Received: November 20, 2024. Revised: December 19, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Long Short-Term Memory with Temporal Pyramid Pooling Layer for Land Use Land Cover Classification

Vinaykumar Vajjanakurike Nagaraju[1]* **Ananda Babu Jayachandra[1]**

[1]*Department of Information Science and Engineering, Malnad College of Engineering, Hassan, and Visvesvaraya Technological University, Belagavi, Karnataka, India*
* Corresponding author's Email: vinays.kumar08@gmail.com

Abstract: The Land Use Land Cover (LULC) classification from the Remote Sensing (RS) images is significant in various land use researchers. However, numerous methods have been developed for LULC classification which failed to capture the multiple deep temporal features. To mitigate this limitation, the Long Short-Term Memory (LSTM) with Temporal Pyramid Pooling (TPP) layer technique is proposed for LULC classification. The TPP layer is incorporated into the LSTM network to capture multiple temporal features and enhance its ability to differentiate the various classes of LULC. The ResNet-50 based feature extraction technique is developed to extract the deep meaningful features which help to differentiate the LULC classes. The performance of LSTM with the TPP layer technique is evaluated on EuroSAT, SIRI-WHU and UCM datasets. The proposed LSTM with TPP layer technique obtained 98.37% accuracy on the EuroSAT dataset, 98.77% accuracy on SIRI-WHU dataset, 99.75% accuracy on the UCM dataset and 99.35% accuracy on NWPU dataset when compared with existing algorithms like optimized self-attention fused Convolutional Neural Networks.

Keywords: Land use land cover, Long short-term memory, Remote sensing, ResNet-50, Temporal pyramid pooling.

1. Introduction

Land Use and Land Cover (LULC) classification from hyperspectral images has significance in different fields and applications, including natural resource management, environmental monitoring [1,2], developing infrastructure, managing disaster, urban planning, food security, conserving biodiversity and climate change adaptation [3]. The correct classification of land cover types enables the efficient management of natural resources [4]. That assisted in monitoring the forest cover, wetlands, water bodies and agricultural areas [5]. Land use refers to the process of land cover and employed to surface cover on earth including urban architecture, vegetation, water, and bare soil that does not define land use, and it varies for lands with the same cover format [6, 7]. The evaluation of LULC is required to monitor, plan and maintain the usage of natural resources [8]. The classification of LULC directly influences the atmosphere, water resources and soil erosion, but that has indirect implications for global landscape problems [9].

Recently, Deep Learning (DL) based algorithms have been used as key role in extracting high-level features and designed as the leading paradigm in the detection of patterns and Computer Vision (CV) [10, 11]. In DL-based algorithms, Convolutional Neural Networks (CNN) are generally used to identify complex patterns and extract the essential features from HSR RSI [12]. Additionally, the sequence of DL-based scene classification algorithms has been developed. Feature selection is the essential phase in pattern recognition and various algorithms have been implemented in recent times [13]. Feature selection algorithms majorly aim to minimize inappropriate data from actual feature subsets and reduce computational time. The significant contributions of the research are given as follows.

- The Long Short-Term Memory (LSTM) with Temporal Pyramid Pooling (TPP) layer is proposed to classify the different classes of LULC.

- The TPP layer is incorporated with the LSTM network to capture multiple temporal features, which helps to differentiate the different classes of LULC.
- The ResNet-50 based feature extraction technique is developed, which captures the deep meaningful features and helps to differentiate LULC classes.

This research paper is organized as follows: Section 2 summarizes the literature review of existing algorithms. Section 3 explains the process of the proposed methodology. Section 4 provides results and a comparison of the proposed methodology. The conclusion of this research paper is given in Section 5.

2. Literature review

In this section, the existing algorithms used for LULC classification are summarized with its advantages and limitations.

Vinay Kumar V.N [14] suggested the Optimal Guidance – Whale Optimization Algorithm (OG-WOA) for choosing appropriate features and minimizing the issue of overfitting. The optimal guidance method enhances the exploitation search process by adjusting the location of search agent relevance to optimal fitness value. This maximization in exploitation supports the choice of appropriate features and ignores the issue of overfitting. At last, chosen features were processed for classification by Bi-directional Long Short-Term Memory (Bi-LSTM). The suggested method improves the search ability in feature selection and helps to improve classification. However, the presented method doesn't capture the multiple temporal features for differentiating the LULC in classification.

Hussain Mobarak Albarakati [15] presented the fully automated optimized self-attention fused Convolutional Neural Network (CNN) structure for LULC classification. The contrast enhancement equation has been developed for data augmentation. Then, the fused self-attention CNN structure was developed. The presented method contains two custom methods such as IBNR65 and Densenet-64 and these methods were based on inverted bottleneck residual mechanism and dense blocks. Then, both methods were integrated by depth-wise concatenation and employed a self-attention layer to extract deep features. The presented method minimized execution time because the presented method selected the optimal features for classification. However, the CNN architecture needed a huge amount of labelled data to obtain reasonable results.

