International Journal of Advanced Technology and Engineering Exploration, Vol 11(120)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2024.111100036

Research Article

OptTransEnsembleNet: deep ensemble learning framework for land use land cover classification

Kavita Devanand Bathe [1, 2*] and Nita Sanjay Patil[3]
Research Scholar, Datta Meghe College of Engineering, Sector-3, Airoli, Navi Mumbai, – 400708, Maharashtra, India[1]
Assistant Professor, K. J. Somaiya Institute of Technology, Somaiya Ayurvihar Complex, Eastern Express Highway, Near Everard Nagar, Sion (East), Mumbai – 400 022, Maharashtra, India[2]
Associate Professor, K. C. College of Engineering and Management Studies and Research, Mithbunder Road, Thane (East), 400603, Maharashtra, India[3]

Received: 06-January-2024; Revised: 05-November-2024; Accepted: 08-November-2024

Abstract
Precise land use land cover (LULC) classification plays a crucial role in urban and regional planning. Policymakers ought to have an in-depth understanding of the LULC of the landscape. Classical remote sensing methods of LULC classification are time-consuming and expensive. In recent years, deep learning (DL) has shown remarkable performances in remote sensing applications especially like LULC classification. In spite of its promising result, generalization ability of DL models, optimization of hyperparameters, and accuracy with limited datasets are major concerns of DL algorithms. In this study, an OptTransEnsembleNet DL framework which integrates deep transfer learning, deep ensemble learning, and optimization is proposed. Initially transfer learning approach is employed to train the 4 base learners on EuroSAT spectral indices derived dataset which is generated from benchmark EuroSAT dataset. The pretrained models InceptionV3, MobileNetV2, Xception, and InceptionResNetV2 are used as feature extractors and further custom layers are added. The hyperparameters of these layers of base learners are optimized with the grid search optimization algorithm. Next, the aforementioned 4 models are used as base learners in deep ensemble learning and a weighted average ensemble is created for LULC. The optimal weights for the weighted average ensemble are found using the grid search algorithm. Overall, OptTransEnsembleNet-ensemble of 4 base learners produce an overall accuracy (OA) of 96.40% which is higher in comparison to individual models such as MobileNetV2 (92.15%), Xception (93.52%), InceptionV3 (91.50%) and InceptionResNetV2 (93.80%). Likewise, the model is applied on Jagatsinghpur district in Odisha state of India for classification followed by generation of its LULC map. The proposed model helps to improve the generalization performance and robustness of a DL model. This study could be beneficial for urban planners and administrators for resource management.

Keywords
Deep learning, Remote sensing, Land use land cover classification, Ensemble learning, Optimization.

1.Introduction
Land use land cover (LULC) is a crucial aspect of landscape and sustainable resources. Land cover refers to physical cover present on the Earth's surface whereas land use refers to land utilization for socioeconomic prospects. LULC classification categorizes the earth observation features in to distinct classes such as water bodies, built-up area, crops, soil, forest etc.

The LULC classification is generally carried out for various application studies such as infrastructure analysis, forest cover analysis, wetland mapping and disaster management. Urbanization, population growth, and economic development have triggered considerable changes in LULC across the globe.

The LULC analysis is essential for administrators and policymakers to understand the environmental change dynamics [1]. Terrestrial mapping is the conventional method of LULC mapping. This method provides accurate and precise mapping; however, it is time-consuming, incomplete,

*Author for correspondence

uneconomical, resource-extensive, and impractical for mapping a larger area [2]. Remote sensing with its large area coverage is widely used for LULC applications. Recent accessibility to multi-source and multi-temporal remote sensing satellite imagery from diverse geographic regions offer new prospects to LULC classification. The multispectral optical data and synthetic aperture radar (SAR) based satellite imagery acquired from satellite sensors like LandSAT-8, Sentinel-2 and Sentinel-1 is widely employed in various LULC mapping studies [3]. In recent decades, machine learning (ML) algorithms have played a vital role in LULC mapping using earth observation data [4, 5]. Pixel-based image classification and object-based image analysis methods are commonly used in various studies [6]. ML techniques concerning LULC fall into various categories, like supervised, unsupervised, and semi-supervised learning. Supervised ML techniques like decision tree (DT), random forest (RF) [7], artificial neural network (ANN) [8], k-nearest neighbour (KNN) [9], gradient tree boosting (GTB) [10], support vector machine (SVM) [11] and the unsupervised learning methods such as fuzzy c-means algorithms [12], iterative self-organizing data (ISO data) [13] has grabbed special attention of researchers for LULC classification. These methods have shown notable performance but with their own set of limitations. These methods need manual features extraction [8, 10]. Recent advancements in computational resources have accelerated the usage of the deep learning (DL) approach for earth observation satellite imagery data [14–16]. DL has shown remarkable performance for LULC mapping tasks. In contrast to ML, DL has the capability of automatic feature extraction. DL techniques like convolution neural networks (CNNs) [17], recurrent neural network (RNN) [18], auto encoders and generative adversarial network (GAN) [19], dual attention guided deep encoder-decoder network [19], and UNet [20] are used in various remote sensing studies. Though DL has shown promising results in multiple studies but they are subject to certain limitations. This includes challenges related to the limited dataset, adverse conditions, privacy, class imbalance problem, optimized hyperparameter selection, model interpretation, and generalization capability. In the related literature, researchers have implemented CNN models from scratch or by adopting a transfer learning approach for LULC classification. Single models limit the performance of DL models. It can be improved by employing an ensemble learning method that combines multiple individual models to enhance the performance. Hyperparameters play a vital role in deep neural networks as they significantly impact the performance of DL models. Most studies in the literature have selected hyperparameters of deep neural networks by trial-and-error method. This is time-consuming and may not give correct results. This gap has motivated the development of a hybrid method that combines approaches such as deep transfer learning, ensemble learning, and optimization. The primary objective of this study is to propose a hybrid DL model that integrates optimized deep transfer learning with the ensemble learning approach to improve the model's performance and to generate the LULC map of the study area.

The main contributions are as follows:

1. The hyperparameters of deep neural networks are optimized using a grid search optimization algorithm.
2. A dataset is generated from spectral indices derived from the publicly available Sentinel-2 13-band EuroSAT dataset.
3. LULC classification is performed using a multi-stage, optimized deep ensemble learning model. Additionally, the deep ensemble learning model, OptTransEnsembleNet, is developed using an optimized weighted average ensemble method.
4. LULC classification for the Jagatsinghpur district in the Odisha state of India is conducted, and a LULC map is generated to evaluate the generalization performance of the proposed model.

The rest of the paper is organized as follows: Section 2 presents the literature review. Section 3 describes the materials and methods. Section 4 showcases the results, followed by a discussion of the results in Section 5. Finally, Section 6 concludes the paper.

2.Literature review

LULC classification is an important aspect of landscape studies. Over the years, researchers have employed numerous methods to accomplish this task. Conventional remote sensing methods are good, but they are time- consuming. ML methods [21] are employed in various studies of LULC classification and mapping. For instance, Amini et al. [22] applied RF algorithm on Landsat 9 time series satellite images to assess urban LULC change analysis of Isfahan region and achieved 94.43% overall accuracy (OA) and 0.93 Kappa value. Here, authors have employed only mono sensor approach with optical data. This poses limitations on performance of the ML model. The results can be further improved by adopting the multisensory approach of Landsat-9,

Sentinel-2 and Sentinel-1. In another study, Avcı et al. [23] have conducted the performance evaluation of SVM and RF algorithms for LULC classification in particular to wetlands.

In this study, RF outperformed SVM and achieved an OA of 93%. This accuracy could be further improved by integrating multiple datasets. Yuh et al. [24] analyzed the performance of various ML algorithms such as ANN, RF, KNN, and SVM on Landsat7 enhanced thematic mapper plus (ETM+) and Landsat 8 operational land imager (OLI) images of northern Cameroon region. The SVM algorithm has utilized radial basis function as kernel function for LULC classification. Further, ANN model architecture utilized 6 input neurons in input layer, one hidden layer and 8 neurons in output class representing 8 landcover classes. The SVM, KNN, ANN algorithms have achieved more than 80% accuracy whereas RF has achieved a remarkable classification accuracy of 90% with 500 number of estimators and 3000 training samples. Though this approach is robust, it does not consider nonlinearities in ecosystem dynamics in the region. This limitation can be handled by employing convergent cross mapping model.

