Deformable Sub-Pixel Convolutional Layer with Convolutional Neural Network for Land Use and Land Cover Classification

Kishore Raju Kalidindi [1,*], Murty Chakka S. V. V. S. N [2], A Srinivasa Reddy [3], Sridevi Gadde [4], and Rambabu Pemula [5]

[1] Department of Information Technology, Sagi Rama Krishnam Raju Engineering College, Bhimavaram, India
[2] Computer Science and Engineering, Aditya University, Surampalem, India
[3] Department of Computer Science and Engineering (Data Science), CVR College of Engineering, Hyderabad, India
[4] Computer Science and Engineering, Raghu Engineering College, Visakhapatnam, India
[5] Department of Information Technology, Vidya Jyothi Institute of Technology, Hyderabad, India
Email: kkishoreraju79@gmail.com (K.R.K.); chsatyamurty@gmail.com (M.C.S.V.V.S.N.); srinivas.asr@gmail.com (A.S.R.); sridevi.gadde@raghuenggcollege.in (S.G.); rpemula@gmail.com (R.P.)
*Corresponding author

***Abstract*—Image classification using Remote Sensing (RS) or satellite images provides data about Land Use and Land Cover (LULC), which is also used in several applications such as environmental monitoring and urban planning. In recent times, Deep Learning (DL)-based approaches have been used for LULC classification and have achieved high effectiveness. However, it remains challenging because of class similarity and mixed-pixel issues, which lead to misclassification and reduced classification performance. In this manuscript, the Deformable Sub-Pixel Convolutional Layer with Convolutional Neural Network (DSPCL with CNN) is developed to effectively classify LULC by mitigating class similarity and mixed-pixel issues. The deformable convolutional layer is incorporated instead of the traditional convolutional layer, which adaptively changes the filters and captures deep features. Then, the sub-pixel convolutional layer is incorporated into the output layer, which separates the pixels into multiple pixels and classifies the classes more effectively. The developed DSPCL with CNN method achieved 98.62% accuracy on the EuroSAT dataset and 99.01% accuracy on the NWPU-RESISC45 dataset compared to conventional techniques.**

***Keywords*—convolutional neural network, deformable convolutional layer, land use and land cover, mixed pixels, sub-pixel convolutional layer**

I. Introduction

Land Use and Land Cover (LULC) classification using Remote Sensing (RS) images is an essential phase in several applications such as forest management, land-use planning, and agricultural practices [1–3]. Land use represents the purpose of land cover and employs areas on Earth that include urban structures, vegetation, water, and others. Land use varies within areas with the same cover type, making the evaluation of LULC crucial for effectively monitoring, planning, and managing natural resource usage [4, 5]. Classification of LULC affects atmospheric, water, and soil erosion, and is indirectly relevant to global landscape problems [6, 7]. RS and its processes provide current and extensive information on surface situations [8]. Generally, existing works focus on extracting discriminative features for LULC classification using RS data [9]. Traditional methods focused on hand-crafted attributes such as color and texture features. Mid-level approaches have been employed to develop representations supporting high-level statistical methods [10].

In the domain of RS, scene classification images are essential and complex issues in real-time applications such as urban growth from High-Spatial-Resolution (HSR) data, geospatial object identification, natural hazard identification, and environment monitoring [11–13]. In many real-time applications of RS, LULC classification is an essential step. The Deep Learning (DL) algorithm plays a major role in capturing high-level attributes and has become an important paradigm in Computer Vision (CV) [14]. In DL methods, the Convolutional Neural Network (CNN) is an efficient and data-driven method designed to identify intricate architectures and extract significant data from HSR RS. Furthermore, a sequence of DL scene classification methods has been introduced in recent times [15]. However, DL-based methods struggle to achieve high classification performance due to class similarity and mixed-pixel issues. To overcome these limitations, deep features are captured, and sub-pixel classification is employed. By capturing deep features, class similarity can be identified, and sub-pixel classification differentiates mixed pixels, thereby improving classification performance. Unlike traditional models that extract features using fixed receptive fields,

Manuscript received August 15, 2025; revised September 23, 2025; accepted November 17, 2025; published April 16, 2026.

the deformable convolutional layer dynamically adjusts spatial sampling positions, enabling adaptive feature learning for spectrally similar classes. The sub-pixel convolutional layer performs pixel-level reconstruction to resolve mixed pixels and refine class boundaries. This integration of adaptive feature extraction and spatial reconstruction in traditional models has not been developed for LULC classification. The model improves discrimination and spatial consistency, achieveing superior performance compared to traditional CNN models. The essential contributions of this manuscript are summarized below:

- The Deformable Sub-Pixel Convolutional Layer with Convolutional Neural Network (DSPCL with CNN) method is developed to effectively classify LULC classes.
- The deformable convolutional layer is included instead of the traditional convolutional layer, using adaptive filter forms to capture deep features that help differentiate various classes.
- The sub-pixel convolutional layer is incorporated in the reconstruction phase, performing pixel-level upscaling and rearrangement. This enables the separation of mixed pixels at finer spatial resolutions, minimizing misclassification in boundary regions.
- The extracted global feature maps are reconstructed by sub-pixel convolution, which rearranges learned feature tensors into spatially consistent classification maps. This dual module ensures spectral discrimination and spatial refinement in LULC classification.

The research paper is organized as follows: Section II analyzes existing research and provides their descriptions. Section III gives details of the proposed method and dataset description. Section IV presents the results and comparison of the developed approach. The conclusion of this research is given in Section V.

II. Literature Review

Aljebreen *et al.* [16] suggested the LULC Classification assigned River Formation Dynamics Algorithm with DL (LULCC-RFDADL) model on RS. In the suggested method, a dense EfficientNet technique was employed to extract features. Hyperparameter tuning of the Dense EfficientNet technique was introduced using the RFDA method. In the classification procedure, the suggested method utilized a Multi-Scale Convolutional AutoEncoder (MSCAE) technique. The Seeker Optimization Algorithm (SOA) was employed for parameter selection of the MSCAE method. The suggested method effectively classified different LULC classes but was sensitive to variantions in data quality and characteristics.

Temenos *et al.* [17] presented an interpretable DL framework to classify LULC in RS using Shapley Additive Explanations (SHAP). The presented method utilized a conventional CNN to classify satellite images and then fed the outcomes into the SHAP deep explainer, which strengthened the classification results. The presented method improved the classification accuracy of each individual class, and interpretability led to better channel estimation. However, the presented method struggled to differentiate between similar classes, causing classification errors.

Albarakati *et al.* [18] implemented a fully optimized self-attention-fused CNN model to classify LULC using RS images. A new contrast enhancement equation was introduced and used in the implemented method for data augmentation. Next, a fused self-attention CNN model was introduced and used for LULC classification. The implemented method achieved good accuracy and precision rates on selected features. However, the method had difficulty differentiating land covers within a pixel, which degraded the classification performance.

Rubab *et al.* [19] introduced a network-based fusion deep structure on a 16-tiny Vision Transformer. In the early stage, data augmentation was applied to resolve data imbalance. Next, a self-attention bottleneck based on the Inception CNN, named SIBNet, was developed. Blocks were introduced using the Inception model, and each Inception block was transformed into bottleneck blocks. The hyperparameters of the introduced method were performed using Bayesian Optimization to better train RS images. The introduced method extracted deep features from the self-attention layer and classified them using a neural network classifier with multiple hidden layers. However, the introduced approach did not scale the data uniformly, which resulted in lower classification performance.