Saddaf Rubab [16] developed the network-level fusion deep structure dependent on 16-tiny Vision Transformer and SIBNet. Initially, data augmentation was performed to resolve the issue of data imbalance. The blocks were developed by utilizing the inception structure and every inception module was developed with blocks of bottleneck. The hyperparameters of the developed method were initialized by utilizing Bayesian Optimization for optimal training. The extracted features were classified by utilizing a neural network with multiple hidden layers. The developed method enhanced the classification accuracy and minimized execution time. However, the developed method failed to scale the pixel values in the image, minimising the performance of LULC.

Abhishek Bhatt and Vandana Thakur Bhatt [17] introduced the Deep-CNN (D-CNN) and ResNet-50 to extract relevant features from pre-processed data. The data were extracted, and the reduction of dimensionality was performed by assigning Principal Component Analysis (PCA) to rule out the count of inappropriate features. After assigning the PCA, the image classification was performed by employing the logistic weight updating hyperparameters tuning Random Forest (RF) technique that classified the extracted features. The introduced method extracted different features from images that help to reduce the dimensionality. However, the introduced method failed to extract the deep temporal features, which minimizes the classification performance.

Vijaykumar P. Yele [18] implemented the Auto-Metric Graph Neural Network to enhance the LU/LC (AMGNN-LU/LC) classification. The classification was performed with the support of EuroSAT dataset. The input image was enhanced with a pre-processing technique known as Anisotropic Diffusion Kuwahara Filtering (ADKF). After the pre-processing, the Hesitant Fuzzy Linguistic Bi-objective Clustering (HFLBC) method was used for segmentation. The Binary Coyote Optimization Algorithm (BCOA) was utilized to optimise the HFLBC method. The segmented image was classified as LU/LC with the support of AMGNN. The implemented method enhanced the visibility and quality of input images. However, the implemented method doesn't extract deep meaningful features for classification.

From the analysis of existing algorithms, these algorithms fail to capture multiple temporal features and deep meaningful features, struggle to scale the pixel values in images and require a large amount of labelled data to obtain reasonable results. These limitations minimize the classification performance and reduce the classification accuracy. To address these limitations, this article developed the LSTM with a TPP layer to effectively classify different classes in LULC. The pixel values in images of the dataset are scaled by using min-max normalization in the pre-processing phase. Then, the deep meaningful features are extracted by using the ResNet-50 technique and the LSTM network is incorporated with TPP layers that capture multiple temporal features and improve the classification performance.

3. Proposed methodology

The effective DL based algorithm is developed to classify the different classes of LULC. The datasets used in this article are EuroSAT, SIRI-WHU and UCM datasets. The input images in the dataset are pre-processed by using the min-max normalization technique. Then, the deep meaningful features are extracted by using the ResNet-50 technique and the features are classified by using LSTM with the TPP layer method. Fig. 1 represents the process of LULC classification to classify different classes.

3.1 Dataset

In this article, three datasets are used for classification such as EuroSAT, SIRI-WHU and UCM, the nature of these datasets is RGB. The detailed description of these datasets is explained as follows.

- EuroSAT dataset [19] - This dataset includes 10 classes, and every image has a resolution size of 64×64 pixels and a sampling distance of 10 m ground. Every image is collected by using Sentinel-2-satellite and there are a total 27,500 number of images.
- SIRI-WHU dataset [20] - This dataset has 12 classes and a total of 2400 images are there. Each class includes 200 images with a resolution size of 200×200 and includes a spatial resolution of 2m.
- UCM dataset [21] – This dataset has 21 images and a total of 2100 images. Every class includes 100 images with a resolution size of 256×256 and includes spatial resolution of 0.3. Below table 1 represents the dataset description, Fig. 2 represents the sample images in the EuroSAT dataset, Fig. 3 represents the sample images in the SIRI-WHU dataset and Fig. 4 represents the sample images in the UCM dataset.
- NWPU dataset [22] – This dataset has 31,500 images and in resolution size of 256×256. It has 45 classes and 700 images per class and spatial resolution of 0.2m. Fig. 5 represents the sample images in NWPU dataset.