In another study, Zhao et al. [25] evaluated the performance of ML algorithms on Sentinel-2 satellite imagery for LULC classification of vegetation, built-up, barren, and water bodies features. Authors utilized blue, green, red, near infrared (NIR) and short-wave infrared (SWIR) band from Sentinel-2 satellite imagery and further the spectral indices like normalized difference vegetation index (NDVI), normalized different built index, normalized difference water index (NDWI) and modified normalized difference water index (MNDWI) are computed and classification and regression tree (CART), SVM and RF algorithms are applied on this. CART algorithm is implemented with classifier.smileCart method and its default parameters available in google earth engine (GEE), while SVM is implemented with radial basis function (RBF) kernel and Classifier.libsvm method of GEE. The RF algorithm that is experimented with 100 to 500 number of estimators outperformed CART and SVM with an accuracy of 98.68% and kappa coefficient of 0.97. This work has shown remarkable result with the RF algorithm; however, its generalization ability is not tested across various regions. While few researchers explored supervised learning approaches, other researchers adopted unsupervised learning methods. For instance, Alshari and Gawali [26] experimented unsupervised learning methods like ISO data, k-means, principal components based spectral algorithm, simple linear iterative clustering, and principal components isometric binning, while other researchers explored semi-supervised learning methods like spectral angular mapper classification, self-trained models based semi-supervised classification, semi-supervised transudative support vector machine (TSVM) based semi-supervised classification, semi-supervised extreme learning machine, semi-supervised SVM algorithm, and graph model based semi-supervised classification. These methods have shown great performance. However, they rely on feature-based classification. The feature extraction is done manually, which may pose some limitations.

In recent years, DL methods [27, 28] have achieved substantial developments in LULC classification. CNN [29], due to its automatic feature extraction capability, is a popular choice among researchers. CNN can extract low-level and high-level features from input satellite images. The convolution operation is applied in the initial layers to extract more general features, whereas it is applied in the last layer to extract more specific features. DL models need a huge amount of training data to achieve better performances. Several datasets have been published over the period [30], but generally it is difficult to acquire a huge amount of training data for the study area. It is tedious to train a CNN model from scratch on a small dataset despite its feature extraction capability; Naushad et al. [31] attempted to address the aforementioned problem by implementing a transfer learning approach.

Helber et al. [32] have fine-tuned the CNN-based pre-trained networks visual geometry group (VGG16) and wide residual networks (WRNs) and implemented them on the EuroSAT dataset. The WRN method has shown promising results in terms of accuracy of 99.17% and is computationally more feasible as compared to VGG16. The approach utilized in this study does not provide model interpretation, which poses constraints for stakeholders to utilize DL for LULC applications.

Ebenezer and Manohar [33] proposed a framework based on a deep convolutional spiking neural network and enhanced Elman spike neural network for LULC classification. The proposed method is implemented on Indian remote sensing (IRS) Satellite Resourcesat-1 LISS-III with Cartosat-1 digital elevation model (DEM) satellite imagery and classifies the features as

forest and nonforest. The obtained classification map has certain misclassification errors and noise. The authors have utilized markov chain random field (MCRF) cosimulation approach to handle aforementioned issue. The proposed model has shown the 97.8% accuracy, 97.9 % F1-score and 0.89 kappa coefficient value. The performance of this method can be enhanced by incorporating spatiotemporal variables.

Albarakati et al. [34] proposed network-level fusion of self-attention architecture for agriculture LULC classification. In this study, authors have used a contrast enhancement technique to increase the number of training samples. The proposed model is implemented on EuroSAT dataset and remote sensing image scene classification 45 (RESISC45) dataset developed by northwestern polytechnical university (NWPU) and has shown remarkable performances in terms of accuracy and computational time. The proposed method has shown accuracy of 98.2% on SIRI-WHU dataset. Though the method is good, it has certain limitations. The proposed architecture needs huge amount of labeled data. The utilization of vision transformer-based architecture may improve the results.

Fayaz et al. [35] performed land-area scene classification (SCL) by leveraging the potential of transfer learning. The authors utilized three pretrained DL models, such as InceptionV3, DenseNet121, and ResNet50. The InceptionV3 emerges as the top performer with an OA of 92%. The proposed method provides effective utilization of limited labelled data for fine tuning. The model's performance could be enhanced by implementing an improvised version of CNN method for LULC classification of urban features. Dastour and Hassan [36] evaluated the performance of 39 deep transfer learning models for LULC classification. The experiments are conducted on EuroSAT dataset which has limited data samples for 10 classes. It has been observed that models such as ResNet50, EfficientNetV2B0 performed better than others. ResNet50 achieved an accuracy of 96.7%. Though authors have tested 39 methods, model interpretations still remain a challenge. In another study, Arrechea-castillo et al. [37] proposed a LeNet based CNN model for LULC classification. Authors have created feature set based on Sentinel-2 images, the spectral indices and DEM, and achieved an OA of 96.51% on validation dataset. This method outperformed the traditional ML methods and DL methods like ResNet, DenseNet, and EfficientNet. Though the representative samples of forest are available in the dataset, it has shown some misclassification for dense forest and mixed forest. Further, this method may miss some cloud and cloud shadows as the scene classification map SCL used in this study is not perfect. Few researchers have explored 'weakly towards a strongly supervised learning approach' for LULC classification. For instance, Zhang et al. [38], proposed DL framework weakly towards strongly (WTS) to deal with the issue of insufficiency of pixel-level annotations in semantic segmentation models. Though the proposed model has provided promising results, it could be experimented to increase the effectivity of different bands of remote sensing images.

Dewangkoro and Arymurthy [39] implemented VGG19 with channel squeeze and spatial excitation block for feature extraction and used twin SVM as a classifier. This hybrid approach led to better results with 94.57% OA on the EuroSAT dataset. In this study, only one dataset is explored with red, green, and blue (RGB) channels. The model is not tested on diverse datasets with multiple bands, ie., more than three channels instead of just three channels of RGB images. The Yang et al. [40] employed deep CNN and multi-scale feature fusion to improve the generalization capability of the DL model in the case of limited datasets. The multiscale feature fusion is utilized to improve the representation ability of classification features. In this study, only mono sensor data is utilized. This may limit the performance of the model. Most of the studies have utilized the EuroSAT dataset with RGB joint photographic experts group (JPEG) images, while few researchers utilized 13 band spectral dataset with RGB bands or other bands. For instance, Sonune [41] implemented ResNet50 on the RGB band dataset and achieved an accuracy of 94.25% dataset. This approach finds optimal hyperparameters based on trial and error and not on the basis of hyperparameter optimization technique. While most of the researchers have utilized the EuroSAT dataset with either RGB, JPEG images or tag image file format (TIFF) dataset with limited spectral bands, Yassine et al. [42] explored a 13-band dataset with the computation of spectral indices. Spectral indices are the mathematical equations that are applied to satellite image bands. The authors have computed spectral indices such as moisture index, normalized difference snow index, norgreen normalized difference vegetation index, blue normalized difference vegetation index, normalized difference red edge, bare soil index (BSI), normalized difference

chlorophyll index, normalized burn ratio, disease-water stress index and chlorophyll vegetation index for prominently identifying the features. Further, the authors have examined the transfer learning approach by applying the pre-trained deep neural network DenseNet201 and achieved an OA of 98.78% on the RGB band dataset and an accuracy of 99.58% on all band datasets. Though the method has shown promising results on the EuroSAT dataset, yet the transfer learning approach is used for experimental work and its generalization capability is not much tested.

Many a times, DL models lead to overfitting due to a greater number of features selected by CNN. Hence, it is essential to select appropriate features. Vinaykumar et al. [43] has handled the aforementioned issue by adopting the optimal guidance-whale optimization algorithm (OG-WOA) technique to select the relevant features and further used Bi-directional long short-term memory for classification. The model has achieved promising results on various datasets and achieved a maximum accuracy of 97.02%. The generalization performance can be further improved by incorporating an attention mechanism. In a few studies, researchers have utilized hyperspectral images for LULC classification. For instance, Tejasree and Agilandeeswari [44], adapted autoencoder for feature extraction, ranking-based band selection model for band selection, and finally deep long short-term memory (LSTM) for LULC classification from hyperspectral images. The authors have achieved remarkable results with an accuracy of 96.72%. But that proposed model is not tested on real images and its time complexity is more than desired.