Vinaykumar *et al.* [20] developed an Optimal Guidance-Whale Optimization Algorithm (OG-WOA) to select appropriate attributes and reduce overfitting. The OG model maximized exploitation of the search method by modifying the location of the search agent to achieve better fitness scores. Input images were normalized and used with the AlexNet-ResNet 50 technique to extract features. The OG-WOA method was employed in feature extraction for selecting appropriate features. Finally, the chosen attributes were performed to classification through a Bi-directional Long Short-Term Memory (Bi-LSTM) model. However, the developed model did not capture deep features, which led to lower classification performance.

Rehman *et al.* [21] developed a hybrid method integrating CNN and transformer models. The model included three primary components: the Spectral Spatial Convolutional Module (SSCM), the Spatial Attention Module (SAM), and the Transformer Module (TM). Shailaja *et al.* [22] implemented an Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) to minimize noise using a super-resolution concept, enhancing image quality by integrating deep learning with adversarial training. The model used the Swin Transformer Convolutional Neural Network (ST-CNN), which employed a self-attention mechanism to extract intricate spatial features and temporal dynamics.

From the analysis of existing techniques, these methods struggled to differentiate various land covers within the same pixel, and some techniques did not scale the data uniformly. These drawbacks reduced classification performance and led to classification errors due to class

similarity and mixed-pixel issues. To mitigate these limitations, in this manuscript, a deformable convolutional layer is integrated with the convolutional layer to capture deep features, helping to differentiate between different land covers at the pixel level. Then, a sub-pixel convolutional layer is included in the output layer to classify different LULC classes effectively.

III. Proposed Methodology

An efficient DL approach is developed to effectively classify various classes of LULC. The dataset utilized for this research are EuroSAT and NWPU-RESISC45. The images in these datasets are preprocessed using the Min-Max normalization approach, which scales the data. The preprocessed images are given to the classification phase, where the developed DSPCL with CNN method is used, which effectively extracts deep features and classifies different LULC classes. Fig. 1 represents the process of LULC classification.

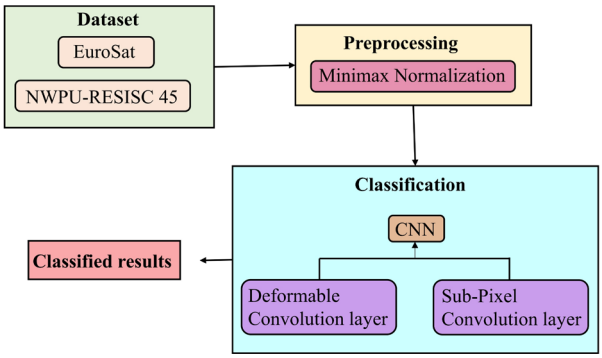

Fig. 1. Process of LULC classification.

A. Dataset

This research used two datasets, (1) EuroSAT: https://www.kaggle.com/datasets/apollo2506/eurosat-dataset (Accessed on November 2025); (2) NWPU-RESISC45: https://www.kaggle.com/datasets/aqibrehmanpirzada/nwpuresisc45 (Accessed on November 2025), for LULC classification. These datasets are publicly available for research purposes in the domain of RS. The description of these two datasets is provided below.

1) EuroSAT

This dataset includes 27,500 images with 10 classes and is used for LULC classification. Each image in the dataset consists of 64×64 pixels and ground sampling distance of 10 m. The dataset is collected through the Sentinel-2 satellite. Fig. 2 represents the dataset description of EuroSAT, and Fig. 3 shows sample images from the EuroSAT dataset.

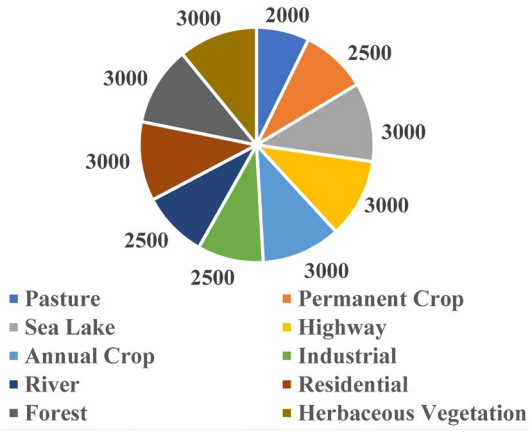

Fig. 2. Dataset description of EuroSAT.

Fig. 3. Sample images in EuroSAT dataset.

2) NWPU-RESISC 45

This dataset was developed through Northwestern Polytechnical University (NWPU) and is publicly available for RS researchers. This dataset includes 700 to 1400 images with 12 classes and a pixel resolution of 256×256. This dataset is challenging due to high image variability, certain discrepancies among scene classes, and similarities across different scenes. Fig. 4 represents the dataset description of the NWPU-RESISC45 dataset, and Fig. 5 shows sample images from the NWPU-RESISC45 dataset.

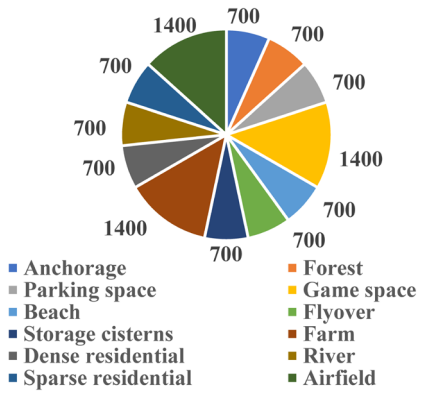

Fig. 4. Dataset description of NWPU-RESISC 45.

Fig. 5. Sample images in NWPU-RESISC 45 dataset.

B. Pre-Processing

Images in the dataset are fed into the preprocessing stage, which normalizes the image into a certain range. In this research, the Min-Max normalization approach is utilized to scale pixel values in the range of 0 to 1 [23]. This ensures that every pixel intensity from the dataset is uniformly scaled and helps the model to learn patterns efficiently. The mathematical expression for Min-Max normalization is shown in Eq. (1).

$$I_{out} = (I_{in} - Min)\frac{newMax - newMin}{Max - Min} + newMin \quad (1)$$

In Eq. (1), Min and Max represent the minimum and maximum pixel intensities, I_{out} represents the normalized output, I_{in} represents the original input image, and $(Max - Min)$ represents the difference between maximum and minimum pixel intensity values in the range of 0 and 1.

C. Classification

The preprocessed images are fed as input to the classification phase, where features are extracted and various LULC classes are classified using the developed DSPCL with CNN. This method includes feature extraction and classification phases. The features are extracted using convolutional layers. The issues of spectral similarity and mixed pixels are mitigated by using deformable and sub-pixel convolution layers, which are incorporated into the CNN for classification. The feature extraction phase, along with the process of the deformable layer and sub-pixel convolution layers, is explained in the sections below.