3.2 Pre-processing

The images in the dataset are given as input to the pre-processing phase to enhance the quality of the image. In this article, the min-max normalization technique is used which scales the pixel intensity values within the range of 0 and 1. This process enhances the convergence rate in the training stage and ensures the method learns the data efficiently. The mathematical formula for min-max normalization is given in Eq. (1),

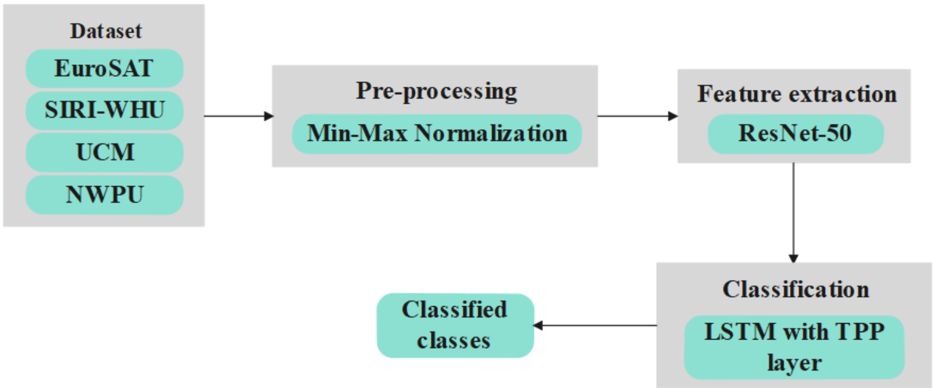

Figure. 1 Process of LULC classification

Table 1. Dataset Description

Datasets	Total images	Classes	Images per class	Resolution size	Spatial resolution
EuroSAT	27,500	10	2000 - 3000	64×64	10
SIRI-WHU	2400	12	200	200×200	2
UCM	2100	21	100	200×200	0.3
NWPU	31,500	45	700	256×256	0.2

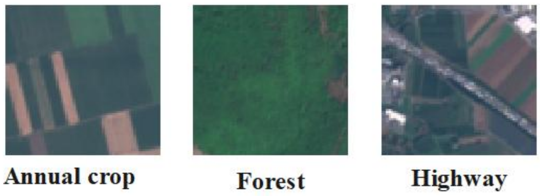

Figure. 2 Sample images in EuroSAT dataset

Figure. 3 Sample images in SIRI-WHU dataset

Figure. 4 Sample images in the UCM dataset

Figure. 5 Sample images in NWPU dataset

$$x' = \frac{x - min\,(x)}{max(x) - min\,(x)} \quad (1)$$

In the above Eq. (1), the $min\,(x)$ represents the minimum value, the $max(x)$ represents the maximum value, the x represents the actual image and the x' represents the normalized image.

3.3 Feature extraction

The pre-processed image is given as input to ResNet-50 to extract the meaningful features from the image which helps to differentiate the various classes of Land use and Land cover. The ResNet 50 method is one of the architectures in CNN that handles the issue of degradation which generally occurs in neural networks. The ResNet 50 includes residual blocks on various stacking convolution layers. The identity mapping in the residual block processes the shortcut connection. The mathematical formula for residual blocks on ResNet is given as Eq. (2),

$$y = F(x, \{W_i\}) + x \quad (2)$$

In the above Eq. (2), the W_i represents the weight of the convolution layer x and y represents the input and output vectors in the layer. The process $F(x, \{W_i\})$ is residual mapping or the result of stacked convolution layer. The essential features are extracted from the average pooling layer and given to the classification phase.

3.4 Classification

The extracted features are given as input to classification phase to classify different classes of land use and land cover. In this article, the Long Short-Term Memory (LSTM) is the kind of Recurrent Neural Network (RNN). By using the LSTM, the issue of sequence learning is resolved through incorporating the recurrent gates which link neurons to other over time. The input sequence is represented as $\{x_1, x_2, \dots, x_r\}$, the hidden state series is represented through $\{h_1, h_2, \dots, h_r\}$. The recurrent gates accommodated input as the x_t, given time is described t and their result value is represented as h_{t-1} in time of $t-1$, next by weighted result values and the mathematical formula is given as Eq. (3),

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b) \quad (3)$$

Here, the weight of the input feature and hidden gate are represented as W_{hx} weight of further time interval with hidden gate and it is represented as W_{hh} , b represents bias and the σ represented as non-linear activation function. However, there is an issue with the training of RNN there is difficulty in learning the long-term dependencies and temporal dependencies. To resolve this issue, the recurrent hidden gate is replaced with a memory cell which is in LSTM. The memory cell retains the feature by self-connected recurrent edge and determines the weight, the gradient is displaced due to various time steps without exploding or disappearing. The four essential components included in LSTM are input, forget, output gates and candidate cell values to compute memory cells and results by using the formula from Eq.s (4) to (9),

$$f_t = \sigma\left(W_{hf}h_{t-1} + W_{xf}x_t + b_f\right) \quad (4)$$

$$i_t = \sigma(W_{hi}h_{t-1} + W_{xi}x_i + b_i) \quad (5)$$

$$C_t \sim= \tanh(W_{hc}h_{t-1} + W_{xc}x_t + b_c) \quad (6)$$

$$C_t = f_c C_{t-1} + i_c \quad (7)$$

$$O_t = \sigma(W_{ho}h_{t-1} + W_{xo}x_t + b_o) \quad (8)$$

$$h_t = O_t \tanh(C_t) \quad (9)$$

In the above equations, the σ represents the sigmoid function, the b_f, b_i, b_c *and* b_o represents the bias term.