Hosseiny et al. [45] proposed WetNet model by incorporating a deep ensemble model with time series Sentinel-1 SAR data and Sentinel-2 multispectral imagery. The two-dimensional (2D) CNN and three dimensional (3D) CNN model is used to extract nonlinear relationship and deep features from Sentinel-2 and Sentinel-1 satellite imagery. Further, LSTM layer is utilized extract multitemporal features from multisensory data. The WeNet model proposed by the authors outperformed the state of art methods InceptionResnetV2, InceptionV3, and DenseNet121with an OA of 0.9581 and kappa coefficient of 0.9503. This model utilizes ensemble learning approach which may be resource extensive. In another study, Yao et al. [46] proposed a multi modal DL framework extended vision transformer (ViT) for LULC classification. The extended ViT achieved the OA of 90.62% and outperformed the models like Co-ViT, Co-CNN and Co-SF. The interpretability of the proposed model still remains a challenge.

Even though a lot of work has been done for LULC classification using DL techniques on satellite imagery, and promising results have been achieved, many challenges remain. Hyperparameter optimization, multimodal approaches, model complexity, limited dataset issues, mixed pixel problems, and model interpretation still require special attention.

3.Materials and methods

The study of LULC is carried out at various stages such as materials and data preprocessing, model construction, model evaluation, model prediction, and LULC map generation. The technical route is depicted in *Figure 1* and stages of process flow are discussed in this section.

3.1Materials

Several LULC datasets are available for researchers to assess the performance of state of art DL algorithms. This study utilizes the publicly available EuroSAT dataset [41]. The preprocessing steps have been already mentioned on this dataset to handle image distortions as well as the resampling of bands. The dataset provides Sentinel-2 satellite images covering 13 spectral bands and consisting of 10 different scene classes with 27000 labeled and geo-referenced sample image patches. Nearly 2000 to 3000 images are available in each class of this dataset. The image patches comprise of 64×64 pixels with a spatial resolution of 10 meters. The dataset is available in two formats, such as RGB and all band TIFF dataset.

Most of the studies used in the literature have employed RGB datasets. This study utilizes all band TIFF datasets that provides 13 spectral reflectance band data. The red (B4), green (B3), blue (B2), visible and near infrared (B8), and short-wave infrared (B11) bands from available Sentinel-2 13 bands are utilized for the creation of ‘spectral indices derived dataset’. In optical remote sensing, spectral indices are used to prominently identify the features from satellite imagery. The multi-band method applies mathematical equations on the spectral reflectance band of satellite images to calculate the spectral indices. Several spectral indices are formulated by researchers to map various features in satellite images. The study area chosen for this study

includes various land covers such as built-up, water bodies, plantation, crops, and others. The spectral indices which help to identify these features prominently are computed. These spectral indices are discussed below.

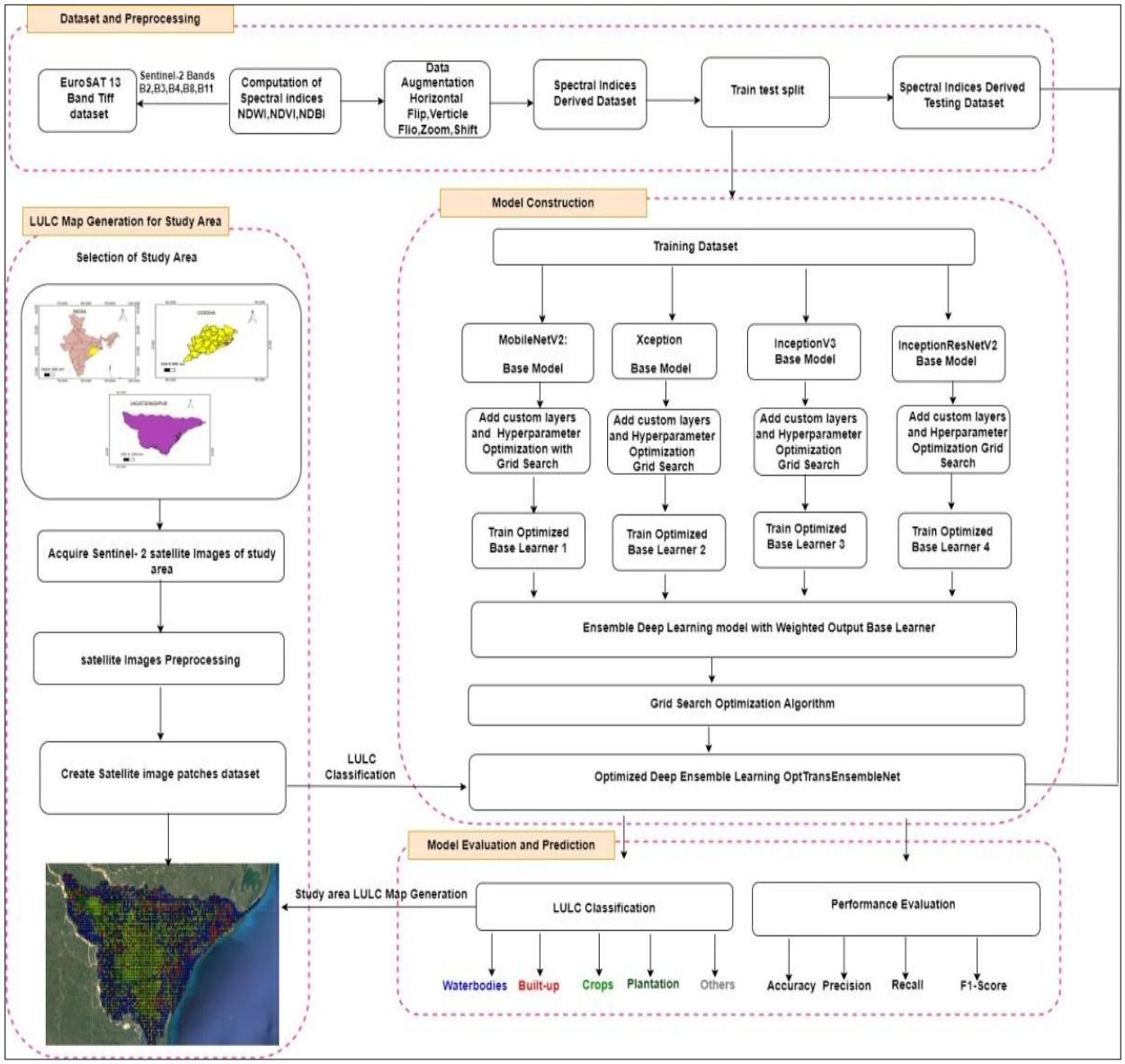

Figure 1 Block diagram of deep ensemble model

3.2Computation of spectral water indices

NDWI

NDWI is computed for delineating the water feature from satellite imagery. It was proposed by McFeeters in 1996 [47]. NDWI for Sentinel-2 is computed using Equation 1. Generally, NDWI index value greater than 0.5 corresponds to a water feature.

$$NDWI = \frac{(B03-B08)}{(B03+B08)} \quad (1)$$

Where, $B03$ represents the green band and $B08$ represents NIR band.

NDVI

NDVI is a simple but effective index. It is generally used for quantifying vegetation. In this study, the NDVI index is computed to delineate the crop and plantation features. NDVI for Sentinel-2 is computed with the formula given in Equation 2. NDVI value ranges from -1 to +1.

$$NDVI = \frac{(B04-B08)}{(B04+B08)} \quad (2)$$

Where, B04 represents red band and B08 represents NIR band.

3. Normalized Difference Built-up index (NDBI)
NDBI index was developed by Zha et al. in 2003 [48]. It is utilized for delineating the built-up and urban features. NDBI for Sentinel-2 is computed with the formula given in Equation 3.

$$NDBI = \frac{(B11-B08)}{(B11+B08)} \quad (3)$$

Where $B11$ represents SWIR band and $B08$ represents NIR band.

Bare soil index (BSI):
BSI is formulated by Rikimaru et al. [49]. BSI can efficiently identify the barren soil feature on Earth surface. In this study, BSI is computed to identify the barren land in satellite imagery of the study location. BSI for Sentinel-2 is computed with the formula given in Equation 4.

$$BSI = \frac{(B11+B4)-(B8+B2)}{(B11+B4)+(B8+B2)} \quad (4)$$

Where, B11 represents SWIR band, B08 represents NIR band, B04 represents red band and B2 represents blue band.