1) Network architecture

Two convolutional layers, Conv1 and Conv2, are used in the feature extraction phase to capture meaningful features. In these convolutional layers, 64 convolutional kernels of size 5×5 are used to ensure low-dimensional feature data. The network considers a Y- channel image as the input to Conv1 and processes feature extraction with 64 convolutions to obtain low-dimensional features. The mathematical formulation is given in Eq. (2).

$$F_i(Y) = tanh(W_i \times Y + B_i) \quad (2)$$

In Eq. (2), B_i and W_i represent the weights and biases of the model. The size of W_i is represented as $c_i \times n_i \times f_i \times f_i$ n_i represents the number of convolutions kernels, and tanh represents the activation function. The Conv3 layer considers the $F_1(Y)$ feature image set, and $F_2(Y)$ represents the convolution branch input, processing convolution with a 5×5 kernel to obtain $F_3(Y)$. Conv4 considers $F_3(Y)$ as input and processes the next mapping to 32 convolutions with a 5×5 kernel to obtain the $F_4(Y)$ feature image. The next layer, Conv5, considers $F_3(Y)$ and $F_4(Y)$ as inputs and processes the third mapping to 32 convolution filters using a 3×3 kernel to acquire features with high-dimension and texture information. Next, Conv6 considers $F_5(Y)$ with 32 convolutions of 5×5 kernel size to obtain a high-dimensional global feature map. The processes of Conv4 and Conv6 are the same as Conv1, and their mathematical formulation is given in Eq. (3).

$$F_i(Y) = ReLU \left[0, W_i \times \left(k_1 \times F_{i-1}(Y) + k_2 \times F_{i-2}(Y)\right) + B_i)\right] \quad (3)$$

In Eq. (3), i represents the layer number, and k_1 and k_2 represent the ratios of the $i-1$ and $i-2$ feature map layers. To attain the final classified output, the deformable layer and sub-pixel convolutional layer are implemented for reconstructing the output image. In the Conv7 layer, the global feature map $F_6(Y)$ is utilized to reconstruct and attain the final image. The number of channels becomes r^2, which is processed through convolutions and a residual layer, and the images are rearranged in the sub-pixel convolution layer for full image reconstruction. The mathematical formulation for this process is given in Eq. (4).

$$I^{SR} = PS(W_i \times F_{i-1}(Y) + B_i) \quad (4)$$

In Eq. (4), the size of W_i is represented as r^2, c represents the number of input channels in the initial image, *PS* represents the sub-pixel convolution process in which channel images are rearranged according to pixel positions, features are associated, and I^{SR} denotes the image patterns after reconstruction.

2) Deformable layer

The primary benefit of the deformable convolution is that it adaptively changes the filter form, which is significant for addressing the issue of class similarity. The default convolution has two major phases: (1) R represents the regular grid utilized for sampling over the input feature map, and (2) the summation is used to add sample scores weighted through w. The resulting feature map, without loss of generality, is represented in Eqs. (5) and (6).

$$y(P_0) = \sum_{P_n \epsilon R} w(p_n) \times x(p_0 + p_n) \quad (5)$$

$$R = \{(-1,-1),\ (-1,0), \ldots,\ (0,1),\ (1,1)\} \quad (6)$$

In Eqs. (5) and (6), p_0 represents the position of the result, and p_n represents a position in the regular sampling grid. For execution, the deformable filter initially convolves on the input feature map and then the learned offsets are applied to the feature map, providing conditional transformation of R input. Here, Δp augments R, where Δp takes real number in an unconstrained range. The mathematical formulation for the output y is given in Eqs. (7) and (8).

$$y(p_0) = \sum_{p_n \in R} w(p_n) \times (p_0 + p_n + \Delta p_n) \quad (7)$$

$$\{\Delta p_n | n = 1,\ \ldots,\ N\},\ N = |R| \quad (8)$$

The grid generator samples R at $p_0 + p_n$ irregularly. The predicted samples are used to generate an irregular sampling grid that includes a group of locations in the input map sampled in a flexible form. Bilinear interpolation is used, and its formulation is given in Eq. (9).

$$x(p_0) = \sum_q G(q,\ p) \times x(q) \quad (9)$$

In Eq. (9), p is the arbitrary fractional position and q is the input at the integral position of x. The bilinear interpolation kernel is represented as G, which includes two dimensions. Its mathematical formulation is given in Eqs. (10) and (11).

$$G(p,q) = g(p_x,\ q_x) \times g(p_y,\ q_y) \quad (10)$$

$$g(a,b) = max(0, 1 - |a - b|) \quad (11)$$

In these equations, there are only small number of non-zero elements within $G(p,q)$. These integrated phases develope the deformable convolution.

3) Sub-pixel convolutional layer

In this layer, initial features are interpolated through a bicubic technique according to the downsampling process used for reconstructing the classified images. Consider a sampling factor represented as r and c, which is the number of colour channels. Let $H \times W \times c$ and $rH \times rW \times c$ represent the dimensions of the actual initial features and the reconstructed High-Resolution (HR) image, respectively. Hence, the sub-pixel convolutional layer is implemented to perform the upsampling process at the output layer of the CNN. The input to this layer is a feature map with r^2 channels for every pixel in the image of size $r \times r$. The features with an actual tensor dimension of $H \times W \times C \times r^2$ are converted to $rH \times rW \times C$. To obtain the reconstruction effect, this layer rearranges the feature vector dimensions according to specific rules. When the process is performed with $r = 2$, pixels in similar locations in the features are rearranged and integrated to form respective blocks, where each group of four pixels forms one area. The feature space employs convolution process with filter W_s of size f_s and weight interval of $\frac{1}{r}$. The partial weights within the filter is activated while the remaining weights of pixel in the calculation remain inactive. The number of activations is r^2, and the weights are activated based on location distribution. In the convolution process, the filter scans images according to sub-pixel location, and scanning weights are activated. Consider $mod(x,r)$ and $mod(y,r)$ as the horizontal coordinate x and vertical coordinate y of the resultant pixel within the feature space. On sub-pixel convolution process, the original upsampling process for $mod(f_s, r) = 0$ is given in Eqs. (12) and (13).

$$F_L(Y) = PS(W_L \times Y + b_L) \quad (12)$$

$$PS(T)_{x,y,c} = T_{[x/r],[y/r],C \times r \times mod(y,r) + C.mod(x,y) + C'} \quad (13)$$

In these equations, Y represents the feature with actual tensor dimensions $H \times W \times C \times r^2$, and W_L represents its size. $PS(T)$ represents the periodic permutation process performed in the sub-pixel convolutional layer.

IV. Experimental Analysis

The developed DSPCL with CNN model was simulated using Python 3.7 in a Windows 10 (64 bit) environment with an i5 processor and 8 GB RAM. Evaluation measures such as accuracy, specificity, recall, precision, and F1-Score were considered to validate the performance of the DSPCL with CNN method. The mathematical formulas for these performance measures are given in Eqs. (14)–(18).

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \times 100 \quad (14)$$

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (15)$$

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

$$Precision = \frac{TP}{TP+FP} \quad (17)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recll} \quad (18)$$

In the above equations, FP and FN denote False Positives (FP) and False Negatives (FN), and TP and TN denote True Positives (TP) and True Negatives (TN), respectively. Table I represents the parameters of the proposed method.

TABLE I. Parameters of the Proposed Model

Parameter	Value
Optimizer	Adam
Learning rate	0.001
Batch size	32
Number of epochs	100
Size of image	224 × 224
Early stopping	10
Split ratio	80:10:10

To ensure fair and consistent validation, all baseline models were reproduced under the same experimental settings as the proposed DSPCL-CNN, using the same optimizer, batch size, and training epochs. For advanced models such as the Vision Transformer (ViT) and Swin Transformer, the outcomes were adopted as these methods were pre-trained and optimized on the same dataset. This ensured that performance differences accurately represented architectural enhancements under experimental conditions.