3.4.1.Temporal Pyramid Pooling

By using numerous pooling and convolutional layers, the feature map with size of $L \times 512$ 5 conv layers. The L represents the length of input than constant, obtaining the activations of variable-length relevance to inputs in the layer. In this article, the Temporal Pyramid Pooling (TPP) layer is used to learn features in fixed dimensions and extract the data from various temporal scales. The adaptive windows are utilized for the pooling process. For the pyramid phase with n bins, the window of max-pooling moves over the feature map by time, where the ith bin is relevant to the feature map in $\left(\left[\frac{i-1}{n}L\right], \left[\frac{i}{n}L\right]\right)$. The four pyramid phases $\{1,2,3,4\}$ are utilized in this article. After this process, concatenate the feature map with various phases and attain the fixed-dimensional vector for the next procedure. The TPP later converts the variable-length feature to the fixed-length result. Without the TPP, the neural network requires fixed-length inputs. To address this issue without altering the feature size, the TPP layer is employed to learn the fixed-dimensional representation. By using the TPP layer, it aggregates the data from various temporal scales and captures the multiple scale features in various classes of LULC. Adaptive global pooling is employed in the TPP layer as one of the pyramid levels. Compared with adaptive global pooling, multi-pyramid pooling keeps much data and identifies numerous pooling phases that help to enhance precision. The L_{in} and L_{out} represent the length of input and output sequences. Initially, sets the pyramid phase M and size of the pyramid is n_i in phase i. Next, the length of the output sequence is given in Eq. (10),

$$L_{out} = \sum_{i=1}^{M} n_i \quad (10)$$

For obtain the set of L_{out} , the TPP layer parameters are given in Eq. (11) and (12),

$$K_{TPP} = S_{TPP} = \left\lceil \frac{L_{in}}{n_i} \right\rceil \quad (11)$$

$$P_{TTP} = \left\lceil \frac{K_{TPP} \times n_i - L_{in} + 1}{2} \right\rceil \quad (12)$$

The above Eq. (10) to (12), the K_{TPP}, S_{TPP} and P_{TTP} represents the size of the kernel, stride and padding of Maxpooling process in the TPP layer. The input feature size is different, but the result representation remains in a similar size, which allows samples with different time window sizes to be fed in the training stage. Fig. 6 represents the process of the TPP layer.

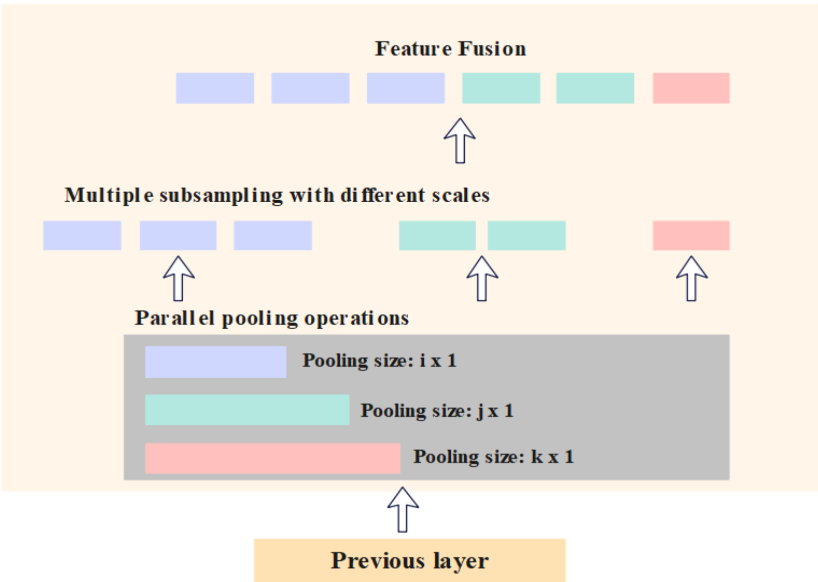

Figure. 6 Process of TPP layer

4. Experimental results

The performance of LSTM with TPP layer method is simulated with MATLAB 2020 b environment and required system configurations are an i5 processor, windows 10 (64 bit) and 8 GB RAM. The performance of the developed method is evaluated with metrics of recall, precision, accuracy, and f1-score. The mathematical formula for metrics is given from Eq. (13) to (16),

