Advances in Artificial Intelligence and Machine Learning; Research (Ahead of Print) Received 25-12-2025; Accepted 19-03-2026; Published online 26-03-2026

Land Cover Classification of Fused Hyperspectral and Multispectral Image Using In-Ception-Resnetv2

Aparna Halbe aparna_halbe@spit.ac.in
Assistant Professor, Computer Science and Engineering
Sardar Patel Institute of Technology,
research scholar, VJTI
Mumbai, Maharashtra 400056, India

Sunil Bhirud vc@coeptech.ac.in
Vice-Chancellor,
COEP Tech. University,
Pune, Maharashtra, 411005, India

Corresponding Author: aparna_halbe@spit.ac.in

Abstract

Remote sensing data from various satellites and multimodal data processing techniques have grabbed the curiosity of geoscience communities. Accurate classification of land cover from remotely sensed satellite images remains a serious challenge. The proposed work develops an Inception-ResNet v2 to classify different land covers of the Konkan region of India. Quick Atmospheric Correction (QUAC), stacking, and merging techniques are used to pre-process multispectral images. SG smoothing and merging techniques are used to pre-process hyperspectral images. The pre-processed spectral information contains multiple wavebands, resulting in a high-dimensional feature set. Therefore, Principal Component Analysis (PCA) with Fennec Fox Optimization (FFO) is applied to extract useful wavebands. Subsequently, the extracted bands are sent to the hybrid NCTCP-DTCWT data fusion algorithm. These fused images are then given to Inception-ResNet v2 for classification. Proposed deep learning model can detect different land covers such as forest region, vegetation land, water, built-up, and barren land in the given fused image. The model achieved 96% accuracy, 91% precision, and 98% specificity. Comparison of different metrics prove that our hybrid model outperformed several other machine learning models explored in earlier work. Further, the proposed model is tested with fused images and non-fused images. The study proves that fused images can generate accurate classes than non-fused images. This study examines whether optimized band selection combined with multi-resolution hyperspectral–multispectral fusion within a unified deep-learning framework can significantly improve land-cover classification in complex coastal environments.

Keywords: Analysis of satellite images, Data fusion, Band selections and fusion technique, Fennec fox optimization, Inception-Resnet v2

Citation: Aparna Halbe And Sunil Bhirud. Land Cover Classification of Fused Hyperspectral and Multispectral Image Using In-Ception-Resnetv2. **Advances in Artificial Intelligence and Machine Learning**. 2026. (Ahead of Print). https://dx.doi.org/10.54364/AAIML.2026.62293

1. INTRODUCTION

Remote sensing is extremely important in today's world. It assists and supports communities in a range of fields such as agricultural activities, monitoring the environment, geology [1]. Remote sensing assists the country's growth by classifying and monitoring land usage, as well as detecting problems in hazardous or remote regions. Using satellite images, it is possible to track a variety of things on a regular basis, including water resources monitoring, soil mapping, crop area estimation, agriculture, urban areas, desertification and monitoring and identification of the characteristics of water, crops, and soil [2]. By quantitatively displaying changes in people, vegetation, and water regions over a period of years, Land Use Land Classification models also emphasize historical changes in an area. [3].

Traditional land cover classification methods primarily rely on multispectral images, but these often face challenges due to limited spectral information and resolution. The advent of hyperspectral imaging has provided advantages in terms of spectral detail, enabling the detection of subtle variations in the land surface[4, 5]. However, hyperspectral images often suffer from lower spatial resolution, making it challenging to precisely delineate land features[6].

To overcome these limitations, the fusing hyperspectral and multispectral images is a promising approach. Hyperspectral images provide high spectral resolution, while multispectral images offer better spatial resolution. By combining the two, it is possible to obtain both the spectral richness and spatial clarity required for accurate land cover classification. This integration boosts the model's ability to differentiate between similar land cover types, a task that would be difficult with only a single type of imaging data. This model combines the strengths of the Inception network, which is efficient at capturing spatial features through multiple filter sizes, and the ResNet architecture, which enables the learning of deeper networks through residual connections. The Inception-ResNet V2 model has demonstrated its capability in handling high-dimensional data, such as hyperspectral images, and has been successful in a variety of remote sensing tasks, making it an ideal choice for land cover classification of fused hyperspectral and multispectral images.

In this study, we present a method for classifying land cover in the Konkan region, situated on India's western coast, by utilizing fused hyperspectral and multispectral images. The fusion of these two types of imagery leverages the complementary strengths of spectral and spatial information, resulting in more accurate land cover classifications. The proposed Inception-ResNet V2 CNN model has proven to be effective in managing complex, high-dimensional data and capturing fine-grained features essential for distinguishing between land cover types. The Konkan region, characterized by its diverse land cover types, including forests, agricultural fields, and urban areas, presents a challenging but ideal case for testing our method. Here multispectral images are obtained from Landsat 8, while hyperspectral images are acquired from Bhuvan, an Indian remote sensing platform. The fusion of these images is done to combine the spatial detail of the multispectral images with the spectral richness of the hyperspectral data.

The focus of this research is to explore the effectiveness of fused hyperspectral and multispectral imagery for land cover classification, and to assess the effectiveness of the Inception-ResNet V2 in classifying the diverse land cover types of the Konkan region. The results of this study aim to provide better understanding of deep learning and data fusion techniques to improve land cover mapping in regions with complex and diverse land characteristics.

According to several studies, there are several spectral and geographical constraints on the LULC classification that reduce its accuracy when employing medium and low-resolution satellite images [7, 8]. Machine learning algorithms, including SVM, RF, RBF, NB, MLC, and DT are employed to overcome the obstacles and produce high-precision LULC pictures. Despite its many benefits, machine learning has many drawbacks. For example, given that the training data, parameter inputs, and machine learning model configuration are all required for successful outcomes, it could be difficult to consistently generate a high-precision LULC map [9]. To address these problems, we have proposed a novel approach for land classification which uses a fused multispectral and hyperspectral image which is given to a deep learning model for identifying different classes in the fused image. For detecting the land covers various new methods are explored. Some of the methods that are now used to identify land cover are examined in the following sections.

Reference [10], suggested the use of PCA as a group of related variables (original bands) is converted into a new set of independent variables (principal components) that preserve most of the original information, though their physical interpretation requires further analysis. Reference [11], has suggested Standard Neural Net-works (NNs) with Principal Component Analysis (PCA). This technique first combines three RGB bands of a multispectral satellite image using the Brovey transform algorithm. However, it requires high cost. The proposed Squirrel search-based Qubit Lattice Model (SbQLM) [12], efficiently selects informative bands by leveraging quantum qubits and squirrel-inspired search behavior, enhancing image quality and noise reduction. This model can be improved with fusion of hyperspectral and multispectral bands. Reference [13], established land cover mapping utilizing a SVM with complicated tree classifiers. According to analysis of [14], the quantity of agricultural land in the focus area strengthened by 1180 km, whereas the area covered by trees and grassland dropped with 828.7 km2 and 444.1 km2, accordingly. Reference [15], created a point-wise deep learning method based on aerial multispectral LiDAR data for LC classification. The proposed approach per-forms well on the test scenarios, with an accuracy rate of 96.9% and a Kappa value of 0.950. A pixel-certainty activity learning (PCAL) is a technique was developed by [16], that learns from textural patterns using information from extended differential patterns (EDP). Using this pattern col-lection, PCAL algorithms were employed to categorize HSI patterns, which are important for distant sensing applications. In India's humid tropical region [17], developed spectral angle mapper and SVM algorithms to classify land use using hyperspectral data. Several concerns are arising in the above reviewed articles during the land cover detection process. The high cost [18], the lack of discussion on the pre-processing effects on the resulting LULC [13], and the time-consuming nature of the process [19], were noted. Consequently, some of the concerns in the works that are currently available include several socio-economic and environmental challenges that have major geographical and temporal implications [16]. To manage these concerns, the proposed study utilizes a modified CNN in the Konkan region by integrating fused hyperspectral and multispectral images.