Table II represents the performance of the developed DSPCL with CNN method evaluated on different classes of the EuroSAT dataset. Evaluation measures such as accuracy, sensitivity, precision, recall, and F1-Score were utilized to validate its performance. The EuroSAT dataset contains 10 classes, and their results and average results are evaluated and presented in Table III. The developed DSPCL with CNN method achieved an average accuracy of 98.62%, sensitivity of 98.38%, precision of 98.14%, recall of 97.89%, and F1-Score of 98.01% on the EuroSAT dataset.

Table III shows the performance of the developed DSPCL with CNN method evaluated on various classes of the NWPU-RESISC45 dataset. The dataset has 12 classes, and their results and average results are represented in Table IV. The developed DSPCL with CNN method achieved an average accuracy of 99.01%, sensitivity of 98.63%, precision of 98.43%, recall of 98.30%, and an F1-Score of 98.36% on the NWPU-RESISC45 dataset.

Table IV validates the performance of the developed DSPCL with CNN method on the EuroSAT dataset using different evaluation measures. The classifiers considered include VGG-16, ResNet50, Multi-Layer Perceptron (MLP), and a traditional CNN. The developed DSPCL with CNN method achieved 98.62% accuracy, 98.38% sensitivity, 98.14% precision, 97.89% recall, and an F1-Score of 98.01%.

TABLE II. CLASS-BASED EVALUATION ON EUROSAT DATASET

Classes	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F1-Score (%)
Pasture	97.28	97.10	96.94	96.87	96.90
Industrial	97.59	97.38	97.27	97.08	97.17
Permanent Crop	97.93	97.84	97.44	97.38	97.40
River	98.31	98.01	97.65	97.54	97.59
Sea Lake	98.52	98.27	98.09	97.82	97.95
Residential	98.76	98.53	98.37	98.03	98.19
Highway	99.19	98.86	98.54	98.20	98.36
Forest	99.41	99.06	98.73	98.47	98.59
Annual Crop	99.55	99.32	99.17	98.76	98.96
Herbaceous Vegetation	99.67	99.48	99.22	98.83	99.02
Average	98.62	98.38	98.14	97.89	98.01

TABLE III. CLASS-BASED EVALUATION ON NWPU-RESISC 45 DATASET

Classes	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F1-Score (%)
Anchorage	97.99	97.46	97.18	97.05	97.11
Beach	98.23	97.68	97.47	97.33	97.39
Dense residential	98.36	97.89	97.63	97.52	97.57
Forest	98.57	98.19	98.01	97.85	97.92
Flyover	98.73	98.39	98.17	98.02	98.09
River	99.02	98.58	98.37	98.16	98.26
Parking space	99.14	98.79	98.52	98.49	98.50
Storage cisterns	99.25	98.94	98.77	98.63	98.69
Sparse residential	99.47	99.18	99.06	98.95	99.00
Game space	99.62	99.35	99.23	99.08	99.15
Farm	99.86	99.49	99.37	99.21	99.23
Airfield	99.91	99.68	99.41	99.35	99.37
Average	99.01	98.63	98.43	98.30	98.36

TABLE IV. PERFORMANCE OF DSPCL WITH CNN METHOD ON EUROSAT DATASET

Methods	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F1-score (%)
VGG-16	96.19	95.83	95.67	95.33	95.49
ResNet 50	96.68	96.27	96.05	95.78	95.91
MLP	97.21	96.85	96.47	96.32	96.39
CNN	97.78	97.18	96.82	96.57	96.69
ViT	97.86	97.52	97.27	97.02	97.16
Swin Transformer	98.44	98.25	98.05	97.66	97.82
DSPCL with CNN	98.62	98.38	98.14	97.89	98.01

Table V validates the performance of the developed DSPCL with CNN method on the NWPU-RESISC45 dataset using the same measures. The developed DSPCL with CNN method achieved 99.01% accuracy, 98.63% sensitivity, 98.43% precision, 98.30% recall, and an F1-Score of 98.36%.

Table VI shows the performance of the proposed DSPCL-CNN model compared to different CNN methods on the EuroSAT and NWPU-RESISC45 dataset, demonstrating consistent improvement achieved by each architectural improvement. The traditional CNN obtained accuracies of 97.78% and 98.76%, which increased with

the addition of the deformable convolution layer, showing its capability to adaptively capture spatially variant features and address class similarity. Similarly, incorporating the sub-pixel convolutional layer enhanced accuracy to 98.21% and 98.84% through refined pixel-level reconstruction and resolution of mixed-pixel problems. The proposed DSPCL-CNN model, integrating both deformable and sub-pixel layers, achieved 98.62% accuracy for EuroSAT and 99.01% for NWPU-RESISC45, demonstrating the strength of adaptive feature extraction and fine spatial reconstruction in improving LULC classification performance.

TABLE V. PERFORMANCE OF DSPCL WITH CNN METHOD ON NWPU-RESISC45 DATASET

Methods	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F1-Score (%)
VGG-16	97.82	97.63	97.47	97.05	96.52
ResNet 50	98.17	97.85	97.61	97.31	97.08
MLP	98.52	98.07	97.86	97.55	97.38
CNN	98.76	98.32	98.05	97.84	98.17
ViT	98.85	98.43	98.21	98.02	98.21
Swin Transformer	98.94	98.57	98.35	98.23	98.30
DSPCL with CNN	99.01	98.63	98.43	98.30	98.36

TABLE VI. ABLATION STUDY OF DSPCL-CNN WITH INDIVIDUAL AND COMBINED COMPONENTS

Dataset	Models	Accuracy (%)
EuroSAT	CNN	97.78
	CNN + Deformable layer	98.32
	CNN + Sub-pixel layer	98.21
	Proposed DSPCL-CNN (Complete model)	98.62
NWPU-RESISC45	CNN	98.76
	CNN + Deformable layer	98.91
	CNN + Sub-pixel layer	98.84
	Proposed DSPCL-CNN (Complete model)	99.01

Table VII represents the cross-dataset validation of the proposed DSPCL-CNN model and the traditional CNN, where training and testing were conducted on different dataset to assess model generalization. When trained on EuroSAT and tested on NWPU-RESISC45, the proposed model achieved 94.82% accuracy and 94.11% F1-Score. When trained on NWPU-RESISC45 and tested on EuroSAT, the proposed model achieved 95.24% accuracy and 94.79% F1-Score outperforming the traditional CNN. These results determine the superior generalization ability of the proposed method, attributed to the deformable convolution layer's capability to extract spatial features and the sub-pixel layer's precision in reconstructing fine-grained spatial patterns. This integration enables the model to maintain consistent performance across dataset with varying resolutions and scene complexities, proving its robustness and transferability for LULC classification.

Table VIII represents the computational analysis of the proposed DSPCL-CNN shows slightly higher computational costs with an increase in training and inference time. These additional costs result from integrating deformable and sub-pixel convolutional layers that improve adaptive feature extraction and fine-grained spatial reconstruction. Despite the increased complexity, the improvement in classification accuracy and spatial consistency determines that the proposed model attains a superior balance between computational efficiency and performance for LULC classification.