3.3Data augmentation and data split

DL needs a huge amount of data to achieve significant results. Data augmentation is applied to increase the dataset size. Image data augmentation in DL refers to a category of data augmentation that creates the transformed version of images in training data. In this study, augmentation is performed on images by applying operations such as horizontal flip, vertical flip, zoom, shift, etc. The distribution of RGB and spectral indices derived dataset is shown in *Table 1*. Further, the dataset is separated into three portions as training set, validation set and test set. A total of 80% data is utilized as training set, 10% data is utilized as validation set, and finally 10% data is used as test set.

Table 1 EuroSAT Dataset: RGB dataset, spectral indices derived dataset details utilized for LULC

LULC class	Spectral indices employed	Number of samples utilized from EuroSAT RGB dataset	Number of samples utilized from spectral indices derived dataset
Waterbodies	NDWI	1000	1000
Built-up	NDBI	1000	1000
Crop	NDVI	1000	1000
Plantation/ Trees	NDVI	1000	1000
Other	BSI	1000	1000
Total		5000	5000

3.4Methodology

The proposed methodology has been utilized in an integrated approach of optimization, transfer learning, and DL ensemble learning to propose an optimized deep ensemble learning model for LULC classification. The methodology section is subdivided into the selection of DL base learners, model construction, model optimization, and building of optimized deep ensemble learners. The pseudo code for the proposed model is depicted in algorithm 1 and steps have been discussed subsequently.

Let dataset SID_DS= (Xi, Yi) where X represents the number of images, Y represents the class category Y= {Built-up, Crop, Other, Plantation, Waterbodies} and SID_DS represents spectral indices derived dataset. Let M = {MobileNetV2, InceptionResNetV2, Xception, InceptionV3} be the set of pre-trained DL models. Let Wi where i=1,…,K be the weight for Mpred.

The detailed algorithm is discussed below.
OptTranEnsembleNet: Pseudocode for proposed optimized deep ensemble learning model.
Input: Spectral indices derived EuroSAT images (X, Y) from dataset SID_DS
Output: Output Class YEnsemble : OCYEnsemble € Y
BEGIN
Step 1: Divide dataset SID_DS into SID_DSTrain ,SID_DSVal and SID_DSTest
Step 2: Resize input images as per pre-trained model architecture
Step 3: Train base learners (BL) with individual DL models M
for i=1,2, 3…….,m
do
Define parameter grid for hyperparameters
Optimize hyperparameters of BLi with Grid search optimization
BLi=Yi (SID_DSTrain)

end for

Step 4: Compute weighted average ensemble for LULC

for i=1,2,……,m

do

Compute prediction Yk for each BLi

Apply grid search to find optimal values of weights for each model BLi

Assign weights to Wi to Mi predictions,

Multiply Wi with Mi predictions,

Compute weighted sum:

weightedsum += Pi × Wi

end for

OptTransEnsembleoutput=Compute optimized weighted average

Step 5: Compute performance evaluation metric

END

Technical route of proposed methodology

1. Selection of base learner:

Ensemble learning refers to combining multiple models to improve the generalization ability of ML models. Choosing the base learners that are most suited for the specific problem is crucial in the deep ensemble learning model. DL needs an enormous amount of data for remarkable performance. Remote sensing needs ground truth data. The process of data collection is time-consuming and expensive. It is challenging to acquire such huge ground truth data through field surveys. Transfer learning is the preferable solution in such a scenario. In this method, a model trained for one task can be used for another slightly similar task. As a result, base learners are chosen using the deep transfer learning approach. Various pre trained models like VGG16, ResNet50, ResNet50V2, ResNet101, ResNet101V2, InceptionV3, VGG19, InceptionResNetV2, MobileNet, DenseNet121, DenseNet201, EfficientNetB7, and MobileNetV2 are experimented and finally best performing state of art pre trained CNN architectures such as InceptionV3, MobileNetV2, Xception, and InceptionResNetV2 are employed as base learners for LULC classification. The pretrained models in base learners are used as feature extractors and 3 custom layers are added for the specific task of LULC classification. The pretrained models consider the weights of ImageNet. The custom layers comprise of dense layer followed by a batch normalization layer and a dropout layer. Lastly, an output layer is added with 5 neurons to represent the land cover classes. The hyperparameters such as a number of input neurons, activation function, and dropout rate are used for these layers. The values of hyperparameters directly impact the performance of the DL model. The hyperparameters of the custom layers are optimized with a grid search optimization algorithm to improve the accuracy of the proposed model. For grid search optimization, initial parameters and their respective values are defined through parameter grid. A grid search is performed for each combination of hyperparameters in the parameter grid, and optimal hyperparameters are selected. The optimal parameters obtained for custom layers are 512 number of neurons, rectified linear unit (ReLU) activation function for custom layers, softmax activation function for output layer, and dropout rate of 0.4. The grid search is tuned for 10 maximum trials and for each trial, 10 epochs is considered. The hyperparameter values provided by top 1 best model are selected finally for model training of each base learner. *Table 2* represents the grid search-parameter grid and optimal values of hyperparameters of the model architecture.

Table 2 Hyperparameters of grid search

Name of the hyperparameter	Parameter grid for grid search	Optimal hyperparameter value
Optimizer	Stochastic gradient descent (SGD), Adam, Adagrad, RMSprop	Adam
Hidden input neuron	Min 32 till Max 512	512
Activation function hidden layer	SoftMax, ReLU, Sigmoid	Relu
Activation function output layer	SoftMax, ReLU, Sigmoid	SoftMax
Dropout rate	min_value: 0.0, max_value: 0.6, step: 0.1	0.4
Learning rate	min_value: 0.00001, max_value: 0.01	0.0001
Epochs	Min_value:10, Max_value:50	25

Construction of optimized deep ensemble model:

All the base learner models are individually trained on the spectral indices derived EuroSAT dataset with optimized hyperparameters and their corresponding test accuracies are Accbl are recorded. The prediction Pm from the selected base learner is a probability vector acquired through the SoftMax activation function. Predictions from different optimized base learners are different. Simple averaging may affect the performance of the deep ensemble model as few

models outperform others. Hence, weights are assigned to each base learner for deep ensemble learning. Initially, random initialization of weights was done. But it did not give the appropriate results. Next, higher weights were assigned to best-performing models, whereas less weights were assigned to least-performing models. This strategy provided better results than random initialization. These results got further improved by identifying the optimal values of these weights. The appropriate weights Wi is computed through Grid search optimization. The parameter grid comprises of weight values from 0 to 1. Grid search algorithm systematically explores hyperparameter space. It also checks each combination of hyperparameters to find the optimal combination within specified range in contrast to random search algorithm, which randomly searches the hyperparameter space. The optimal weights Wi identified with grid search algorithm is multiplied with Pi to get a weighted sum and weighted average is computed likewise. The proposed deep ensemble model successfully classifies the land cover in accordance with the classes using weighted average ensemble strategy. This model outperforms the standalone models even on small datasets.

4.Results

4.1Experimental setup

All the experiments are conducted in Google Colab. The computational resources, like hardware accelerator T4GPU are used. Spectral indices computation as well as implementation of proposed model OptTransEnsembleNet is done in python using keras library.

4.2Performance evaluation of deep ensemble learning approaches

Performance evaluation of the DL model is essential to assess its applicability in real-time applications. The confusion matrix or error matrix is generally used to assess performance. The confusion matrix comprises 4 major fields such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The performance is evaluated based on accuracy, precision, recall, and F1-score metrics. The formula for different metrics is as mentioned in Equation 5 to Equation 8.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1-Score = \frac{2\ X\ Precision\ X\ Recall}{Precision+Recall} \quad (8)$$

Performance evaluation of spectral indices-based TIFF dataset

Spectral indices play a vital role in prominently identifying the features on the earth's surface. In this study, spectral indices like NDWI, NDVI, NDBI and BSI are computed to highlight various landscape features (land cover classes) like waterbodies, built-up, plantation, crop, and others. Various approaches, like the optimization of hyperparameters of the transfer learning model, a weighted average of the deep ensemble learning model, and an integrated approach where hyperparameters of base learners and weights of weighted average deep ensemble learning model are optimized with grid search optimization algorithm. *Figure 2* shows the accuracy graph of training accuracy and validation accuracy. *Figure 3* highlights the performance evaluation analysis.