The primary contributions of this study are listed below:

- We propose a feature extraction mechanism with an optimization algorithm to select the optimal bands from satellite imagery.
- Secondly, we introduce a fusion technique called the Hybrid NCTCP-DTCWT approach to fuse multispectral and hyperspectral images.

- We have built a deep learning model to detect different land covers in the image generated by fusing MSI and HSI.

As per the results, our model outperforms the existing models currently in use.

2. GENERATION OF DATA

Hyperspectral images of the Konkan are obtained from the Bhuvan website (Bhuvan) [20], a platform that provides data captured in 64 spectral bands within the Visible and Near-Infrared (VNR) range, a spatial resolution of 500 meters and a swath width of 128 kilometres. These images offer detailed information across a broad spectrum, enabling comprehensive analysis of the region's land cover and vegetation. Additionally, multispectral satellite images from Landsat 8 are sourced from the U.S. Geological Survey's Earth Explorer website [21], offering valuable data for further study. Landsat 8 provides a panchromatic resolution of 15 meters and multispectral resolution of 30 meters, with a swath width of 185 kilometres, allowing for detailed observation of the land surface over a larger area. For the study, a total of 500 samples are generated through an augmentation process, effectively increasing the dataset's size and diversity for robust analysis. Of these 500 samples, 80% (or 400 images) are utilized for training machine learning models, while the remaining 20% (100 images) are set aside for testing, ensuring an effective evaluation of the model's performance. Additionally, India's land classification data, available at a 100-meter resolution on a decadal basis [22], is used as a reference for land cover categorization across the region. This combination of hyperspectral, multispectral, and land classification data forms the foundation for the analysis.

3. METHODOLOGY

In this study, both hyperspectral and multispectral images are utilized for land cover classification, focusing on the Konkan region of Maharashtra. The process begins with applying the QUAC method to the multispectral image to correct atmospheric distortions. Following this, stacking and merging techniques are applied to produce high-resolution multispectral images. For the hyperspectral data, a Savitzky-Golay (SG) smoothing technique is used as a preprocessing step to enhance data quality. After smoothing, various bands from the hyperspectral images are merged to create a high-dimensional feature set with multiple wavebands, which is critical for comprehensive analysis.

We apply optimization algorithm FFO along with PCA on high dimensional feature set to extract useful wavebands. The optimized features are then fused using Nonlocal Coupled Tensor CP Decomposition and Dual-Tree Complex Wavelet Transform for further processing. The result is a final fused image that contains complementary information from both hyperspectral and multispectral sources.

Following fusion, the image is processed by a deep learning classifier Inception-ResNet V2 model, to identify land cover types. This model has been shown to achieve high accuracy in identifying various land classes from complex image data, making it ideal for this type of analysis. Using advanced preprocessing, feature selection, and fusion techniques, this study demonstrates a com-

prehensive approach to land cover classification, leveraging the power of both hyperspectral and multispectral imaging to produce high-accuracy results. This is shown in FIGURE 1.

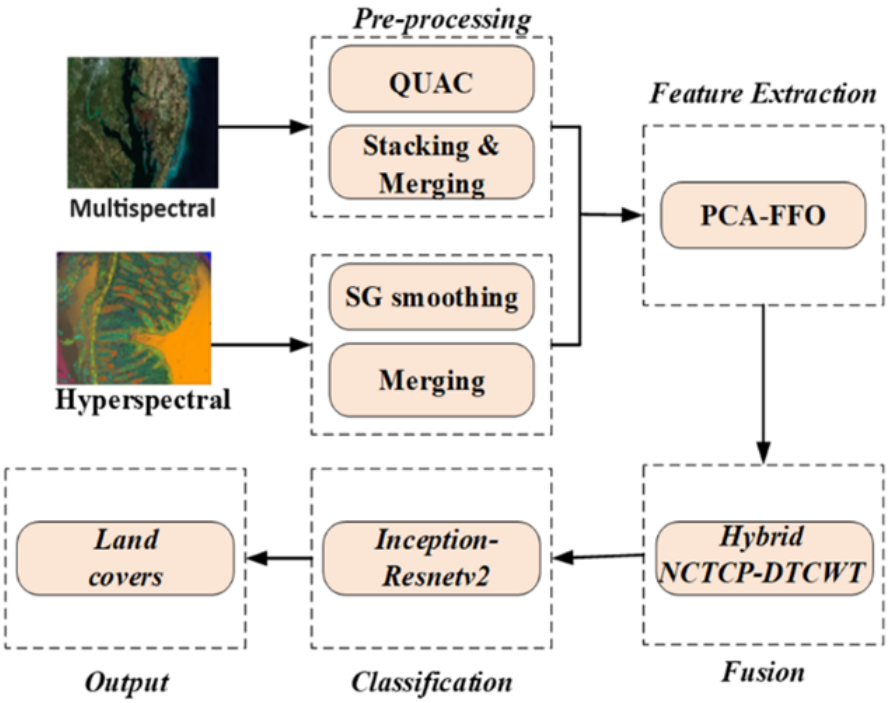


Figure 1: Land Cover Classification of fused HSI and MSI image

3.1 Multispectral Image Pre-processing

The Landsat satellite has total of 8 bands. To correct radiometric error, each band of multispectral image undergoes QUAC image preprocessing technique. Multispectral image has low resolution bands. To get high resolution image, seven bands are stacked and finally it is merged with the eighth band.

3.1.1 Quick Atmospheric Correction (QUAC)

Analytical approaches often do not require any field measurements and rely heavily on the picture data (Basith, 2020). Additionally, using the offset and gain parameters, QUAC can recover surface reflectance inside the picture. These parameters are stated in 1

$$\rho_{\mathrm{SUP}} = g\,(L_{\mathrm{TOA}} - o) \tag{1}$$

3.1.2 Layer stacking

The process of connecting various bands to generate a multi band image is called layer stacking. The spectral subset bands of Landsat-8 have different spatial resolutions. Spatial resolution is of 30 meters from 1st to 5th Band and 7th band, while the TIR and panchromatic (PAN) band, band 8, have spatial resolutions of 60 meters and 15 meters, respectively [15]. In the proposed method, bands 1–7 of the Landsat-8 MSI are merged.

3.1.3 Resolution merge

The resolution merging process combines a multiband image of low-quality with a single-band image of high-quality to produce an image with high-resolution hyperspectral image.

3.2 Hyperspectral Image Pre-Processing

As a hyperspectral image pre-processing technique, SG smoothing and merging are applied.

3.2.1 Savitzky-Golay (SG) smoothing

This smoothing method eliminates noise from signals of suddenly fluctuating and non-stationary data. The signal width decreases with increasing polynomial order value. Consider the time series x(n) to have n elements. Let w(n) stands for the noise of the y(n) signal, as the observed time series. By using 2, the output of the SG filter $\hat{x}(n)$ is determined

$$\hat{x}(n) = \sum_{k=-M}^{M} h(n)y(n-k) \tag{2}$$

Where M is the parameter of SG filter and h(n) is the impulse response of the filter [23].

3.3 Merging

The merging process combines a multi band image of low-quality with a single-band image of high-quality to produce an image with high-resolution hyperspectral image. All the bands present in the hyperspectral dataset are merged to produce a single image.