- Accuracy

The accuracy is referred to as the amount of correctly classified classes (TN+TP) to whole classes of the dataset (TP+TN+FP+FN). The mathematical formula for accuracy is given in Eq. (13),

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

- Precision

The precision is referred to as the ratio of appropriate recognized classes (TP) to whole classes which is accurately classified (TP+FP). The mathematical formula for precision is given in Eq. (14),

$$Precision = \frac{TP}{FP+TP} \tag{14}$$

- Recall

The recall estimated the whole number of correct positive classes in all positive classes. The mathematical formula for the recall is given in Eq. (15),

$$Recall = \frac{TP}{FN+TP} \tag{15}$$

- F1-score

The f1-score is referred to as the average value of recall and precision. The mathematical formula for the f1-score is given in Eq. (16),

$$F1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{16}$$

In the above Eq. (13-16), the TP, FP, TN and FN describe the True Positive, False Positive, True Negative and False Positive.

Table 2 evaluates the performance of the feature extraction technique using different metrics on three datasets like EuroSAT, SIRI-WHU and UCM. The developed ResNet-50 based feature extraction technique obtained 98.37% accuracy, 97.51% precision, 97.04% recall and 97.27% f1-score on EuroSAT dataset. The developed ResNet-50 based feature extraction technique obtained 98.77% accuracy, 98.41% precision, 98.28% recall and 98.34% f1-score on the SIRI-WHU dataset. The developed ResNet-50 based feature extraction technique obtained 99.75% accuracy, 99.43% precision, 99.21% recall and 99.31% f1-score on the UCM dataset.

Table 3 evaluates the performance of the classifier with different metrics on three datasets EuroSAT, SIRI-WHU and UCM. The developed LSTM with TPP layer technique obtained 98.37% accuracy, 97.51% precision, 97.04% recall and 97.27% f1-score on EuroSAT dataset.

Table 2. Performance of ResNet-50 based feature extraction technique

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EuroSAT dataset				
AlexNet	97.19	96.32	95.85	96.08
VGG-16	97.46	96.83	96.08	96.45
VGG-19	97.79	97.02	96.34	96.67
ResNet 18	98.07	97.28	96.67	96.97
ResNet 50	98.37	97.51	97.04	97.27
SIRI-WHU dataset				
AlexNet	97.54	97.25	97.12	97.18
VGG-16	97.83	97.56	97.31	97.43
VGG-19	98.05	97.81	97.57	97.68
ResNet 18	98.32	98.06	97.83	97.94
ResNet 50	98.77	98.41	98.28	98.34
UCM dataset				
AlexNet	97.28	97.08	96.72	96.89
VGG-16	97.93	97.72	97.34	97.52
VGG-19	98.16	98.01	97.85	97.92
ResNet 18	98.83	98.42	98.22	98.31
ResNet 50	99.75	99.43	99.21	99.31

Received: November 20, 2024. Revised: December 19, 2024.

Table 3. Performance of LSTM with TPP layer technique

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EuroSAT dataset				
MLP	97.02	96.77	96.20	96.48
CNN	97.43	96.95	96.41	96.67
RNN	97.71	97.03	96.65	96.83
LSTM	98.04	97.27	96.89	97.07
Proposed LSTM with TPP layer	98.37	97.51	97.04	97.27
SIRI-WHU dataset				
MLP	97.03	96.87	96.47	96.66
CNN	97.49	97.18	96.82	96.99
RNN	97.87	97.56	97.35	97.45
LSTM	98.23	97.93	98.02	97.97
Proposed LSTM with TPP layer	98.77	98.41	98.28	98.34
UCM dataset				
MLP	98.05	97.89	97.66	97.77
CNN	98.43	98.22	98.03	98.12
RNN	98.77	98.51	98.32	98.41
LSTM	99.19	99.03	98.85	98.93
Proposed LSTM with TPP layer	99.75	99.43	99.21	99.31

The developed LSTM with TPP layer technique obtained 98.77% accuracy, 98.41% precision, 98.28% recall and 98.34% f1-score on the SIRI-WHU dataset. The developed LSTM with TPP layer technique obtained 99.75% accuracy, 99.43% precision, 99.21% recall and 99.31% f1-score on the UCM dataset.

4.1 Analysis of NDVI

The NDVI is the key metric commonly utilized in remote sensing to assess and monitor vegetation health, coverage and density. The NDVI is measured from satellite images through compared with reflectance values in Near-Infrared (NID) and red (RED) bands of the electromagnetic spectrum. The band of NIR reflected the healthy vegetation when RED band absorbed the light. The mathematical formula for NDVI is given in Eq. (17),

$$NDVI = \frac{(NIR-RED)}{(NIR+RED)} \tag{17}$$

The NIR represents the quantity of near-infrared light reflected through vegetation. The healthy vegetation reflected numerous NIR lights because of internal architecture. The RED represents the quantity of red light absorbed through chlorophyll in vegetation. The healthy plants absorbed the red light, whereas less healthy vegetation reflected much red light. Table 4 represents the threshold value of different VIs.