TABLE VII. CROSS-DATASET GENERALIZATION ANALYSIS FOR DSPCL-CNN MODEL

Training and Testing Dataset	Models	Accuracy (%)	F1-Score (%)
Trained on EuroSAT and tested on NWPU-RESISC45	CNN	92.03	91.62
	Proposed DSPCL-CNN	94.82	94.11
Trained on NWPU-RESISC45 and tested on EuroSAT	CNN	93.14	92.33
	Proposed DSPCL-CNN	95.24	94.79

TABLE VIII. COMPUTATIONAL ANALYSIS OF PROPOSED MODEL

Methods	Training Time (s)	Inference Time Per Image (ms)	Parameters ($\times 10^6$)	FLOPs ($\times 10^6$)
CNN	1840	3.2	1.8	42.5
Resnet-50	2470	4.5	2.4	65.2
ViT	3120	5.1	3.2	71.6
Swin-Transformer	3290	5.5	3.5	78.3
Proposed DSPCL-CNN	3560	5.8	3.9	84.1

Figs. 6 and 7 show the confusion matrices representing the classification performance of the proposed DSPCL-CNN model on the NWPU-RESISC45 and EuroSAT dataset. Fig. 6 determines strong class-wise consistency across 45 classes, representing the model's capability to differentiate visually similar land-cover types. Similarly, Fig. 7 shows high classification precision across 10 classes with minimal confusion among adjacent land types. This superior performance results from the adaptive feature extraction of the deformable convolution layer, which overcomes class similarity, and the fine-grained spatial reconstruction of the sub-pixel convolution layer, which resolves mixed-pixel ambiguities.

Figs. 8 and 9 represent the Receiver Operating Characteristic (ROC) curves for the NWPU-RESISC45 and EuroSAT dataset, respectively.

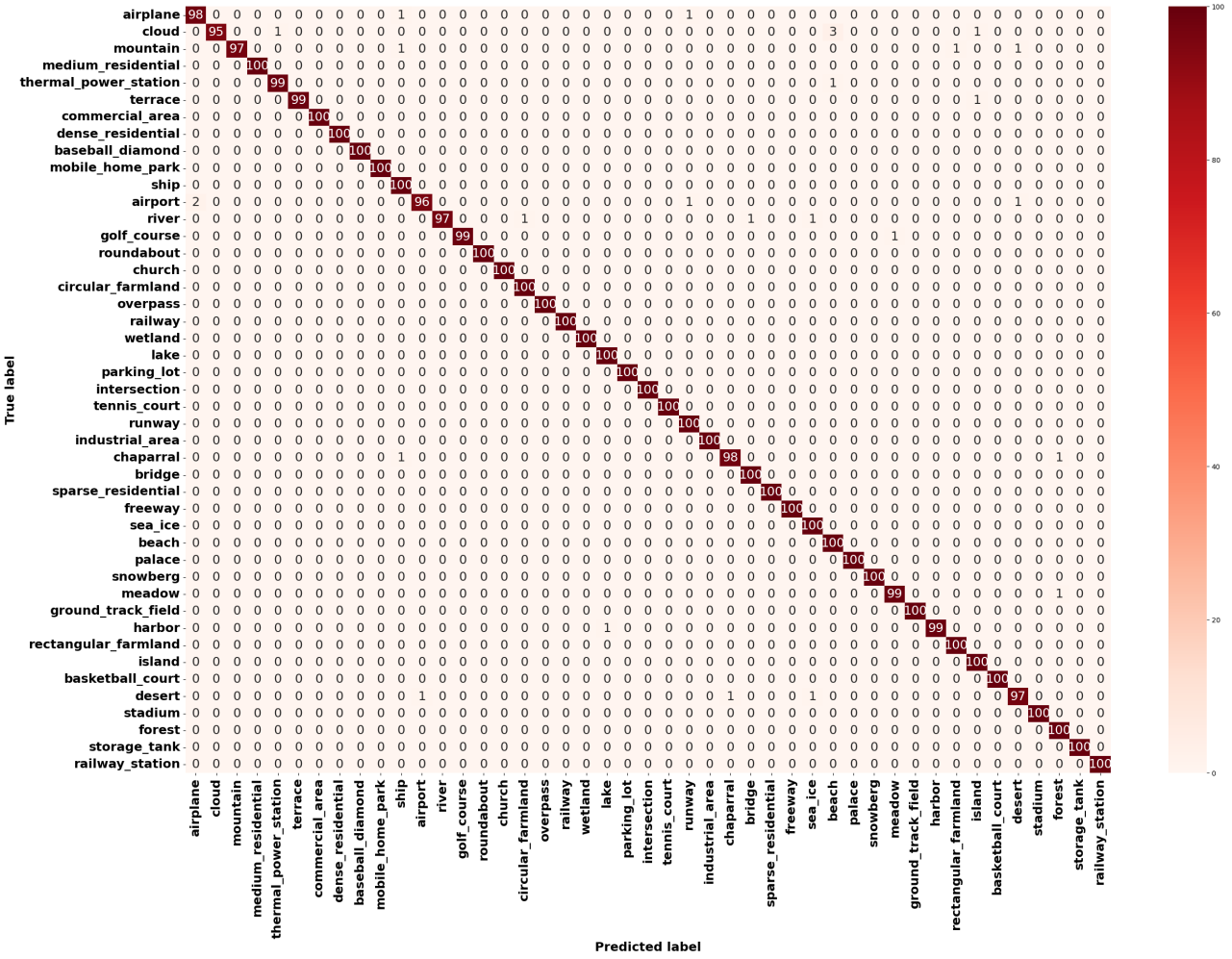

Fig. 6. Confusion matrix of NWPU-RESISC45 dataset.

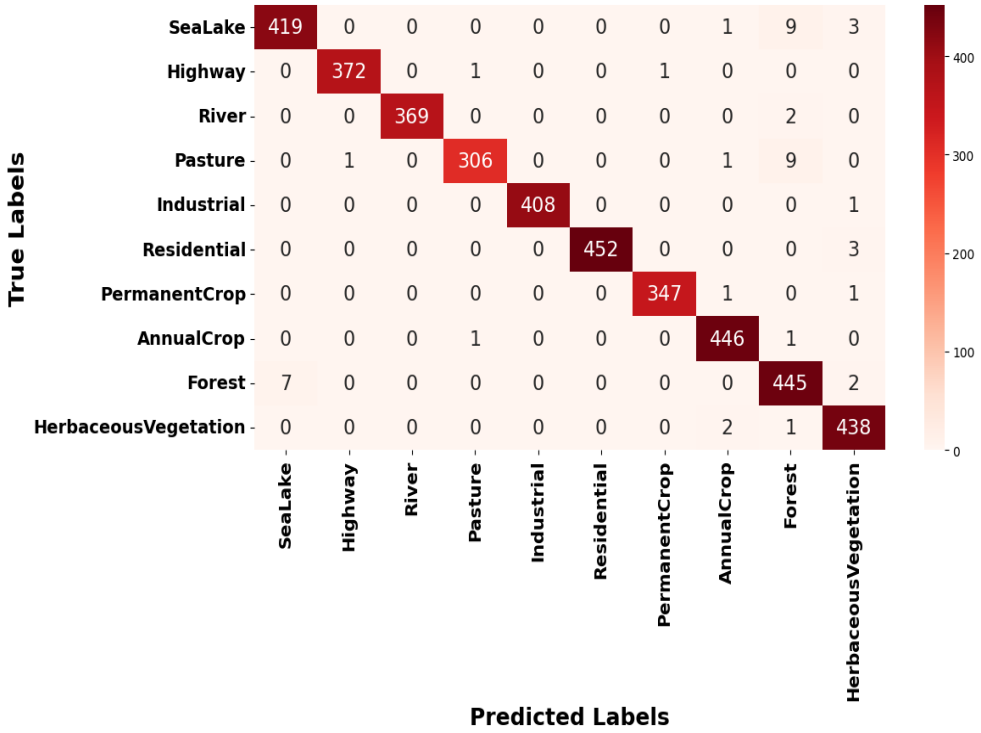

Fig. 7. Confusion matrix of EuroSAT dataset.