3.4 Feature Extraction Using PCA-FFO

Pre-processed spectral data has multiple wavebands; the feature sets are high-dimensional. Therefore, it is required to reduce the feature set's size while keeping most of the dataset's information [24]. The proposed method employs a hybrid PCA-FFO strategy for feature selection. Features are

typically chosen based on Fennec Fox Optimization for better accuracy. This optimization ensures that PCA selects the most key features from the input data.

3.4.1 Fennec fox optimization algorithm

A fox belonging to the fennec resides in Sinai Peninsula in Egypt and North Africa. The fennec fox, the smallest species of canid, is recognized for its large ears. This type of fox can sense any kind of motion even under the ground through their special ears [24]. Nature inspired mathematical model of FFO for identifying the optimal components is explained below.

Step 1: Initialization Fennec foxes are generated at random in the search space shown in (3)

$$I = \{N_1, N_2, \ldots, N_n\} \tag{3}$$

Step 2: Fitness Function Fitness function is represented as shown in (4)

$$E = \frac{1}{2}\sum_{n=1}^{N}(y_n - z_n)^2 \tag{4}$$

Where y_n is displayed value, z_n is the actual value, E represents Mean Squared Error (MSE).

Step 3: Update the value. According to the fennec fox's behavior, the value changes with each iteration until it achieves its ideal level. The mathematical simulation of this stage for updating FFO members makes use of (5) and (6).

$$Q_N = \begin{cases} Q_N^{p1}, & H_N^{p1} < H_i \\ Q_N, & \text{otherwise} \end{cases} \tag{5}$$

$$Q_N = \begin{cases} Q_N^{p2}, & H_N^{p2} < H_i \\ Q_N, & \text{otherwise} \end{cases} \tag{6}$$

Where, H_N^{p1} are its j^{th} dimension and value of the objective function according to initial phase, Q_N^{p1} is the i^{th} fennec fox's new projected state. The next planned state of i^{th} fennec fox is represented by Q_N^{p2} in the second phase, and H_N^{p2} represents the value of its goal function.

Step 4: Termination The procedure ends when the best solution has been found. This feature extracted image is then fed into fusion process.

3.5 Fusion

The output image obtained from the feature extraction technique is given as an input for the fusion technique. This method uses the Hybrid NCTCP-DTCWT approach to combine the features of both hyperspectral and multispectral images. A single algorithm alone does not capture sufficient details,

as it extracts fewer features. In this hybrid approach, NCTCP and DTCWT are applied separately to fuse hyperspectral and multispectral images, and the resulting images from both methods are then merged to create a single, comprehensive fused image. This hybrid fusion approach offers more detailed information from both hyperspectral and multispectral images compared to other fusion methods.

3.5.1 NCTCP (Nonlocal Coupled Tensor CP Decomposition) Approach

Tensor Representation of HSI-MSI Fusion: Tensor X of high-resolution Hyperspectral image (HR-HSI) consists of three orders. To get low resolution Hyperspectral image (LR-HSI) within the tensors, the down sampling operator (H) and the spatial blurring operator (S) are used. The representation of LR-HSI is expressed as (7)

$$\tilde{X} = XSH + \varepsilon_h \tag{7}$$

The representation of HR-MSI tensor format is expressed in (8)

$$y = RX + \varepsilon_m \tag{8}$$

If ε_h and ε_m are independent noise terms with identical distributions (i.i.d.). According to nonlocal similarity, many regions with a similar structure may exist inside a single picture. To do this HR-HSI has been divided into two sets of overlapping spatial-spectral blocks cubes as $\{P_{ij}\}_{1\le i\le W-d_w, 1\le j\le H-d_h} \subset \mathbb{R}^{d_w d_h \times B}$. A 3D cube's width and height are represented by the variables d_w and d_h. In linguistic order, set up each band into a column vector to unfold each cube into its matrix structure. Each of these 3D cubes is now transformed into a collection of 2D patches as $\Omega = \{P_i \in \mathbb{R}^{d_w d_h \times B}\}$, where $N = (W - d_w + 1) \times (H - d_h + 1)$ corresponds to the count of local patches obtained from HR-HSI. K clusters are often formed from the 2D patch set Ω. Make a 3-order tensor out of each cluster. Define $G_p X$ as in (9) and define $D_p^m\{\}$ serving to extract the m-th spatial patch (size $d_w \times d_h$) with B bands from the p-th cluster.

$$G_p X := \left(D_p^1\{X\}, D_p^2\{X\}, ..., D_p^{N_p}\{X\}\right) \in \mathbb{R}^{d_w d_h \times N_p \times B} \tag{9}$$

In this case, the p-th cluster's total number of nonlocal patches is represented by N_p. From 3D cubes to 2D patches, the arrangement of the cubes may be smoothly unfolded in this place. An acceptable model in this case is the CP decomposition of $G_p X$, which may be expressed as (10).

$$G_p X = [[A_p, B_p, C_p]] \tag{10}$$

$$X = \left(\sum\nolimits_p G_p^T G_p\right)^{-1} \sum\nolimits_p G_p^T [[A_p, B_p, C_p]] \tag{11}$$

According to (11), the perfect X is determined by adding up the values from each cluster and averaging the results. It integrates the nonlocal tensor information into a coupled tensor CP decomposition issue. It is referred as NCTCP decomposition. This HR-HSI undisturbed spectrum structure, represented in a nonlocal tensor format, demonstrates that it is feasible to transform HR-HSI to HR-MSI with this representation [25].

3.5.2 Dual Tree Complex Wavelet Transform (DT CWT) Approach

It generates frequency sub bands by repeatedly applying distinct spatial filters. A sparse and efficient computational filter bank represents a wavelet. A more accurate representation of the discontinuity can be obtained by creating basis $\psi(x)$ using the function of scaling $v(x)$.

$$\psi_{a,b} = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-a}{b}\right) \tag{12}$$

$$f(x) = \sum_{n=-\infty}^{+\infty} c(n)\cdot\phi(x-n) + \sum_{j=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} d(j,n)2^{\frac{j}{2}}\cdot\psi(2^j x-n) \tag{13}$$

(13) expresses the mother wavelet kernel, a and b are scalars in this instance, while $\psi(x)$ is the wavelet basis. (14) provides the wavelet-decomposed version.

$$\begin{aligned}\phi_{j,k}(x) &= 2^{\frac{j}{2}}\phi(2^j x-k)\\ \psi_{j,k}(x) &= 2^{\frac{j}{2}}\psi(2^j x-k)\end{aligned} \tag{14}$$

With the help of (14), the discrete wavelet transform (DWT) generates its scaling basis function and wavelet basis function. A signal is transformed from one form to another showing a particular representation of the original signal. Following the discovery of wavelet functions, each of the registered input pictures is first converted into a set of DTCWT is employed for calculating both coefficient of high-frequency N as well as coefficient of low frequency M as shown in (15).

$$(M,N) = \text{DTCWT}(I) \tag{15}$$

In general, low frequency coefficients are named as M1, M2 and the high frequency coefficients are named as N1, N2 and are handled separately to produce fused image [26], shown in equation (16).

$$\begin{aligned}M_F &= \phi_M(M_1,M_2)\\ N_F &= \phi_N(N_1,N_2)\end{aligned} \tag{16}$$

In the fusion framework, ϕ_M and ϕ_N refer to the respective fusion weights for the low- and high-frequency components. The resulting fused coefficients are denoted by M_F for low-frequency and N_F for high-frequency domains.