The below figure 7 is confusion matrix for EuroSAT dataset. The below figure 8 is confusion matrix for SIRI-WHU dataset. The below figure 9 is confusion matrix for UCM dataset. The below figure 10 is confusion matrix for NWPU dataset.

Table 4. Threshold value of VIs

NDVI values	Description
0.1 or less	Low NDVI
0.2 to 0.5	Moderate NDVI
0.6 to 0.9	High NDVI

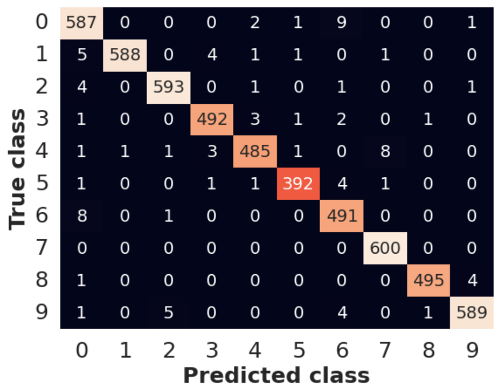

Figure. 7 Confusion matrix for EuroSAT

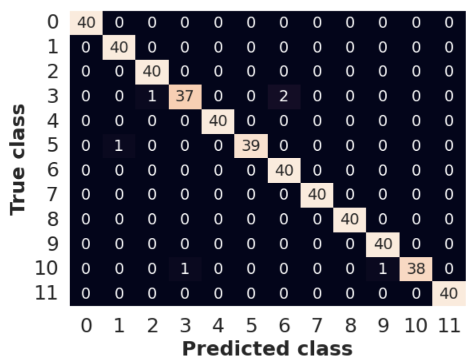

Figure. 8 Confusion matrix for SIRI-WHU

Received: November 20, 2024. Revised: December 19, 2024.

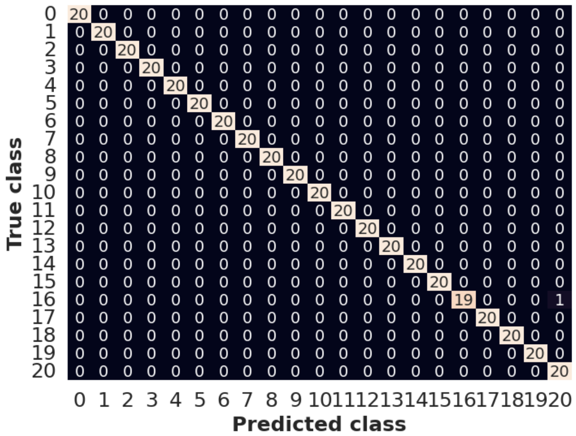

Figure. 9 Confusion matrix for UCM

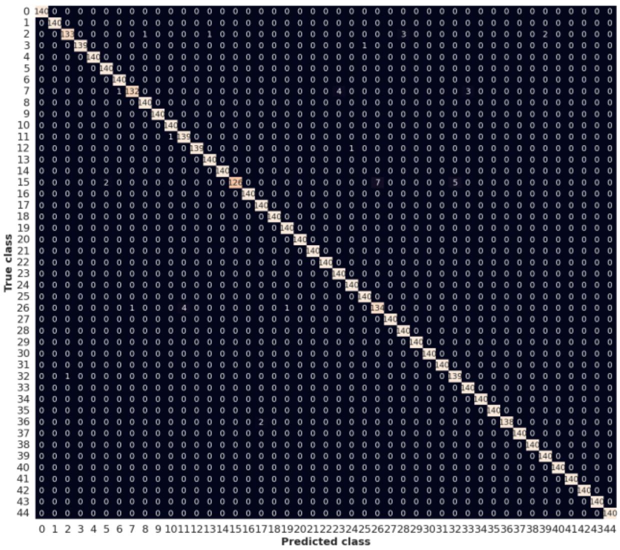

Figure. 10 Confusion matrix for NWPU

4.2 Comparative analysis

The performance of developed LSTM with TPP layer technique is compared with existing algorithms like Optimized self-attention fused CNN [15], 16-tiny Vision transformer and SIBNet [16], OG-WOA and Bi-LSTM [14], D-CNN [17] and AMGNN-LU/LC [18] on EuroSAT, SIRI-WHU, UCM, NWPU datasets. The proposed LSTM with TPP layer technique obtained 98.37% accuracy on EuroSAT dataset, 98.77% accuracy on SIRI-WHU dataset, 99.75% accuracy on the UCM dataset, 99.35% accuracy on NWPU dataset. Table 5 represents the comparative analysis of developed LSTM with the TPP layer technique.

