of Integrative Agriculture*, 2024, **23**, 711-723, doi: 10.1016/j.jia.2023.05.032.
[28] B. Kasraei, B. Heung, D. D. Saurette, M. G. Schmidt, C. E. Bulmer, W. Bethel, Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning, *Environmental Modelling & Software*, 2021, **144**, 105139, doi: 10.1016/j.envsoft.2021.105139.
[29] S. van der Westhuizen, G. B. M. Heuvelink, D. P. Hofmeyr, L. Poggio, Measurement error-filtered machine learning in digital soil mapping, *Spatial Statistics*, 2022, **47**, 100572, doi: 10.1016/j.spasta.2021.100572.
[30] I. K. Nti, A. Zaman, O. Nyarko-Boateng, A. F. Adekoya, F. Keyeremeh, A predictive analytics model for crop suitability and productivity with tree-based ensemble learning, *Decision Analytics Journal*, 2023, **8**, 100311, doi: 10.1016/j.dajour.2023.100311.
[31] S. Dasgupta, S. Debnath, A. Das, A. Biswas, D. C. Weindorf, B. Li, A. Kumar Shukla, S. Das, S. Saha, S. Chakraborty, Developing regional soil micronutrient management strategies through ensemble learning based digital soil mapping, *Geoderma*, 2023, **433**, 116457, doi: 10.1016/j.geoderma.2023.116457.
[32] A. Subeesh, N. Chauhan, Biotic stress management in soil-less agriculture systems: a deep learning approach for identification of leaf miner pest infestation, *Procedia Computer Science*, 2024, **233**, 371-380, doi: 10.1016/j.procs.2024.03.227.
[33] W. Shangguan, Y. Dai, Q. Duan, B. Liu, H. Yuan, A global soil data set for earth system modeling, *Journal of Advances in Modeling Earth Systems*, 2014, **6**, 249-263, doi: 10.1002/2013MS000293.
[34] B. Song, K. Park, Detection of aquatic plants using multispectral UAV imagery and vegetation index, *Remote Sensing*, 2020, **12**, 387, doi: 10.3390/rs12030387.
[35] S. S. Somvanshi, M. Kumari, Comparative analysis of different vegetation indices with respect to atmospheric particulate pollution using Sentinel data, *Applied Computing and Geosciences*, 2020, **7**, 100032, doi: 10.1016/j.acags.2020.100032.
[36] C. Qiu, G. Liao, H. Tang, F. Liu, X. Liao, R. Zhang, Z. Zhao, Derivative parameters of hyperspectral NDVI and its application in the inversion of rapeseed leaf area index, *Applied Sciences*, 2018, **8**, 1300, doi: 10.3390/app8081300.
[37] S. Gao, R. Zhong, K. Yan, X. Ma, X. Chen, J. Pu, S. Gao, J. Qi, G. Yin, R. B. Myneni, Evaluating the saturation effect of vegetation indices in forests using 3D radiative transfer simulations and satellite observations, *Remote Sensing of Environment*, 2023, **295**, 113665, doi: 10.1016/j.rse.2023.113665.
[38] L. P. Venancio, E. C. Mantovani, C. H. do Amaral, C. M. U. Neale, I. Z. Gonçalves, R. Filgueiras, F. C. Eugenio, Potential of using spectral vegetation indices for corn green biomass estimation based on their relationship with the photosynthetic vegetation sub-pixel fraction, *Agricultural Water Management*, 2020, **236**, 106155, doi: 10.1016/j.agwat.2020.106155.
[39] P. Lemenkova, Hyperspectral vegetation indices calculated by QGIS using Landsat TM image: a case study of northern Iceland, *Advanced Research in Life Sciences*, 2020, **4**, 70-78, doi: 10.2478/arls-2020-0021.
[40] L. M. dos Santos, G. A. E. S. Ferraz, D. B. Marin, M. A. de Figueiredo Carvalho, J. E. L. Dias, A. de Oliveira Alecrim, M. de Lourdes Oliveira e Silva, Vegetation indices applied to suborbital multispectral images of healthy coffee and coffee infested with coffee leaf miner, *AgriEngineering*, 2022, **4**, 311-319, doi: 10.3390/agriengineering4010021.
[41] K. Yan, S. Gao, H. Chi, J. Qi, W. Song, Y. Tong, X. Mu, G. Yan, Evaluation of the vegetation-index-based dimidiate pixel model for fractional vegetation cover estimation, *IEEE Transactions on Geoscience and Remote Sensing*, 2021, **60**, 4400514, doi: 10.1109/TGRS.2020.3048493.
[42] M. K. Senapaty, A. Ray, N. Padhy, A decision support system for crop recommendation using machine learning classification algorithms, *Agriculture*, 2024, **14**, 1256. doi: 10.3390/agriculture14081256.
[43] J. Chang, D. E. Clay, C. G. Carlson, D. Malo, S. A. Clay, J. Lee, M. Ellsbury, Precision farming protocols: part 1. grid distance and soil nutrient impact on the reproducibility of spatial variability measurements, *Precision Agriculture*, 1999, **1**, 277-289, doi: 10.1023/A:1009921024451.
[44] E. E. Hussein, B. Zerouali, N. Bailek, A. Derdour, S. S. M. Ghoneim, C. A. G. Santos, M. A. Hashim, Harnessing explainable AI for sustainable agriculture: SHAP-based feature selection in multi-model evaluation of irrigation water quality indices, *Water*, 2025, **17**, 59, doi: 10.3390/w17010059.
[45] J. Liu, Y. Wu, L. Zhou, A. Zhang, S. Wang, Y. Liu, D. Yang, S. Wang, Influence of flowering on the anatomical structure, chemical components and carbohydrate metabolism of Bambusa tuldoides culms at different ages, *Frontiers in Plant Science*, 2023, **14**, 1260302, doi: 10.3389/fpls.2023.1260302.
[46] N. Mahesh, S. Balakumar, S. Shyamalagowri, J. Manjunathan, M. K. S. Pavithra, P. S. Babu, M. Kamaraj, M. Govarthanan, Carbon-based adsorbents as proficient tools for the removal of heavy metals from aqueous solution: A state of art-review emphasizing recent progress and prospects, *Environmental Research*, 2022, **213**, 113723, doi: 10.1016/j.envres.2022.113723.
[47] M. R. Granados, R. B. Thompson, M. D. Fernández, C. Martínez-Gaitán, M. Gallardo, Prescriptive-corrective nitrogen and irrigation management of fertigated and drip-irrigated vegetable crops using modeling and monitoring approaches.

Agricultural Water Management, 2013, **119**, 121-134, doi: 10.1016/j.agwat.2012.12.014.