Fusion Rules:
The most chosen picture fusion rule is the generic and practical rule for fusing high frequency coefficients.

$$N_F = \begin{cases} N1 & abs(D1) \geq abs(D2) \\ N2 & Otherwise \end{cases} \tag{17}$$

It is essential to fuse a low-frequency coefficient effectively because it preserves global qualities that enhance the fused image's natural appearance. (17) describes Pixel-level weighted averaging

fusion strategy. (18) describes the fused low-frequency coefficient M_F.

$$M_F = w_1 \times M_1 + w_2 \times M_2 \tag{18}$$

Where w_1, w_2 have values ranging from 0 to 1. Typically, the weights are set to 0.5. The pixel-based weighted average fusion rule considerably reduces the contrast in the fused image and alters the intensity range when merging multi-modal images with varying dynamic ranges. However, it performs effectively with images of the same modality.

3.6 Classification

The selection of training and testing datasets representing diverse surface cover types is grounded in previously identified real-world land cover classes. The suggested method for categorization uses a modified CNN termed Inception-Resnet V2 to detect these areas.

3.6.1 Inception-Resnet V2

The Inception network architecture aims to boost network capacity and extract a wider range of information from different scales by using convolutional kernels of various sizes.

The architecture of CNNs employs inception structural designs for feature extraction. Inception design uses effective factorization techniques to reduce constraints and enhance processing efficiency. For instance, the Inception v2 model factorizes a 5 × 5-pixel convolution unit into multiple 7 × 7-pixel convolutions, enabling faster calculations. Factorization into smaller convolutions is beneficial because of spatial filters of large size, such as 5 × 5 or 7 × 7, often incur high computational costs. While larger filters like 5 × 5 can capture broader patterns in signals, breaking them into smaller convolutions with a multi-layer network can achieve similar results with reduced computational expense. Essentially, this approach simplifies the computation by dividing a 5 × 5 convolution into smaller, interconnected layers that collectively replicate the effects of the larger filter.

Decomposition of Spatial Data into Asymmetric Convolutions: Convolution having a 3 × 3 dimension may be achieved with merely a set of 3 x 3 convolutional layers thereby it is not certainly essential. Convolutions of size 2 x 2 can be used as a substitute. It rotates out as well, which may still be enhanced by 2 x 2 using uneven convolutions, for example, n × 1. A 2-layer network is dragged with the identical available area that in a 3 × 3 convolution if a 3 × 1 convolution is followed by a 1 × 3 convolution. For an equivalent number of output filters, the cost of the two-layer approaches has decreased by 33%.

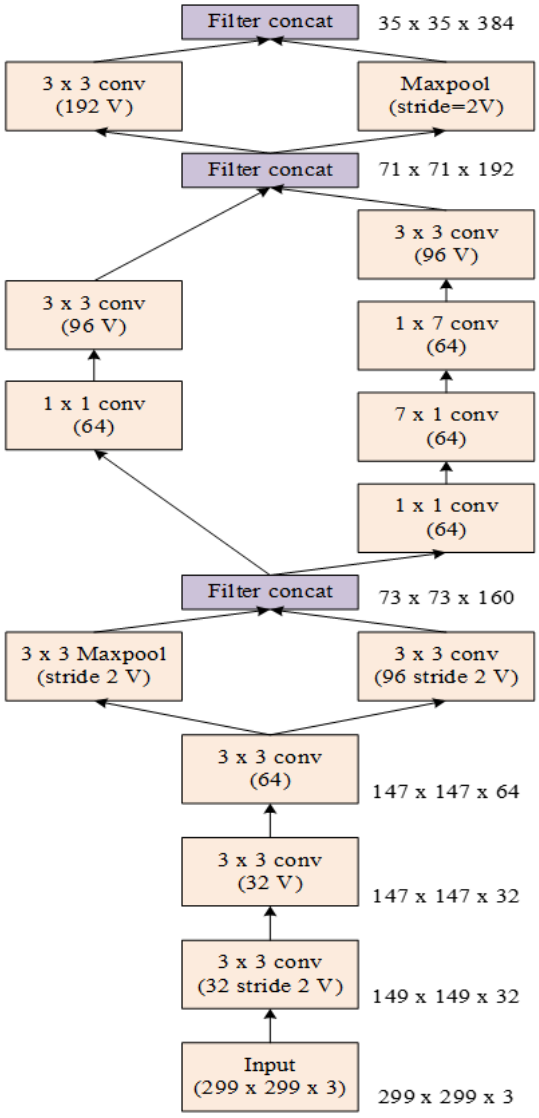


Figure 2: Architecture of Inception-ResNet v2, adapted from GeeksforGeeks [27].

By employing evaluation, total processing savings from a 3 × 3 convolution being factored into two 2 × 2 convolutions is approximately 11% [28]. FIGURE 2 displays the Inception-ResNet V2's hierarchical structure. The suggested model classifies the five major forms of land cover including forest region, vegetation land, water body, built-up land, and bare land. The hyperparameters such as learning rate and batch size of Inception-ResNet V2 model are automatically selected using random search validation process. Using this validation process, best values are selected for the hyperparameters which increases the overall performance. When there are multiple hyperparameters with a finely tuned grid of values, random search is a highly helpful tool. It receives a reasonably decent set of values for the hyperparameters by using a subset consisting of 5–100 randomly picked points.

3.7 Proposed Multimodal Fusion–Based Land-Cover Classification Algorithm

Algorithm 1 summarizes the complete pipeline for land-cover classification using fused hyperspectral and multispectral imagery. The inputs are the raw multispectral image I_{MSI} and hyperspectral image I_{HSI}. The algorithm performs atmospheric correction and spectral smoothing, followed by dimensionality reduction using PCA and band selection through Fennec Fox Optimization (FFO). The selected hyperspectral features are stacked with corrected multispectral pixels and normalized. A hybrid NCTCP–DTCWT fusion scheme is then applied to generate fused feature representations, which are finally fed to an Inception–ResNet v2 network for pixel-wise land-cover classification. The output is the predicted label map Y_{pred}.

Algorithm 1 Land classification using fused hyperspectral and multispectral images with Inception–ResNet v2

Input: Raw multispectral image I_{MSI}; raw hyperspectral image I_{HSI}; PCA variance threshold τ; trained network parameters θ

Output: Predicted land-cover map Y_{pred}

Stage 1: Pre-processing
for each pixel p in MSI **do**
$I_{MSI}^{corr}(p) \leftarrow I_{MSI}(p) - \widehat{I}_{atm}(p)$
end for
for each spectral band b in HSI **do**
$I_{HSI}^{smooth}(x,b) \leftarrow \sum_{k=-m}^{m} c_k I_{HSI}(x+k,b)$
end for
Stage 2: Dimensionality reduction and band selection
while retained variance > threshold **do**
$I_{PCA} \leftarrow U^T(I_{HSI}^{smooth} - \mu)$
end while
$I_{FFO} \leftarrow \arg\max_S f(S),\ S \subseteq I_{PCA}$
Stage 3: Feature stacking and normalization
for each pixel p **do**
$I_{stacked}(p) \leftarrow [I_{FFO}(p), I_{MSI}^{corr}(p)]$
$I_{merged}(p) \leftarrow \dfrac{I_{stacked}(p) - I_{min}}{I_{max} - I_{min}}$
end for
Stage 4: Hybrid NCTCP–DTCWT fusion
for each pixel p **do**
$C_{NCTCP}(p) \leftarrow \sum_{i,j} w_{ij}\, I_{merged}^{ij}(p)$
$C_{DTCWT}(p) \leftarrow \sum_{k=1}^{K} \psi_k\, C_{NCTCP}(p)$
end for
for each pixel p **do**
$I_{fused}(p) \leftarrow \sum_{m,n} C_{DTCWT}^{m,n}(p)\, \phi_{m,n}$
end for
Stage 5: Classification
for each pixel p **do**
$Y_{pred}(p) \leftarrow \arg\max_c P(c \mid I_{fused}(p); \theta)$
end for
return Y_{pred}

4. RESULT AND DISCUSSION

This study proposes a land classification approach for the Konkan area using fused Multispectral and Hyperspectral images with an Inception-ResNet V2 model. The method integrates both image types to enhance classification accuracy by harnessing the complementary strengths of spectral and spatial data. The implementation of the proposed method utilizes a system with the following configuration: an Intel(R) Core i5-3450S processor running at 2.80GHz, supported by an NVIDIA

GTX 1650 4 GB (GDDR6) GPU with 16 GB RAM and 64-bit OS. Detailed simulation parameters for the FFO algorithm and Inception-ResNet V2 model are provided in TABLE 1, and TABLE 2 for reference.