4.3 Discussion

The performance of LSTM with the TPP layer is evaluated with three datasets such as EuroSAT, SIRI-WHU and UCM datasets. The developed method is evaluated with different algorithms like AlexNet, VGG-16, VGG-18, ResNet-18, MLP, CNN, RNN and traditional LSTM. Moreover, the developed LSTM with TPP layer is compared with existing methods like Optimized self-attention fused CNN [15], 16-tiny Vision transformer and SIBNet [16], OG-WOA and Bi-LSTM [14], D-CNN [17] and AMGNN-LU/LC [18] on EuroSAT, SIRI-WHU, UCM and NWPU datasets. These existing algorithms failed to capture multiple temporal features and deep meaningful features, struggled to scale the pixel values in the image, and needed a huge amount of labelled data to obtain reasonable results.

Table 5. Comparative analysis of developed LSTM with TPP layer technique

Datasets	Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EuroSAT	Optimized self-attention fused CNN [15]	89.50	88.63	88.65	88.63
	16-tiny Vision transformer and SIBNet [16]	97.8	97.00	96.97	96.98
	D-CNN [17]	98.63	97.64	97.11	NA
	AMGNN-LU/LC [18]	98.03	99	99.2	99
	Proposed LSTM with TPP layer	98.37	97.51	97.04	97.27
SIRI-WHU	Optimized self-attention fused CNN [15]	98.2	98.23	98.20	98.21
	Proposed LSTM with TPP layer	98.77	98.41	98.28	98.34
UCM	OG-WOA and Bi-LSTM [14]	99.34	NA	99.44	NA
	Proposed LSTM with TPP layer	99.75	99.43	99.21	99.31
NWPU	OG-WOA and Bi-LSTM [14]	96.73	NA	97.21	NA
	Optimized self-attention fused CNN [15]	91.70	91.91	91.44	91.67
	16-tiny Vision transformer and SIBNet [16]	98.9	98.26	98.13	98.19
	Proposed LSTM with TPP layer	99.35	98.71	98.27	98.54

Received: November 20, 2024. Revised: December 19, 2024.

These limitations minimize the classification performance and reduce the classification accuracy. To overcome these limitations, this article developed the LSTM with the TPP layer method to classify different classes in LULC. The pixel values in images of the dataset are scaled by using min-max normalization in the pre-processing phase. Then, the deep meaningful features are extracted by using the ResNet-50 technique and the LSTM network is incorporated with TPP layers that capture multiple temporal features and improve the classification performance.

5. Conclusion

The effective DL-based algorithm is developed to classify the different LULC classes with high classification accuracy. The datasets used in this research are EuroSAT, SIRI-WHU, UCM and NWPU dataset and the images are scaled into uniform range by using min-max normalization technique. Then, the deep meaningful features are extracted by using the ResNet-50 technique which helps to differentiate the various LULC classes. Finally, the extracted features are classified by using the LSTM with TPP layer which captures the multiple temporal features and classifies the LULC patterns with high classification accuracy. The performance of LSTM with the TPP layer technique is evaluated on EuroSAT, SIRI-WHU, UCM and NWPU datasets. The proposed LSTM with TPP layer technique obtained 98.37% accuracy on EuroSAT dataset, 98.77% accuracy on SIRI-WHU dataset, 99.75% accuracy on the UCM dataset and 99.35% accuracy on NWPU dataset. In future, different DL-based algorithms can be used to further improve the LULC classification.

Notations

Notations	Description
$min\ (x)$	Minimum Value
$max(x)$	Maximum Value
x	Actual Image
x'	Normalized Image.
W_i	Weight of the Convolution Layer
$F(x, \{W_i\})$	Residual Mapping
W_{hx}	Weight of the Input Feature and Hidden Gate
W_{hh}	Weight of Further Time Interval with Hidden Gate
b	Bias
σ	Non-Linear Activation Function
L_{out}	Length of the Output Sequence
K_{TPP}	Size of the Kernel
S_{TPP}	Stride
P_{TTP}	Padding Of Maxpooling Process

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author.

References

[1] A.A. Darem, A.A. Alhashmi, A.M. Almadani, A.K. Alanazi, and G.A. Sutantra, "Development of a map for land use and land cover classification of the Northern Border Region using remote sensing and GIS", *The Egyptian Journal of Remote Sensing and Space Science*, Vol. 26, No. 2, pp. 341-350, 2023.