In the *Figure 2*, Model 1 represents MobileNetV2, Model 2 represents Xception, Model3 represents InceptionV3, Model4 represents InceptionResNetV2 and Lastly, Model5 represents the proposed model OptTransEnsembleNet. It is observed that proposed model outperforms others.

Evaluation of proposed algorithm on EuroSAT RGB dataset

EuroSAT dataset is available in two formats RGB dataset and TIFF dataset. RGB dataset consists of JPEG images, whereas the TIFF dataset is available in 13 bands. Performance evaluation of the proposed algorithm is done on RGB and spectral indices derived dataset. It is observed that features are more prominently exhibited in contrast to RGB images. *Figure 4* depicts the accuracy achieved on spectral indices derived EuroSAT dataset and RGB-based EuroSAT dataset. The experimental results show that the spectral indices derived dataset used with deep ensemble learning outperforms the RGB dataset. *Figure 4* depicts the performance evaluation based on the accuracy of various DL models.

The spectral indices like NDWI, NDVI, NDBI and BSI have played a significant role in identifying the features. The NDWI has prominently identified the water features, whereas the NDVI has helped in identifying the crops and plantations. The NDBI index has played significant role in identifying the built-up area. BSI index has identified barren land. It is observed that DL models when applied on the dataset generated with these indices has shown better results as compared to RGB dataset.

Comparison with other research studies

Researchers have utilized the EuroSAT dataset for LULC classification using various DL approaches. It is observed that most of the studies are based on EuroSAT RGB datasets. Few researchers have also implemented DL method on the EuroSAT TIFF dataset. *Table 3* shows a comparison of the results of EuroSAT dataset. *Table 3* shows that proposed model has shown significant performance on all band dataset. *Figure 5* is the performance evaluation deep ensemble learning model on EuroSAT RGB dataset.

Performance evaluation of proposed model on study area (Jagatsighpur district):

The proposed model OptTransEnsembleNet is evaluated on the study area to assess its generalization ability. The study area Jagatsighpur is a district in Odisha state of India. It is located at 20.2395121 Latitude and 86.1615742 Longitude. The area has various land covers such as built-up, crops, plantation, waterbodies and others. The proposed OptTransEnsembleNet is evaluated in the study area. Initially, Sentinel-2 satellite images of the study area are acquired from ESA Copernicus.

The images are acquired for July 2023. The satellite images are huge in size and complex in nature. The satellite image of the study area is clipped to reduce the processing time. It provides 13 band data with 10-50m resolution. One pixel in a satellite image is equivalent to 10m on Earth's surface. The entire satellite image cannot be provided as an input to the DL model.

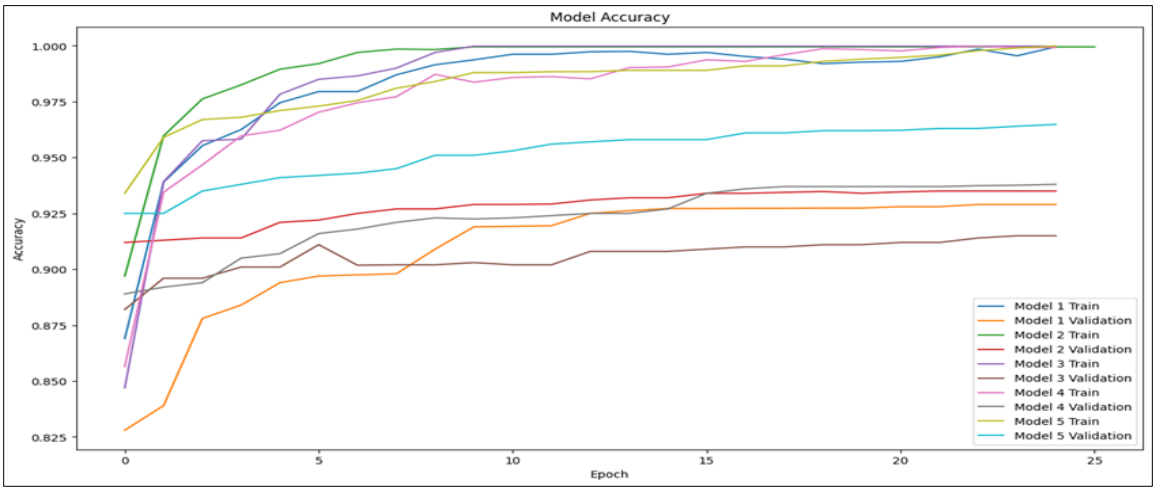

Figure 2 Training and validation accuracy plots

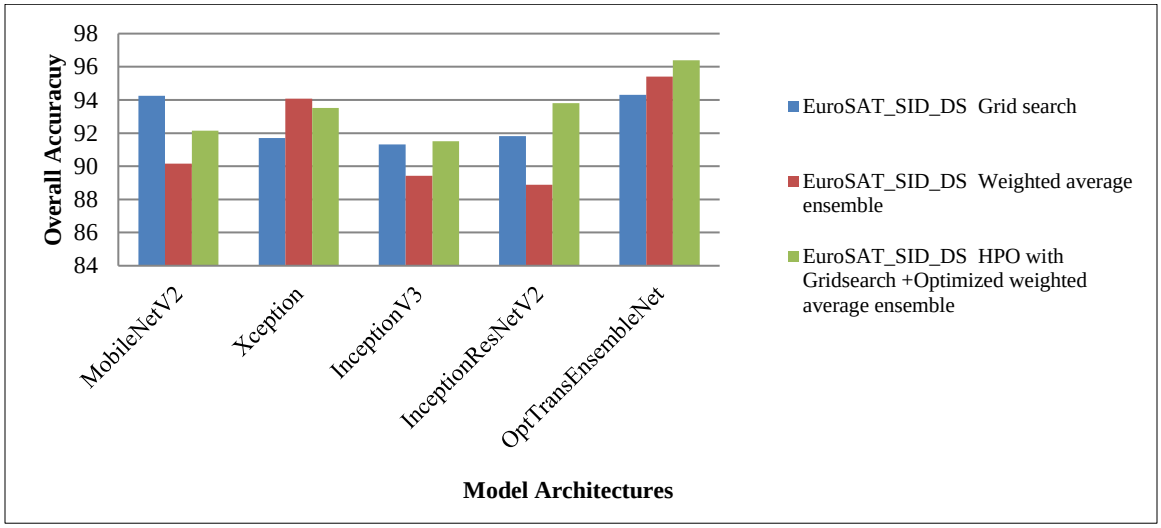

Figure 3 Comparison of OptTransEnsembleNet with other architectures

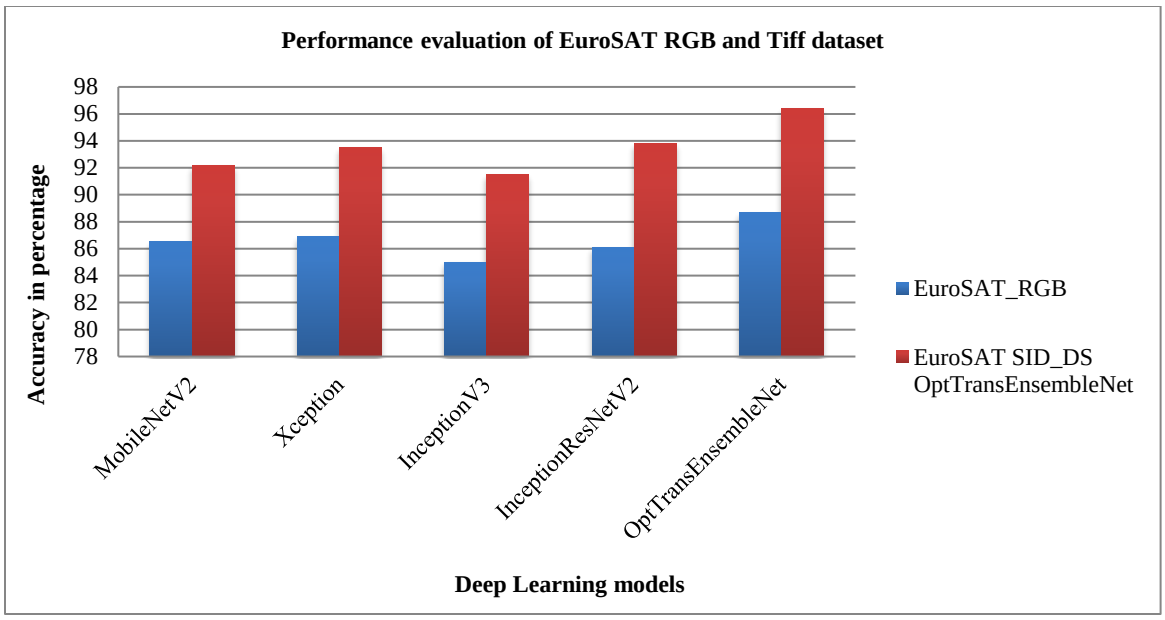

Figure 4 Performance evaluation based on accuracy of various DL models

Table 3 Performance evaluation of model architectures accuracies on EuroSAT dataset