Table 1: Simulation parameter for Inception-ResNet V2

Optimizer	sgdm
SquaredGradientDecayFactor	0.99
MiniBatchSize	128
MaxEpochs	10
Validation Frequency	50
L2Regularization	0.0001
InitialLearnRate	0.001
VerboseFrequency	50
LearnRateDropPeriod	10
Momentum	0.9
LearnRateDropFactor	0.1

Table 2: Parameter used in FFO Simulation

Sample size	50
Termination limit	500
Lower bound search space	90
Upper bound search space	120

Landsat data provides 15m panchromatic and 30m multispectral resolutions across a 185km swath. India's Land Use and Land Cover classification is provided at a resolution of 100 meters (Decadal). FIGURE 3 depicts the Konkan region that was taken from the map of India. By using latitude and longitude we extracted the Konkan region from the India map which represents land covers of the Konkan region.

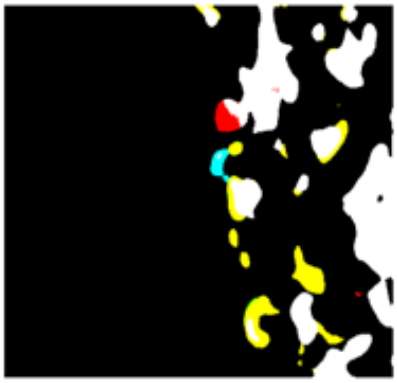
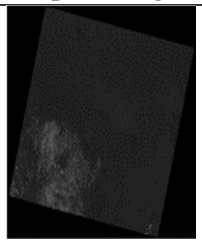
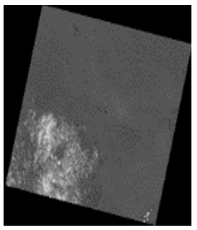
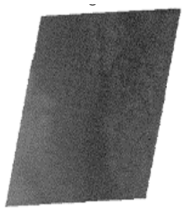
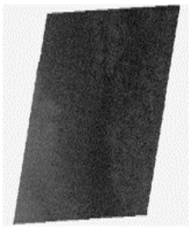
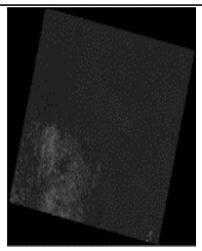
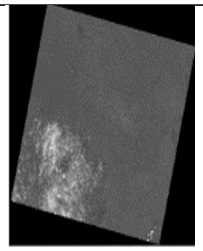
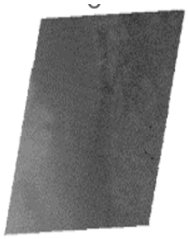
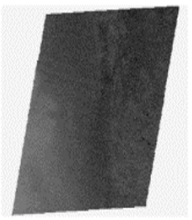
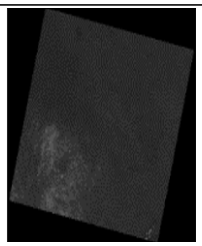
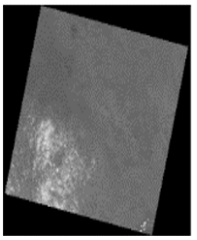
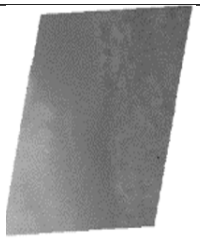
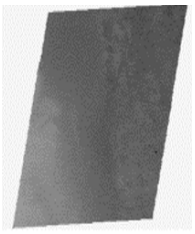
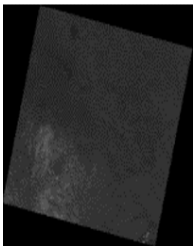
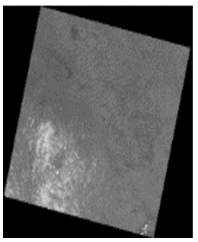
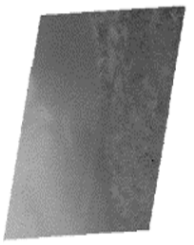
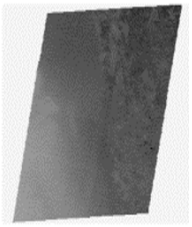
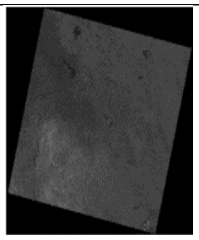
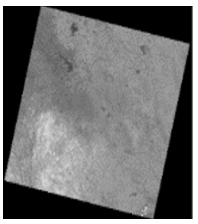
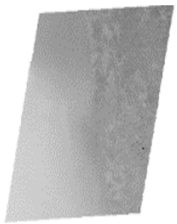
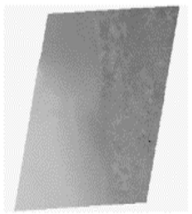
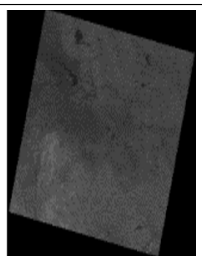
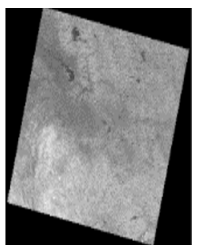
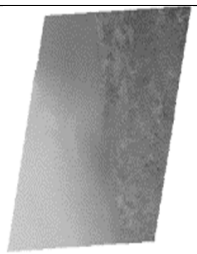
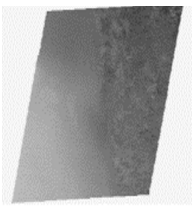
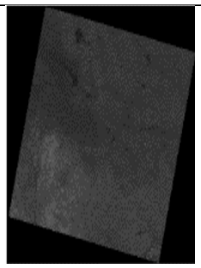
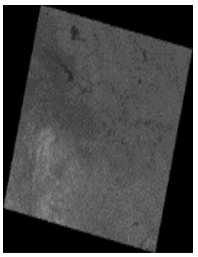
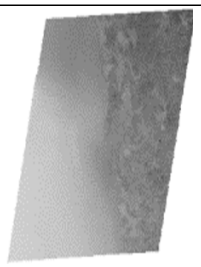
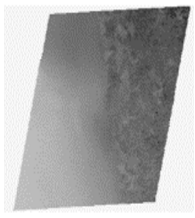
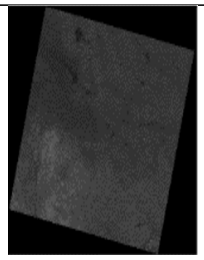
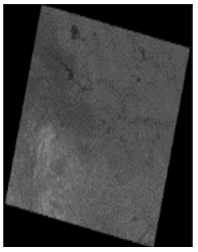
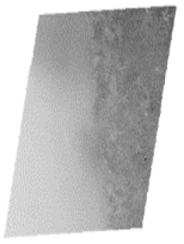
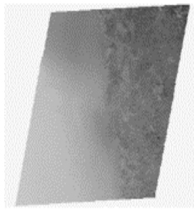
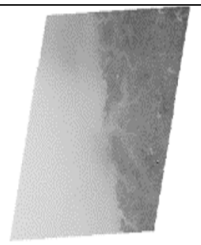
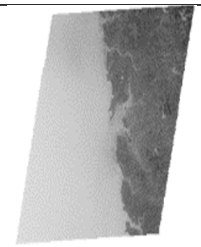
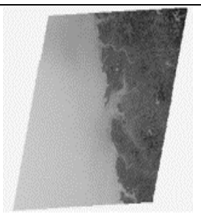
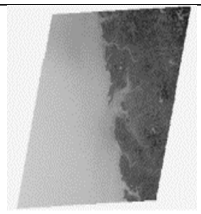
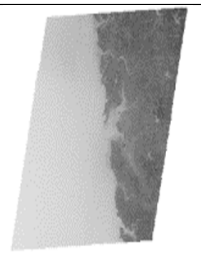
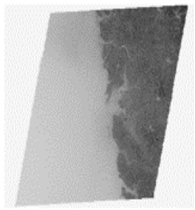
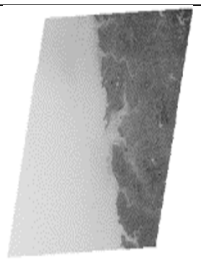
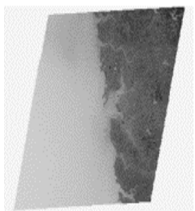