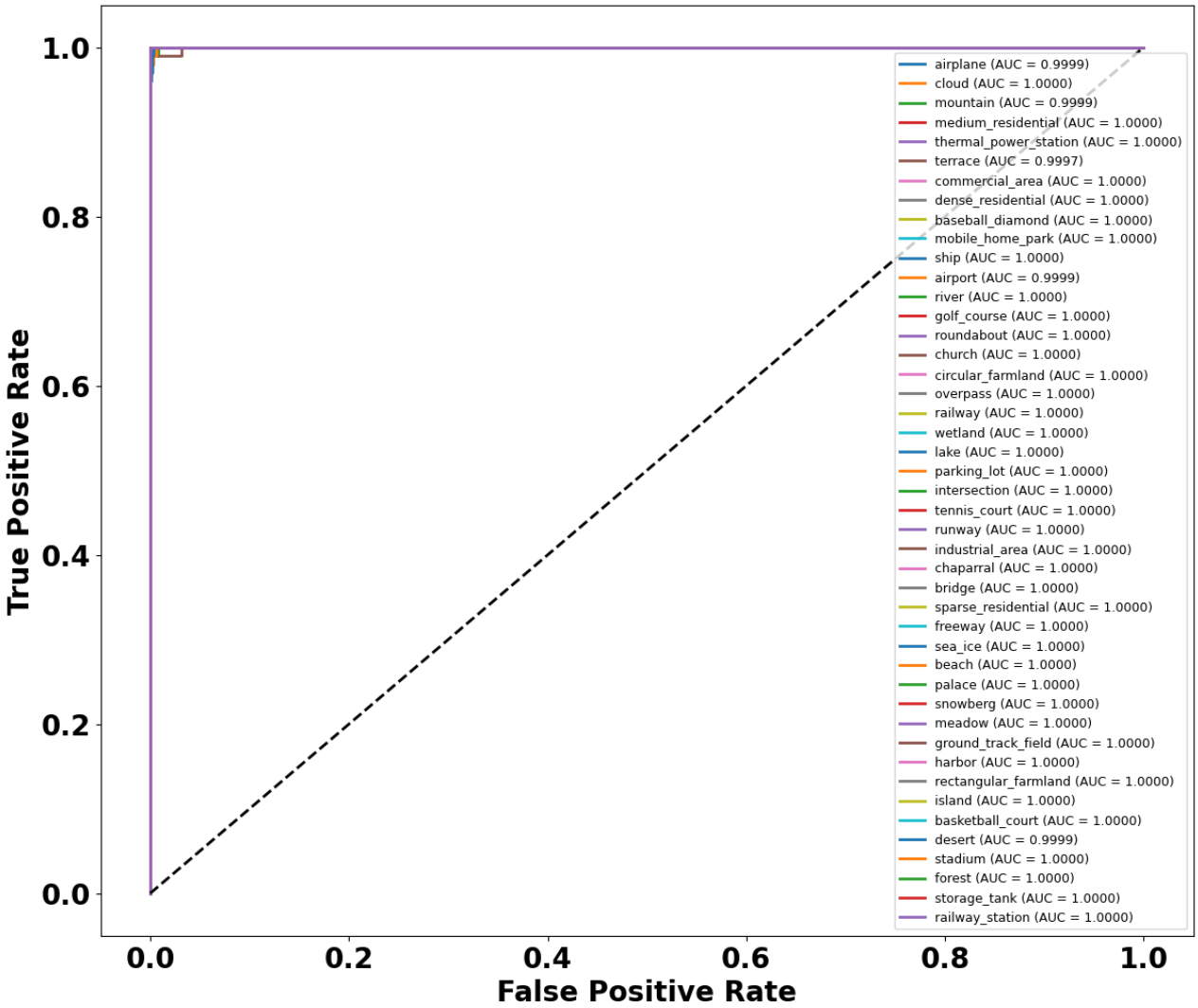

Fig. 8. ROC curve of NWPU-RESISC45 dataset.

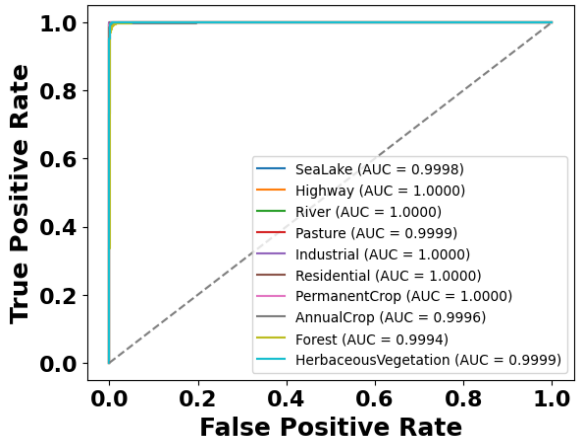

Fig. 9. ROC curve of EuroSAT dataset.

Figs. 10 and 11 illustrate the qualitative visualization outcomes of the proposed DSPCL-CNN model on the EuroSAT and NWPU-RESISC45 dataset, showing correctly and incorrectly classified samples. Each image pair represents the true land-cover class and the predicted class, demonstrating how the model effectively identifies different surface types such as industrial areas, residential zones, water bodies, highways, and vegetation. The activation maps show that the model focuses on important spatial and textural regions related to each class. Misclassifications, such as between Pasture and Forest, mainly occur in visually overlapping regions, indicating the inherent challenge of mixed pixels in RS images. The deformable convolutional layer enables adaptive feature learning, while the sub-pixel layer refines boundaries and improves fine-grained spatial discrimination.