Authors	DL model	Sentinel-2 bands	Accuracy (%)
Helber et al. [32]	ResNet50	SWIR	97.05
Sonune [41]	VGG19	RGB	97.66
Sonune [41]	ResNet50	RGB	94.25
Chong [50]	VGG16	RGB	94.5
Yamashkin et al. [51]	Shallow convolutional neural network	RGB	88.66
Temenos et al. [52]	Convolutional neural network	red, green, NIR, SWIR	94.72
Ours	OptTransEnsembleNet	red, green, blue, NIR,SWIR	**96.40**

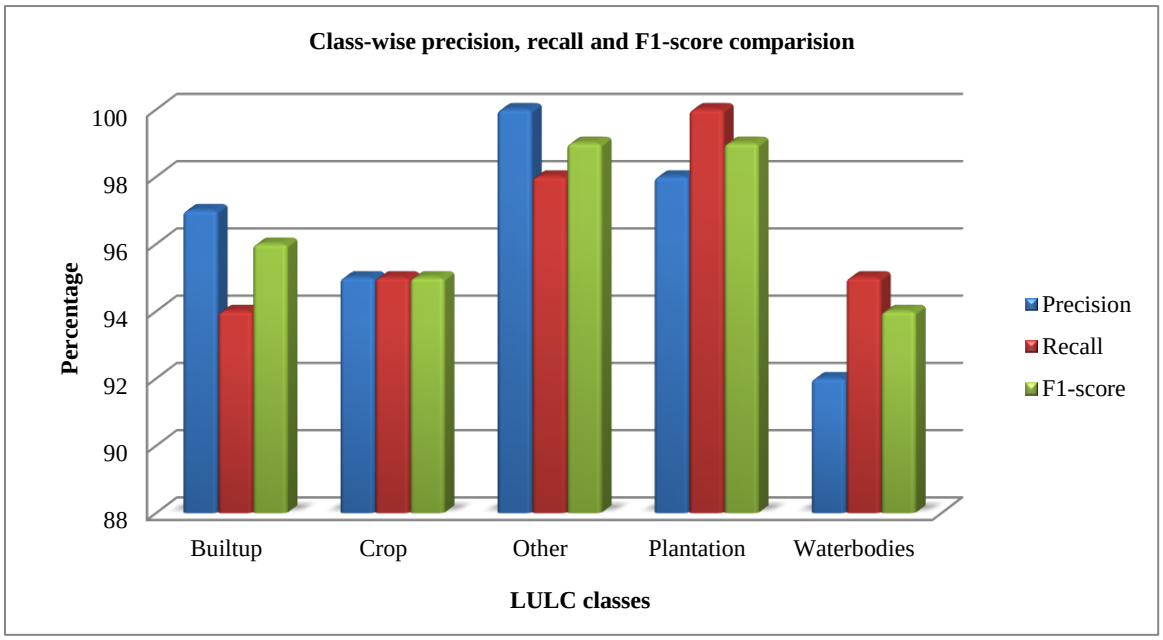

Figure 5 Class-wise precision, recall, and F1-score comparison for proposed model

Hence, satellite image is divided into small tiles of 8×8. These images/ tiles are provided as an input to the proposed OptTransEnsembleNet DL model to perform LULC of study area. *Figure 6* and *Figure 7* depict the LULC map of the study area generated through the proposed model. The built-up area and water bodies are depicted by red and blue colors respectively. The crop feature is represented in light green color, whereas plantation features is in dark green color and the other classes in gray color.

Figure 6 Sentinel 2 satellite image of the study area

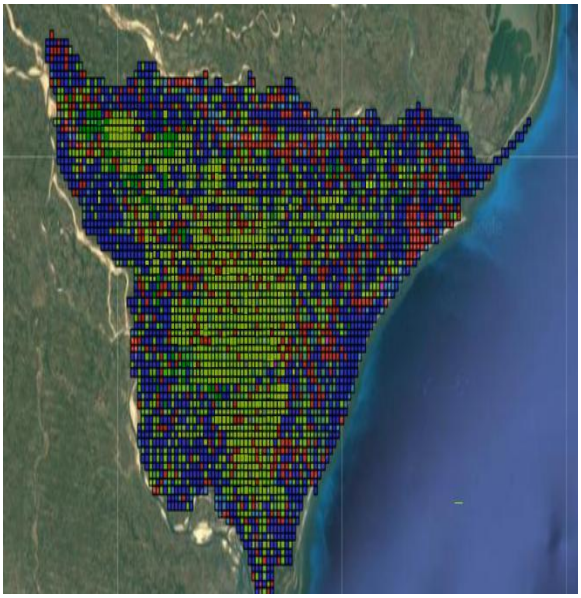

Figure 7 LULC of Jagatsinghpur district using Sentinel-2 and DL-OptTransEnsembleNet

5.Discussion

In this study, hybrid DL model OptTransEnsembleNet is presented by leveraging the potential of transfer learning, optimization, and deep ensemble learning techniques for LULC classification. Initially, experiments are conducted on the EuroSAT spectral indices derived dataset. The quantitative experimental results obtained with OptTransEnsembleNet have also been compared with those obtained from other state-of-the-art complex DL architectures, namely, MobileNetV2, Xception, InceptionV3 and InceptionResNetV2. The experiments are carried out on same dataset and with same computational resources to ensure fair comparison. The optimal hyperparameters for these models are identified with grid search optimization algorithm. It is observed that the InceptionResNetV2 outperformed the other models with 93.80% accuracy. The depth wise separable convolution of MobileNetV2 model assisted to build lightweight model with 92.62% accuracy. Though this model performed well, it is unable to capture diverse and hierarchical feature representations. In InceptionResNetV2, the inception module helps to capture multi scale features and residual connections helps to deal with vanishing gradient problem. The InceptionResNetV2 has shown better performance in comparison to MobileNetV2. The complex architecture of InceptionResNetV2 captures diverse features, resulting in better accuracy as compared with MobileNetV2. The InceptionV3 model has shown the least performance of 91.50%. Though InceptionV3 is a powerful architecture, lack of residual connections makes training deep network challenging. The combination of Inception module and residual module in InceptionResNetV2 resulted in efficient parameter utilization as compared to InceptionV3. The Xception module utilizes depth wise separable convolution, which results in reduced computation. However, it has not shown better generalization ability as InceptionResNetV2. Additionally, ensemble learning approach is explored. Several strategies adopted show different accuracy in different scenarios. The ensemble model created with simple weighted average has shown accuracy of 95.40% whereas the ensemble model created with simple average ensemble has shown least performance with 94.30% accuracy. The range of weight considered for weighted average is 0 to 1. In this study, the optimal hyperparameters obtained through grid search optimization algorithm are utilized for weighted average ensemble. A state-of-the-art accuracy of 96.45% was achieved by using OptTransEnsembleNet. The computational cost of

proposed model is higher as compared to individual learners. The utilization of potential of various base learners for ensemble model and the two-stage optimization has enhanced the performance of OptTransEnsembleNet. The proposed model is tested on EuroSAT RGB dataset and EuroSAT all band spectral indices derived dataset. It is observed that proposed model performs better on spectral indices derived dataset as compared to EuroSAT RGB dataset.