Figure 3: LULC classification for Konkan region

Preprocessing MSI and HSI input image is illustrated in FIGURE 4. To rectify the radiometric error, eight bands of a multispectral image are pre-processed using the QUAC approach. The SG smoothing technique is used to pre-process seventeen bands of hyperspectral pictures to improve contrast.

Bands	**Multispectral Image**		**Hyperspectral Image**	
	Input Image	**QUAC**	**Input Image**	**SG smoothing**
Band One				
Band Two				
Band Three				
Band Four				
Band Five				
Band Six				

Band Seven				
Band Eight				
Band Nine				
Band Ten				
Band Eleven				
Band Twelve				

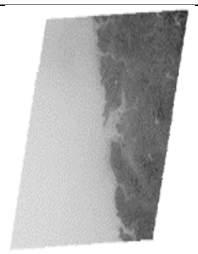
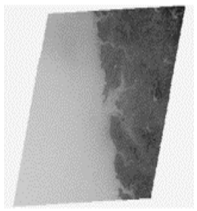
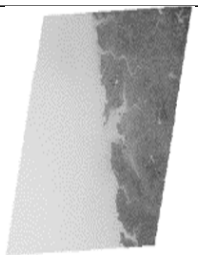
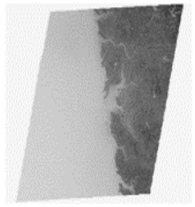
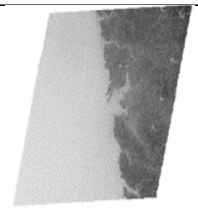
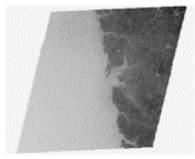
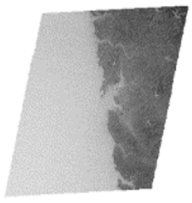
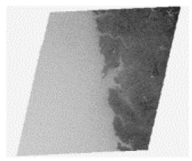
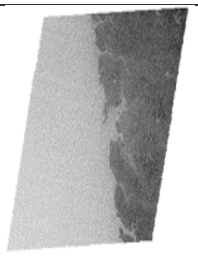
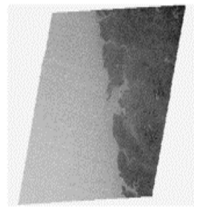
Band Thirteen				
Band Fourteen				
Band Fifteen				
Band Sixteen				
Band Seventeen				

Fig. 4: Pre-processing of Multispectral and Hyperspectral data

To create a single, high-resolution image, these processed bands are then delivered for stacking and merging technique. After performing radiometric correction and image enhancement, the merging step is carried out on both Multispectral and hyperspectral image. FIGURE 5 (a) represents the merged image of Multispectral image. Eight different bands of MSI are first stacked and then merged to produce a single high-resolution image. The merged image of the Hyperspectral image is illustrated in FIGURE 5 (b).

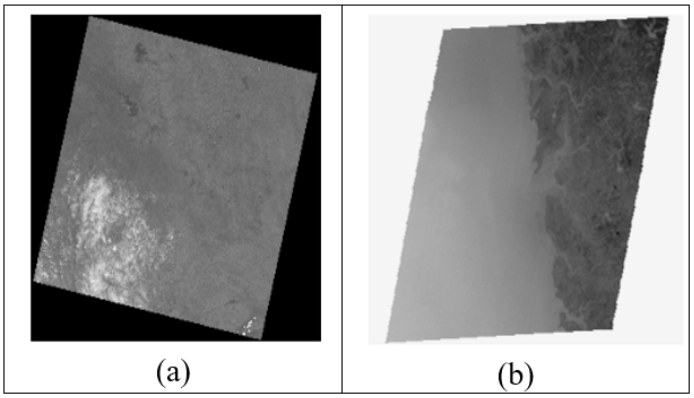


Figure 5: Merged images of (a) Multispectral (b) Hyperspectral

Seventeen different bands of hyperspectral images are merged to produce a single image. FIGURE 6 illustrates the feature extraction and fusion process of pre-processed image.

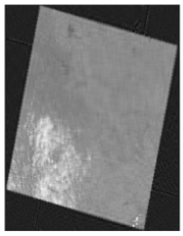
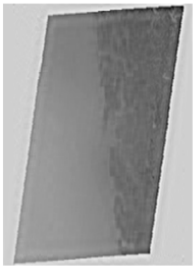
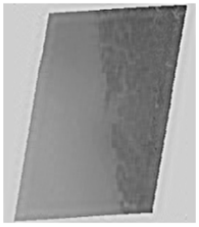
Image types	Feature extraction Image	Fusion Image
Multi spectral		
Hyper spectral		

Figure 6: Feature extraction and fused image

The suggested method uses a PCA-FFO strategy to accomplish the feature extraction process and fusion is achieved using hybrid NCTCP-DTCWT fusion. Inception-Resnet v2 classifier uses this fused image for the final prediction of various land coverings. FIGURE 7 represents the different indications of land covers. The green color in FIGURE 7a, represents the forest region present in the overall Konkan area. Similarly, yellow (FIGURE 7b), blue (FIGURE 7c), red (FIGURE 7d), and white (FIGURE 7e) color indicates the vegetation land, water body, built-up land, and bare land respectively. Whereas the remaining black color area represents other land area. FIGURE 8 presents the convergence plot of the FFO, where the fitness function is represented by the error.