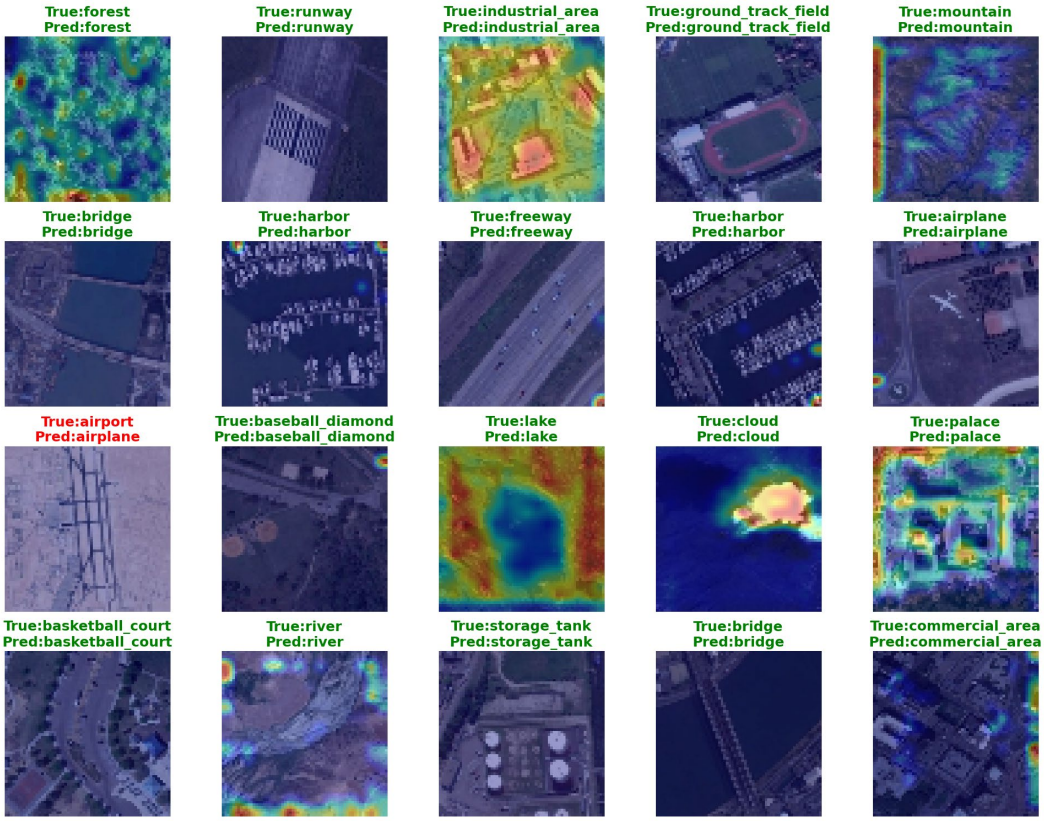

Fig. 10. Misclassified samples of NWPU-RESISC45 dataset.

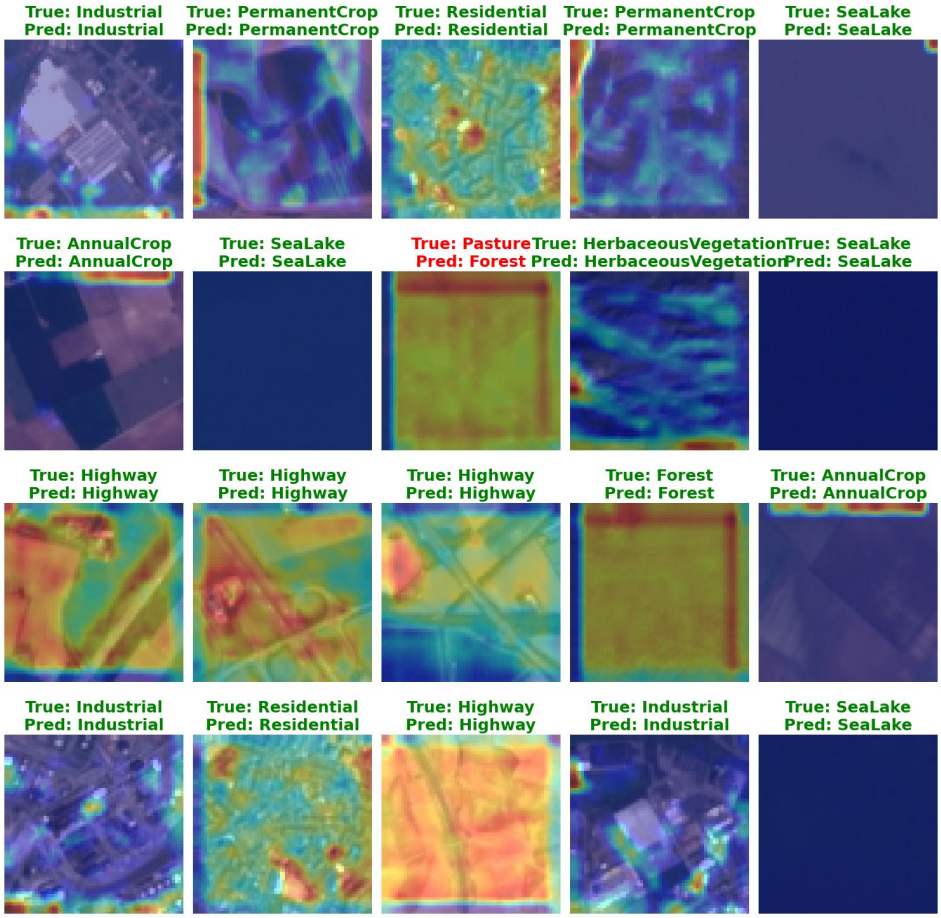

Fig. 11. Misclassified samples of EuroSAT dataset.

A. Comparative Analysis

The performance of the DSPCL with CNN model was compared with existing techniques such as LULCC-RFDADL [16], Deep XAI based on SHAP [17], Optimized Self-attention fused CNN [18], and Self-attention bottleneck-based inception CNN [19] on the EuroSAT and NWPU-RESISC45 dataset. By evaluating the performance, the DSPCL with CNN method demonstrated improved results compared with existing methods. Incorporating the deformable convolutional layer instead of the standard convolutional layer mitigated class similarity between images.

The sub-pixel convolutional layer, incorporated into the output layer, separated pixels and helped differentiate land-use and land-cover categories. These processes improved LULC classification performance compared with existing techniques. Table IX presents the comparative analysis of the DSPCL with CNN model on the EuroSAT and NWPU-RESISC45 dataset. The superior performance of the proposed DSPCL-CNN model is attributed to the integration of the deformable convolutional layer and the sub-pixel convolutional layer, which improve spatial adaptability and fine-grained reconstruction. The deformable convolutional layer dynamically adjusts its sampling positions in accordance with geometric and spectral variations in remote-sensing scenes. This mechanism allows the model to adaptively capture informative regions, enhancing differentiation between spectrally similar classes in the dataset. The sub-pixel convolutional layer restructures feature maps in the output phase and enables pixel-level upscaling and refinement. This process minimizes mixed-pixel effects, especially in boundary-rich classes, and results in clear class separation. These two modules strengthen spectral discrimination and spatial consistency of classification maps, which explains the notable enhancements compared with traditional models. Additionally, deformable convolution minimizes sensitivity to local distortions and texture irregularities, when sub-pixel reconstruction ensures spatial coherence, enhancing robustness under varying image resolutions. The incorporation of these modules increases computational complexity; however, the resulting gain in classification performance determines that the model improvements directly contribute to the performance benefit of the proposed DSPCL-CNN model.

TABLE IX. COMPARATIVE ANALYSIS OF DSPCL WITH CNN METHOD

Datasets	Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EuroSAT	LULCC-RFDADL [16]	98.15	90.79	90.79	90.76
	Deep XAI based on SHAP [17]	94.72	-	-	-
	Optimized Self-attention fused CNN [18]	89.50	88.63	88.65	88.63
	Self-attention bottleneck-based inception CNN [19]	97.8	97.0	96.97	96.98
	Proposed DSPCL with CNN	98.62	98.14	97.89	98.01
NWPU-RESISC 45	Optimized Self-attention fused CNN [18]	91.70	91.91	91.44	91.67
	Self-attention bottleneck-based inception CNN [19]	98.9	98.26	98.13	98.19
	Proposed DSPCL with CNN	99.01	98.43	98.30	98.36

B. Discussion

The experimental outcomes shows that the proposed DSPCL-CNN method efficiently addressed two major challenges in LULC classification: class similarity and mixed-pixel ambiguity. By combining a deformable convolutional layer and a sub-pixel convolutional layer, the proposed model adaptively learns spatially variant features and refines pixel-level boundaries, resulting in superior accuracy across multiple datasets. The ablation and cross-dataset analyses show that this model significantly improves generalization and robustness compared with existing models. Beyond numerical improvement, the findings have broad implications for practical applications. The proposed model supports large-scale land-cover monitoring and urban planning by offering more spatially coherent and spectrally discriminative classification maps. Its ability to adaptively capture fine-grained features makes it suitable for different geographic and climatic regions, enhancing interpretability and reliability of LULC maps. Despite its promising performance, the proposed DSPCL-CNN method has several challenges. The incorporation of deformable and sub-pixel layers increases computational complexity and memory requirements, which limit real-time deployment. Additionally, the dataset utilized are relatively balanced and clean, which do not completely represent the heterogeneous real-world data affected by noise or sensor variations. For future research, the model will be extended by integrating lightweight models to minimize complexity while maintaining accuracy. Evaluating the model on large-scale, multi-sensor dataset and developing domain-adaptive training strategies can further improve its generalization.