[2] A.E. Al-Dousari, A. Mishra, and S. Singh, "Land use land cover change detection and urban sprawl prediction for Kuwait metropolitan region, using multi-layer perceptron neural networks (MLPNN)", *The Egyptian Journal of Remote Sensing and Space Science*, Vol. 26, No. 2, pp. 381-392, 2023.

[3] S. Thirumaladevi, K.V. Swamy, and M. Sailaja, "Remote sensing image scene classification by transfer learning to augment the accuracy", *Measurement: Sensors*, Vol. 25, p. 100645, 2023.

[4] B. Ahmed, T. Akram, S.R. Naqvi, A. Alsuhaibani, Y.N. Altherwy, and U. Masud, "A Novel Deep Learning Framework with Meta-Heuristic Feature Selection for Enhanced Remote Sensing Image Classification", *IEEE Access*, Vol. 12, pp. 91974-91998, 2024.

[5] J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot, "Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 61, pp. 1-15, 2023.

[6] A. Temenos, N. Temenos, M. Kaselimi, A. Doulamis, and N. Doulamis, "Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP", *IEEE Geoscience and Remote Sensing Letters*, Vol. 20, pp.1-5, 2023.

[7] M. Aljebreen, H.A. Mengash, M. Alamgeer, S.S. Alotaibi, A.S. Salama, and M.A. Hamza, "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.

[8] M. Jean Bosco, R. Jean Pierre, M.S.A. Muthanna, K. Jean Pierre, A. Muthanna, and A.A. Abd El-Latif, "MGFEEN: a multi-granularity feature encoding ensemble network for remote sensing image classification", *Neural Computing and Applications*, Vol. 36, No. 12, pp. 6547-6558, 2024.

[9] M. Mukhedkar, C. Kaur, D.S. Rao, S. Bandhekar, M.S. Al Ansari, M. Syamala, and Y.A.B. El-Ebiary, "Enhanced Land Use and Land Cover Classification Through Human Group-based Particle Swarm Optimization-Ant Colony Optimization Integration with Convolutional Neural Network", *International Journal of Advanced Computer Science & Applications*, Vol. 14, no. 11, pp. 404-419, 2023.

[10] N. Guo, M. Jiang, D. Wang, X. Zhou, Z. Song, Y. Li, L. Gao, and J. Luo, "Scene classification for remote sensing image of land use and land cover using dual-model architecture with multilevel feature fusion", *International Journal of Digital Earth*, Vol. 17, No. 1, p. 2353166, 2024.

[11] M. Stoimchev, J. Levatić, D. Kocev, and S. Džeroski, "Semi-Supervised Multi-Label Classification of Land Use/Land Cover in Remote Sensing Images with Predictive Clustering Trees and Ensembles", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.62, 2024.

[12] Wang, C., Li, J., Tanvir, A., Yang, J., Xie, T., Ji, L. and Zhang, T., "Zero-Shot Remote Sensing Scene Classification Based on Local-Global Feature Fusion and Weight Mapping Loss", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 17, pp. 2763-2776, 2023.

[13] V.P. Yele, S. Alegavi, and R.R. Sedamkar, "Effective segmentation of land-use and land-cover from hyperspectral remote sensing image", *International Journal of Information Technology*, Vol. 16, pp. 2395-2412, 2024.

[14] 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 Journal on Advances in Signal Processing*, Vol. 2023, No. 1, p. 13, 2023.

[15] H.M. Albarakati, M.A. Khan, A. Hamza, F. Khan, N. Kraiem, L. Jamel, L. Almuqren, and R. Alroobaea, "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*, Vol. 17, pp. 6338-6353, 2024.

[16] S. Rubab, M.A. Khan, A. Hamza, H.M. Albarakati, O. Saidani, A. Alshardan, A. Alasiry, M. Marzougui, and Y. Nam, "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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 17, pp. 13135-13148, 2024.

[17] A. Bhatt, and V.T. Bhatt, "Dcrff-Lhrf: an improvised methodology for efficient land-cover classification on eurosat dataset", *Multimedia Tools and Applications*, Vol. 83, No. 18, pp. 54001-54025, 2024.

[18] V.P. Yele, S. Alegavi, and R.R. Sedamkar, "Hybrid hesitant fuzzy linguistic bi-objective binary coyote clustering based segmentation and classification for land use land cover in hyperspectral image", *International Journal of Information Technology*, Vol. 16, No. 1, pp. 525-534, 2024.

[19] EuroSAT dataset: https://www.kaggle.com/datasets/apollo2506/eurosat-dataset.

[20] SIRI-WHU dataset: https://www.kaggle.com/datasets/lzsy0226/siri-whu-data-set.

[21] UCM dataset: http://weegee.vision.ucmerced.edu/datasets/landuse.html.

[22] NWPU dataset: https://www.kaggle.com/datasets/aqibrehmanpirzada/nwpuresisc45