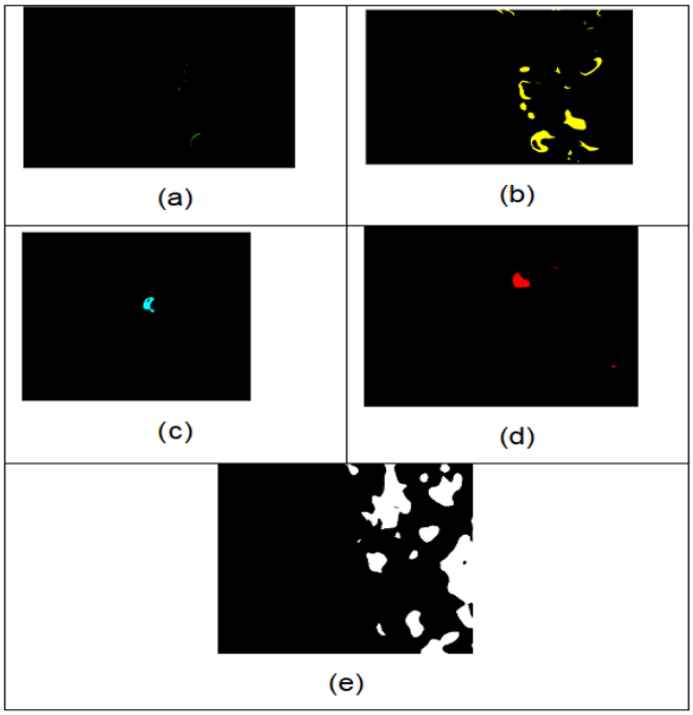


Figure 7: Different indications of land covers

Confusion matrix is shown in FIGURE 9. It is divided into five classes: built-up terrain, barren land, forest, vegetation, and aquatic bodies. Diagram indicates the low number of incorrectly predicted samples in every class. FIGURE 10 illustrates the ROC and AUC plots for the proposed approach. The technique demonstrates strong class discrimination capability, as evidenced by an AUC value approaching 1. The ROC curve, a probabilistic representation, reflects the model's effectiveness in distinguishing between different classes.

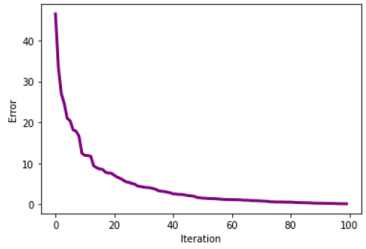


Figure 8: Convergence plot for the proposed FFO algorithm

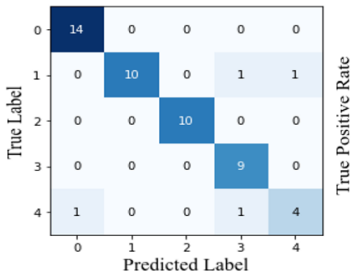


Figure 9: Confusion matrix

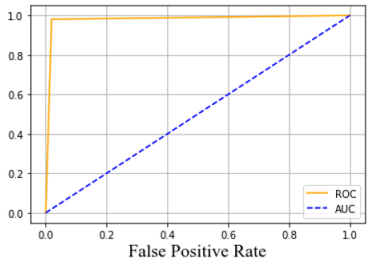


Figure 10: AUC and ROC plot for the proposed approach

The proposed Inception-ResNet v2 classifier is compared with several existing methods, including Alexnet, VGG16, Resnet50, and Inception-V2. The performance measures accuracy, precision, recall, specificity, NPV, FNR, FPR, and F1_Score are used to compare the proposed method with

the existing one. Comparison of proposed model with existing deep learning models is shown in FIGURE 11. This result shows that proposed model is superior and outperforms compared to existing machine learning models. Proposed model is better suited for Hyperspectral Image land classification than existing models because it effectively combines the advantages of both the Inception and ResNet architectures. The Inception module within Inception-ResNet V2 utilizes multiple filter sizes in parallel, allowing the model to capture a wide range of spatial and spectral features at different scales. This is particularly useful for hyperspectral data, which contains rich spectral information across many bands that can vary in spatial resolution.

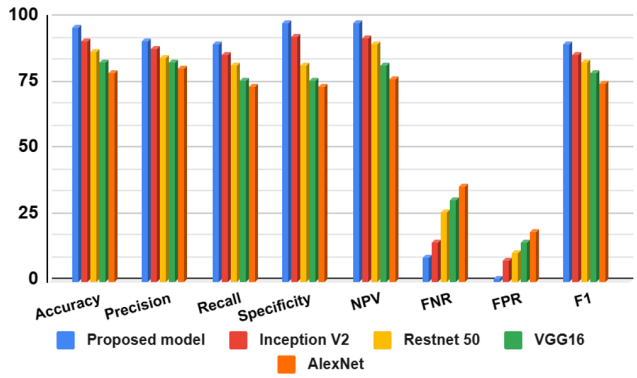


Figure 11: Comparison of deep learning models

Table 3: Comparison of deep learning models

Model	Acc	Prec	Rec	Spec	NPV	FNR	FPR	F1
	%	%	%	%	%			%
Inception Resnet V2 (Proposed model)	**96**	**91**	**90**	**98**	**98**	**9**	**1**	**90**
Inception V2	91	88	86	93	92	15	8	86
ResNet50	87	85	82	82	90	26	11	83
VGG16	83	83	76	76	82	31	15	79
AlexNet	79	81	74	74	77	36	19	75

The ResNet component introduces residual connections that allows proposed Inception-ResNet V2 to handle the high-dimensional data typical of hyperspectral images, which often include hundreds of spectral bands. Such deep architectures help capture complex patterns and subtle variations in land cover that simpler networks like AlexNet or VGG16 might miss due to their limited depth or inability to handle the intricacies of high-dimensional data. Additionally, proposed Inception-ResNet V2 benefits from faster convergence and better generalization compared to earlier models, making it more efficient in training. These combined features make Inception-ResNet V2 particularly effective in classifying land cover types from hyperspectral data. The proposed method uses Fused image (Landsat-8 Multispectral fused with Hyperspectral image). Multispectral image and hyperspectral image performance is compared with fused image. This result is illustrated in TABLE 3.

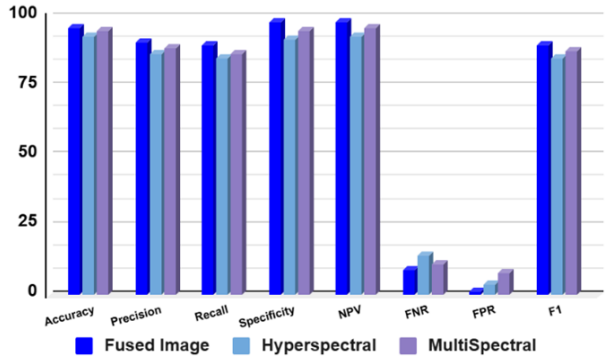


Figure 12: Comparison of HIS, MSI and fused image

Table 4: Comparison of HIS, MSI and Fused Image

Images	Acc	Prec	Rec	Spec	NPV	FNR	FPR	F1
	%	%	%	%	%			%
Fused Image	**96**	**91**	**90**	**98**	**98**	**9**	**1**	**90**
Hyperspectral Image	93	87	85	92	93	14	4	85
Multispectral Image	95	89	87	95	96	11	8	88

The performance metrics used to compare these images are accuracy, precision, recall, specificity, FPV, FNR, FPR, F1_Score. The comparison of HIS, MSI and fused image is shown in FIGURE 12. Results show that the fusion of Hyperspectral Imaging (HSI) and Multispectral Imaging (MSI) provides enhanced accuracy and interpretability in land cover classification tasks. By fusing HSI with MSI, we benefit from the complementary strengths of each: the fine spectral resolution provided by HSI and the sharp spatial details from MSI. This fusion leads to a more comprehensive understanding of the landscape, improving classification accuracy. The fusion allows the model to make more informed decisions, using both spatial and spectral data to classify land cover types more precisely.