V. CONCLUSION

A novel method, DSPCL with CNN, was developed in this research by including a deformable convolutional layer and a sub-pixel convolutional layer. The deformable convolution layer, used instead of the standard convolution layer, captured deep features and supported differentiation among various land-cover classes. The sub-pixel convolutional layer, integrated into the output layer of the CNN, classified different land-cover classes with high accuracy. In the preprocessing phase, the Min-Max normalization approach was utilized to scale data within a certain range, improving classification performance. The developed DSPCL with CNN method achieved 98.62% accuracy on the EuroSAT dataset and 99.01% accuracy on the NWPU-RESISC45 dataset. As future work, various CNN variants can be explored to further enhance the classification performance of LULC in RS images.

Although the proposed DSPL-CNN achieved high results on the EuroSAT and NWPU-RESISC45 dataset, these datasets are relatively balanced and clean, which may not fully represent real-world challenges such as noise, class imbalance, and heterogeneous resolutions. Furthermore, the additional computational cost of deformable and sub-pixel layers may limit their use in resource-constrained settings. Future research can focus on validating the model on more complex dataset, developing lightweight variants, and integrating transformer-based models to improve classification robustness.

CODE AVAILABILITY STATEMENT

The source code developed for the proposed DSPCL-CNN model will be made publicly available as open-source on GitHub within three months after the publication of this paper, following the completion of code optimization. Currently, it can be obtained via the author's email address.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, validation, KRK and MCSVVSN; methodology, software, formal analysis, data curation, writing—original draft preparation, visualization, KRK; investigation, ASR; resources, project administration, SG; writing—review and editing, funding acquisition, RP; supervision, MCSVVSN; all authors had approved the final version.

REFERENCES

[1] M. Khan, A. Hanan, M. Kenzhebay *et al.,* "Transformer-based land use and land cover classification with explainability using satellite imagery," *Sci. Rep.*, vol. 14, no. 1, 16744, 2024.

[2] I. Papoutsis, N. I. Bountos, A. Zavras *et al.,* "Benchmarking and scaling of deep learning models for land cover image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 195, pp. 250–268, 2023.

[3] M. Maze, S. Attaher, M. O. Taqi *et al.,* "Enhanced agricultural land use/land cover classification in the Nile Delta using Sentinel-1 and Sentinel-2 data and machine learning," *ISPRS J. Photogramm. Remote Sens.*, vol. 229, pp. 239–253, 2025.

[4] S. N. MohanRajan, A. Loganathan, P. Manoharan *et al.,* "Fuzzy Swin transformer for land use/land cover change detection using LISS-III satellite data," *Earth Sci. Inf.*, vol. 17, no. 2, pp. 1745–1764, 2024.

[5] L. Martinez-Sanchez, L. See, M. Yordanov *et al.,* "Automatic classification of land cover from LUCAS in-situ landscape photos using semantic segmentation and a Random Forest model," *Environ. Modell. Softw.*, vol. 172, 105931, 2024.

[6] C. Acuña-Alonso, M. García-Ontiyuelo, D. Barba-Barragáns *et al.,* "Development of a convolutional neural network to accurately detect land use and land cover," *MethodsX*, vol. 12, 102719, 2024.

[7] M. Fayaz, J. Nam, L. M. Dang *et al.,* "Land-cover classification using deep learning with high-resolution remote-sensing imagery," *Applied Sciences*, vol. 14, no. 5, 1844, 2024.

[8] V. Pushpalatha, P. B. Mallikarjuna, H. N. Mahendra *et al.,* "Land use and land cover classification for change detection studies using convolutional neural network," *Appl. Comput. Geosci.,* vol. 25, 100227, 2025.

[9] S. Sawant and J. K. Ghosh, "Land use land cover classification using Sentinel imagery based on deep learning models," *J. Earth Syst. Sci.*, vol. 133, no. 2, pp. 1–23, 2024.

[10] G. Tejasree and L. Agilandeeswari, "Land Use/Land Cover (LULC) classification using deep-LSTM for hyperspectral images," *The Egypt. J. Remote Sens. Space. Sci.,* vol. 27, no. 1, pp. 52–68, 2024.

[11] G. Delogu, E. Caputi, M. Perretta *et al.,* "Using PRISMA hyperspectral data for land cover classification with artificial intelligence support," *Sustainability*, vol. 15, no. 18, 13786, 2023.

[12] A. Tzepkenlis, K. Marthoglou, and N. Grammalidis, "Efficient deep semantic segmentation for land cover classification using Sentinel imagery," *Remote Sens.*, vol. 15, no. 8, 2027, 2023.

[13] A. Irfan, Y. Li, E. Xinhua *et al.,* "Land use and land cover classification with deep learning-based fusion of SAR and optical data," *Remote Sens.*, vol. 17, no. 7, 1298, 2025.

[14] S. Sharma, R. Sedona, M. Riedel *et al.,* "Sen4Map: Advancing mapping with Sentinel-2 by providing detailed semantic descriptions and customizable land-use and land-cover data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2024.

[15] D. Kavran, D. Mongus, B. Žalik *et al.,* "Graph neural network-based method of spatiotemporal land cover mapping using satellite imagery," *Sensors*, vol. 23, no. 14, 6648, 2023.

[16] M. Aljebreen, H. A. Mengash, M. Alamgeer *et al.,* "Land use and land cover classification using river formation dynamics algorithm with deep learning on remote sensing images," *IEEE Access*, vol. 12, pp. 11147–11156, 2024.

[17] A. Temenos, N. Temenos, M. Kaselimi *et al.,* "Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 8500105, 2023.

[18] H. M. Albarakati, M. A. Khan, A. Hamza *et al.,* "A novel deep learning architecture for agriculture land cover and land use classification from remote sensing images based on network-level fusion of self-attention architecture," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 6338–6353, 2024.

[19] S. Rubab, M. A. Khan, A. Hamza *et al.,* "A novel network level fusion architecture of proposed self-attention and vision transformer models for land use and land cover classification from remote sensing images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2024.

[20] V. N. Vinaykumar, J. A. Babu, and J. Frnda, "Optimal guidance whale optimization algorithm and hybrid deep learning networks for land use land cover classification," *EURASIP J. Adv. Signal Process.*, vol. 2023, no. 1, 13, 2023.

[21] M. Z. U. Rehman, S. M. S. Islam, A. Ulhaq *et al.,* "Effective land use classification through hybrid transformer using remote sensing imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 18, pp. 2252–2268, 2024.

[22] P. Shailaja, P. M. Kumar, N. Nikhitha *et al.,* "LCC-Net: Swin transformer-CNN hybrid for enhanced land cover classification in natural disaster monitoring," *Syst. Soft Comput.,* vol. 7, 200303, 2025.

[23] A. Stateczny, S. M. Bolugallu, P. B. Divakarachari *et al.,* "Multiplicative long short-term memory with improved mayfly optimization for LULC classification," *Remote Sens.*, vol. 14, no. 19, 4837, 2022.