All the state of art models are compared on the basis of computational training time, parameters and accuracy. Depending upon the experimental results, it is evident that OptTransEnsembleNet obtained better accuracy, but the time taken to training the model is more in comparison to base learners. The number of trainable parameters is more in Xception, InceptionV3 and OptTransEnsembleNet, whereas less trainable parameters are required for MobileNetV2 and InceptionResNetV2. The proposed model provided better accuracy. However, computational training time and number of trainable parameters are greater. The comparison has been extended to additional studies conducted on the same datasets, different bands, and experimental settings, in an effort to increase the comprehensiveness of the research. This comparison is presented in *Table 3*. The proposed model has shown better performance in comparison to VGG16, ResNet50, CNN and shallow CNN. However, performance of DenseNet201 is found better than proposed model on all band dataset. Further, class wise comparative analysis is done based on precision, recall and F1-score evaluation metrics. While other landcover class has shown highest precision followed by built up and plantation class, water bodies class has shown the least precision value. Comparative analysis on the basis of F1-score shows that plantation class has shown highest F1-score as compared to remaining classes. The waterbodies class has exhibited the lowest F1-score. On the other hand, highest recall value is observed for plantation class whereas lowest value is observed for built-up class. In this study, the LULC classification of Jagatsinghpur district in Odisha state of India is performed by employing optimized deep ensemble learning approach. The proposed model has efficiently generated the LULC map of study region. The generated map is validated with ESA world cover data map (https://viewer.esa-worldcover.org/worldcover). Though the proposed model has efficiently generated the LULC map of study region, it has some limitations associated with it. The built up, plantation and crop features are identified correctly whereas certain misclassifications are observed in water bodies class. Due to mixed pixel problem in study area, few discrepancies are also observed in the plantation and built-up area. The spectral indices derived dataset created in this research work is limited to only 5000 data samples. This may not accurately reflect the variability present in the larger population and impacts model's capacity to generalize across different regions and datasets. This study utilizes optical remote sensing data from Sentinel-2. It has certain limitations of cloud cover. Sentinel-1 based SAR satellite imagery that is available in all weather conditions and day-night can be utilized for better performance. Further, the evaluation of the model shows that there are certain misclassifications of the earth observation features classes. Specifically for the waterbodies class. Few misclassifications are also observed in plantation and built-up class in case of mixed pixels. The model can further be improved by reducing the size of the model for resource constrained environment. The model can be tested on diverse benchmark datasets.

A complete list of abbreviations is listed in *Appendix I.*

6.Conclusion and future work

DL has shown promising results in remote sensing applications. However, inappropriate selection of hyperparameters and small dataset size significantly affect its performance. An optimized ensemble learning model called OptTransEnsembleNet was proposed to perform LULC classification on the EuroSAT TIFF dataset. Initially, RGB, NIR, and SWIR bands are used to compute spectral indices such as NDWI, NDVI, NDBI, BSI, and spectral indices-derived dataset is created from the EuroSAT dataset. The proposed optimized ensemble learning model consists of four base learners as pre-trained models with added custom layers: MobileNetV2, InceptionV3, Xception, and InceptionResNetV2. The hyperparameters are optimized with the grid search optimization algorithm. The performance of the ensemble model is evaluated using both averaging and weight-averaging methods. Alternatively, a weighted average ensemble assigns higher weights to models with greater accuracy of 96.40%, which is higher than the model averaging approach. Therefore, optimized weighted average ensemble models perform much better than the individual pre-trained models for LULC classification. The proposed DL model is evaluated on the study area and a LULC map is generated. Accurate LULC mapping plays crucial role in the ecosystem. The LULC maps

generated through proposed model will assist policy makers and administrators for urban planning and land resource management. Generally, DL model is considered as black box model. This poses limitations on its usage in operational systems as stockholders are keen to understand why the model is giving particular output. In the future, Explainable AI-based model-agnostic methods, such as SHapley Additive exPlanations (SHAP) and local interpretable model-agnostic explanations (LIME), are aimed to be used for LULC model interpretation.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

Data availability
The EuroSAT dataset utilized in this study is publicly available and can be found at https://github.com/phelber/eurosat.

Author's contribution statement
Kavita Devanand Bathe: Conceptualization, data collection, writing – original draft, analysis, and interpretation of results. **Nita Sanjay Patil:** Study conception, design, supervision, investigation of challenges, and draft manuscript verification.

References
[1] Wang SW, Gebru BM, Lamchin M, Kayastha RB, Lee WK. Land use and land cover change detection and prediction in the Kathmandu district of Nepal using remote sensing and GIS. Sustainability. 2020; 12(9):1-18.

[2] Langat PK, Kumar L, Koech R, Ghosh MK. Monitoring of land use/land-cover dynamics using remote sensing: a case of Tana River basin, Kenya. Geocarto International. 2021; 36(13):1470-88.

[3] Chen Z, Wang J. Land use and land cover change detection using satellite remote sensing techniques in the mountainous three gorges area, China. International Journal of Remote Sensing. 2010; 31(6):1519-42.

[4] Li W, Dong R, Fu H, Wang J, Yu L, Gong P. Integrating google earth imagery with landsat data to improve 30-m resolution land cover mapping. Remote Sensing of Environment. 2020; 237:111563.

[5] Bathe KD, Patil NS. Assessment of land use-land cover dynamics and its future projection through Google earth engine, machine learning and QGIS-MOLUSCE: a case study in Jagatsinghpur district, Odisha, India. Journal of Earth System Science. 2024; 133(2):111.

[6] Amalisana B, Hernina R. Land cover analysis by using pixel-based and object-based image classification method in Bogor. In conference series: earth and environmental science 2017 (pp. 1-7). IOP Publishing.

[7] Šandera J, Štych P. Selecting relevant biological variables derived from Sentinel-2 data for mapping changes from grassland to arable land using random forest classifier. Land. 2020; 9(11):1-20.

[8] Alshari EA, Abdulkareem MB, Gawali BW. Classification of land use/land cover using artificial intelligence (ANN-RF). Frontiers in Artificial Intelligence. 2023; 5:964279.

[9] Thanh NP, Kappas M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors. 2017; 18(1):1-20.

[10] Zhang S, Yang P, Xia J, Wang W, Cai W, Chen N, et al. Remote sensing inversion and prediction of land use land cover in the middle reaches of the Yangtze River basin, China. Environmental Science and Pollution Research. 2023; 30(16):46306-20.

[11] Loukika KN, Keesara VR, Sridhar V. Analysis of land use and land cover using machine learning algorithms on Google earth engine for Munneru River basin, India. Sustainability. 2021; 13(24):1-15.

[12] Sharma A, Gulati T. Change detection in remotely sensed images based on modified log ratio and fuzzy clustering. In information and communication technology for intelligent systems 2018 (pp. 412-9). Springer International Publishing.

[13] Hu Y, Raza A, Syed NR, Acharki S, Ray RL, Hussain S, et al. Land use/land cover change detection and NDVI estimation in Pakistan's Southern Punjab province. Sustainability. 2023; 15(4):1-21.

[14] Campos-taberner M, García-haro FJ, Martínez B, Izquierdo-verdiguier E, Atzberger C, Camps-valls G, et al. Understanding deep learning in land use classification based on Sentinel-2 time series. Scientific Reports. 2020; 10(1):17188.

[15] Bhakthan SM, Loganathan A. A hyperspectral unmixing model using convolutional vision transformer. Earth Science Informatics. 2024; 17(3):2255-73.

[16] Devanand BK, Patil NS. Leveraging potential of deep learning for remote sensing data: a review. Intelligent systems and human machine collaboration: select proceedings of ICISHMC 2022 (pp. 129-45). Springer.

[17] Chen H, Wu C, Du B, Zhang L, Wang L. Change detection in multisource VHR images via deep siamese convolutional multiple-layers recurrent neural network. IEEE Transactions on Geoscience and Remote Sensing. 2019; 58(4):2848-64.

[18] Naik N, Chandrasekaran K, Sundaram VM, Panneer P. Dual attention guided deep encoder-decoder network for change analysis in land use/land cover for Dakshina Kannada district, Karnataka, India. Environmental Earth Sciences. 2023; 82(1):33.

[19] Ansith S, Bini AA. Land use classification of high resolution remote sensing images using an encoder based modified GAN architecture. Displays. 2022; 74:102229.

[20] Clark A, Phinn S, Scarth P. Optimised U-Net for land use–land cover classification using aerial photography. PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science. 2023; 91(2):125-47.
[21] Aryal J, Sitaula C, Frery AC. Land use and land cover (LULC) performance modeling using machine learning algorithms: a case study of the city of Melbourne, Australia. Scientific Reports. 2023; 13(1):13510.
[22] Amini S, Saber M, Rabiei-dastjerdi H, Homayouni S. Urban land use and land cover change analysis using random forest classification of landsat time series. Remote Sensing. 2022; 14(11):1-23.
[23] Avcı C, Budak M, Yağmur N, Balçık F. Comparison between random forest and support vector machine algorithms for LULC classification. International Journal of Engineering and Geosciences. 2023; 8(1):1-10.
[24] Yuh YG, Tracz W, Matthews HD, Turner SE. Application of machine learning approaches for land cover monitoring in northern Cameroon. Ecological Informatics. 2023; 74:101955.
[25] Zhao Z, Islam F, Waseem LA, Tariq A, Nawaz M, Islam IU et al. Comparison of three machine learning algorithms using google earth engine for land use land cover classification. Rangeland Ecology & Management. 2024; 92:129-37.
[26] Alshari EA, Gawali BW. Development of classification system for LULC using remote sensing and GIS. Global Transitions Proceedings. 2021; 2(1):8-17.
[27] MohanRajan SN, Loganathan A, Manoharan P, Alenizi FA. Fuzzy swin transformer for land use/land cover change detection using LISS-III satellite data. Earth Science Informatics. 2024; 17(2):1745-64.
[28] Qi X, Peng J, Wang Y, Qi X, Peng Y. High-resolution land-cover mapping based on a cross-resolution deep learning framework and available low-resolution labels. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2023; 17:1839-56.
[29] Aljebreen M, Mengash HA, Alamgeer M, Alotaibi SS, Salama AS, Hamza MA. Land use and land cover classification using river formation dynamics algorithm with deep learning on remote sensing images. IEEE Access. 2024; 12:11147-56.
[30] Sawant S, Garg RD, Meshram V, Mistry S. Sen-2 LULC: land use land cover dataset for deep learning approaches. Data in Brief. 2023; 51:109724.
[31] Naushad R, Kaur T, Ghaderpour E. Deep transfer learning for land use and land cover classification: a comparative study. Sensors. 2021; 21(23):1-13.
[32] Helber P, Bischke B, Dengel A, Borth D. Eurosat: a novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2019; 12(7):2217-26.
[33] Ebenezer PA, Manohar S. Land use/land cover change classification and prediction using deep learning approaches. Signal, Image and Video Processing. 2024; 18(1):223-32.
[34] Albarakati HM, Khan MA, Hamza A, Khan F, Kraiem N, Jamel L, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2024; 17:6338-6353.
[35] Fayaz M, Nam J, Dang LM, Song HK, Moon H. Land-cover classification using deep learning with high-resolution remote-sensing imagery. Applied Sciences. 2024; 14(5):1-15.
[36] Dastour H, Hassan QK. A comparison of deep transfer learning methods for land use and land cover classification. Sustainability. 2023; 15(10):1-18.
[37] Arrechea-castillo DA, Solano-correa YT, Muñoz-ordóñez JF, Pencue-fierro EL, figueroa-casas A. Multiclass land use and land cover classification of Andean sub-basins in Colombia with sentinel-2 and deep learning. Remote Sensing. 2023; 15(10):1-20.
[38] Zhang W, Tang P, Corpetti T, Zhao L. WTS: a weakly towards strongly supervised learning framework for remote sensing land cover classification using segmentation models. Remote Sensing. 2021; 13(3):1-18.
[39] Dewangkoro HI, Arymurthy AM. Land use and land cover classification using CNN, SVM, and channel squeeze & spatial excitation block. In conference series: earth and environmental science 2021 (pp. 1-8). IOP Publishing.
[40] Yang Z, Mu XD, Zhao FA. Scene classification of remote sensing image based on deep network and multi-scale features fusion. Optik. 2018; 171:287-93.
[41] Sonune N. Land cover classification with EuroSAT dataset [Internet]. 2020.
[42] Yassine H, Tout K, Jaber M. Improving LULC classification from satellite imagery using deep learning–eurosat dataset. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2021; 43:369-76.
[43] Vinaykumar VN, Babu JA, Frnda J. Optimal guidance whale optimization algorithm and hybrid deep learning networks for land use land cover classification. EURASIP Journal on Advances in Signal Processing. 2023; 2023(1):13.
[44] Tejasree G, Agilandeeswari L. Land use/land cover (LULC) classification using deep-LSTM for hyperspectral images. The Egyptian Journal of Remote Sensing and Space Sciences. 2024; 27(1):52-68.
[45] Hosseiny B, Mahdianpari M, Brisco B, Mohammadimanesh F, Salehi B. WetNet: a spatial–temporal ensemble deep learning model for wetland classification using Sentinel-1 and Sentinel-2. IEEE Transactions on Geoscience and Remote Sensing. 2021; 60:1-4.
[46] Yao J, Zhang B, Li C, Hong D, Chanussot J. Extended vision transformer (ExViT) for land use and land cover classification: a multimodal deep learning

framework. IEEE Transactions on Geoscience and Remote Sensing. 2023; 61:1-5.
[47] Mcfeeters SK. The use of the normalized difference water index (NDWI) in the delineation of open water features. International Journal of Remote Sensing. 1996; 17(7):1425-32.
[48] Zha Y, Gao J, Ni S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. International Journal of Remote Sensing. 2003; 24(3):583-94.
[49] Rikimaru A, Roy PS, Miyatake S. Tropical forest cover density mapping. Tropical Ecology. 2002; 43(1):39-47.
[50] Chong E. EuroSAT land use and land cover classification using deep learning. 2020 [Internet].
[51] Yamashkin SA, Yamashkin AA, Zanozin VV, Radovanovic MM, Barmin AN. Improving the efficiency of deep learning methods in remote sensing data analysis: geosystem approach. IEEE Access. 2020; 8:179516-29.
[52] Temenos A, Temenos N, Kaselimi M, Doulamis A, Doulamis N. Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP. IEEE Geoscience and Remote Sensing Letters. 2023; 20:1-5.

Kavita Devanand Bathe is working as an Assistant Professor at K.J. Somaiya Institute of Technology, Mumbai. She is currently pursuing her PhD. She is actively involved in research and consultancy work. Her current research interests include Remote sensing, and Machine learning.
Email: kavitag@somaiya.edu

Nita Sanjay Patil is working as an Associate professor at K. C. College of Engineering and Management Studies and Research, Thane, India. She is actively involved in research and development. Her current research interests include Blockchain, and Machine Learning.
Email: nsp.cm.dmce@gmail.com

Appendix I

S. No.	Abbreviation	Description
1	2D	Two-Dimensional
2	3D	Three Dimensional
3	ANN	Artificial Neural Network
4	BSI	Bare Soil Index
5	CART	Computed and Classification and Regression Tree
6	CNN	Convolution Neural Network
7	DEM	Digital Elevation Model
8	DT	Decision Tree
9	DL	Deep Learning
10	FN	False Negative
11	FP	False Positive
12	ETM+	Enhanced Thematic Mapper Plus
13	GAN	Generative Adversarial Network
14	GEE	Google Earth Engine
15	GTB	Gradient Tree Boosting
16	ISO data	Iterative Self-Organizing Data
17	IRS	Indian Remote Sensing
18	JPEG	joint photographic experts group
19	KNN	K-Nearest Neighbour
20	LIME	Local Interpretable Model-Agnostic Explanations
21	LSTM	Long Short-Term Memory
22	LULC	Land Use Land Cover
23	MCRF	Markov Chain Random Field
24	ML	Machine Learning
25	MNDWI	Modified Normalized Difference Water Index
26	NDBI	Normalized Difference Built-Up Index
27	NDVI	Normalized Difference Vegetation Index
28	NDWI	Normalized Difference Water Index
29	NIR	Near Infrared
30	NWPU	Northwestern Polytechnical University
31	OA	Overall Accuracy
32	OLI	Operational Land Imager
33	OG-WOA	Optimal Guidance-Whale Optimization Algorithm
34	RBF	Radial Basis Function
35	RESISC45	Remote Sensing Image Scene Classification 45
36	ReLU	Rectified Linear Unit
37	RF	Random Forest
38	RGB	Red, Green, and Blue
39	RNN	Recurrent Neural Network
40	SAR	Synthetic Aperture Radar
41	SCL	Scene Classification
42	SGD	Stochastic Gradient Descent
43	SHAP	SHapley Additive exPlanations
44	SWIR	Short-Wave Infrared
45	SVM	Support Vector Machine
46	TIFF	Tag Image File Format
47	TN	True Negative
48	TP	True Positive
49	TSVM	Transudative Support Vector Machine
50	VGG16	Visual Geometry Group
51	ViT	Vision Transformer
52	WRNs	Wide Residual Networks
53	WTS	Weakly Towards Strongly