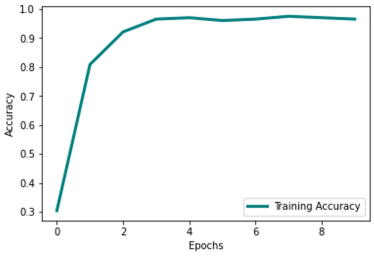


Figure 13: Training

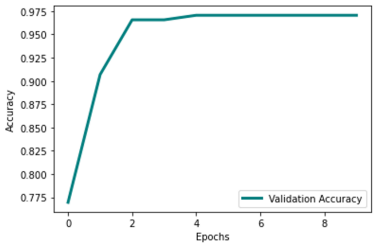


Figure 14: Validation

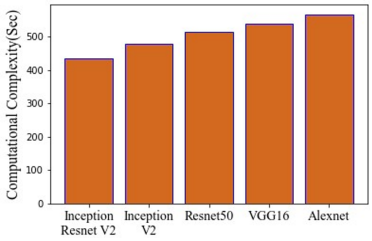


Figure 15: Computational complexity

The training accuracy and validation accuracy is illustrated in FIGURE 13 and FIGURE 14, respectively.

Validation accuracy is taken by varying epoch numbers. Through this analysis, it is suggested that techniques used in this proposed approach produce better outcomes when compared to other existing techniques. Time taken for Inception Reset V2 is 435 sec, which is low when compared to other approaches such as Inception V2 - 482 sec, Resnet 50 - 507 sec, VGG 16 - 536 sec and Alexnet - 597 sec. The time complexity is illustrated in FIGURE 15. The superior performance of the proposed framework compared to other deep learning models can be attributed to three main design choices: optimized band selection using PCA–FFO, hybrid NCTCP–DTCWT fusion, and the use of Inception–ResNet v2 for classification. PCA–FFO reduces spectral redundancy while preserv-

ing discriminative information, which stabilizes training and improves generalization. The fusion stage enhances spatial–spectral consistency, particularly for built-up and vegetation classes that exhibit similar spectral signatures in individual modalities. Inception–ResNet v2 further benefits from multi-scale feature extraction and residual learning, enabling more robust class separation. Results in TABLE 4 and FIGURE 12, clearly demonstrate that fused images outperform individual hyperspectral and multispectral inputs across all metrics.

5. FUTURE WORK

Future work could focus on integrating other remote sensing sources, such as LiDAR or Synthetic Aperture Radar (SAR), to further enhance land cover classification accuracy by providing complementary information on terrain and surface features. Additionally, exploring advanced deep learning models like Transformer networks, which are highly effective in capturing long-range dependencies in data, could improve model performance. Another avenue for future research involves leveraging transfer learning techniques, leading to faster convergence and improved results, particularly in regions with limited labeled data.

6. CONCLUSION

This research demonstrates how deep learning models are effective in classifying different land covers. This research also discusses the effectiveness of use of fused image in classifying different land covers. The study shows that the fusion technique outperformed compared to only the Multispectral or Hyperspectral dataset in a proposed deep learning technique. Though the proposed model performs better it faces certain difficult in selecting the number of band as it does not imply calculation of energy function based on spatial and spectral resolution. Also, it is validated using limited satellite images. This study can further be extended using AI based band selection techniques based on entropy of each band. Finally, the researchers can explore the methods used in the study on other different region images.

7. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

References

[1] Viana CM, Girão I, Rocha J. Long-Term Satellite Image Time-Series for Land Use/Land Cover Change Detection Using Refined Open-Source Data in a Rural Region. Remote Sens. 2019;11:1104.

[2] Saing Z, Djainal H, Deni S. Land Use Balance Determination Using Satellite Imagery and Geographic Information System: Case Study in South Sulawesi Province, Indonesia. Geod

Geodyn. 2021;12:133-147.

[3] Allam M, Bakr N, Elbably W. Multi-Temporal Assessment of Land Use/Land Cover Change in Arid Region Based on Landsat Satellite Imagery: Case Study in Fayoum Region, Egypt. Remote Sens Appl Soc Environ. Apr. 2019;14:8-19.

[4] Bhat IF. Impact of Land-Use Land-Cover Changes on Ecosystem Services of Jammu and Kashmir, India. Agro-Econ. 2022;9.

[5] Alam A, Bhat MS, Maheen M. Using Landsat Satellite Data for Assessing the Land Use and Land Cover Change in Kashmir Valley. GeoJournal. 2019;85(6):1529-1543.

[6] Carranza-García M, García-Gutiérrez J, Riquelme JC. A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks. Remote Sens. 2019;11:274.

[7] Talukdar S, Singha P, Mahato S, Shahfahad, Pal S, et al. Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review. Remote Sens. 2020;12:1135.

[8] Ali U, Esau TJ, Farooque AA, Zaman QU, Abbas F, et al. Limiting the Collection of Ground Truth Data for Land Use and Land Cover Maps With Machine Learning Algorithms. ISPRS Int J Geo-Inf. 2022;11:333.

[9] Digra M, Dhir R, Sharma N. Land Use Land Cover Classification of Remote Sensing Images Based on the Deep Learning Approaches: A Statistical Analysis and Review. Arab J Geosci. 2022;15.

[10] Fedoseeva NV, Lvov AL. Principal Component Analysis Applied to Satellite Imagery for Volcanic Plume Detection Over Kuril and Kamchatka Region. Eur J Nat Hist. 2022:4.

[11] Rai AK, Mandal N, Singh A, Singh KK. Landsat 8 OLI Satellite Image Classification Using Convolutional Neural Network. Procedia Comput Sci. 2020;167:987-93.

[12] Halbe A, Bhirud S. QuantumHSIBands: A Quantum Enabled Band Selection Technique in Hyperspectral Images. Int J Intell Eng Syst. Aug. 2024;17:188-198.

[13] Jamali A. Land Use Land Cover Mapping Using Advanced Machine Learning Classifiers: A Case Study of Shiraz City, Iran. Earth Sci Inform. 2020;13:1015-1030.

[14] Merga BB, Moisa MB, Negash DA, Ahmed Z, Gemeda DO. Land Surface Temperature Variation in Response to Land-Use and Land-Cover Dynamics: A Case of Didessa River Sub-Basin in Western Ethiopia. Earth Syst Environ. Mar. 2022;6:803-815.

[15] Ahmed S. Comparison of Satellite Images Classification Techniques Using Landsat-8 Data for Land Cover Extraction. Int J Intell Computing Inf Sci. 2021:1-15.

[16] Yadav CS, Pradhan MK, Gangadharan SM, Chaudhary JK, Singh J, et al. Multi-Class Pixel Certainty Active Learning Model for Classification of Land Cover Classes Using Hyperspectral Imagery. Electronics. 2022;11:2799.

[17] Gopinath G, Sasidharan N, Surendran U. Land Use Classification of Hyperspectral Data by Spectral Angle Mapper and Support Vector Machine in Humid Tropical Region of India. Earth Sci Inform. 2020;13:633-640.

[18] Basith A, Prastyani R. Evaluating Acomp, Flaash and Quac on Worldview-3 for Satellite Derived Bathymetry (Sdb) in Shallow Water. Geod Cartogr. 2020;46:151-158.

[19] Li D, Shen X, Guan H, Yu Y, Wang H, et al. AGFP-Net: Attentive Geometric Feature Pyramid Network for Land Cover Classification Using Airborne Multispectral LiDAR Data. Int J Appl Earth Obs Geoinf. 2022;108:102723.

[20] https://bhuvan-app3.nrsc.gov.in/data/download/index.php

[21] https://earthexplorer.usgs.gov/

[22] https://daac.ornl.gov/VEGETATION/guides/Decadal_LULC_India.html

[23] Qiao X, Liu X, Wang F, Sun Z, Yang L, Pu X et al. A Method of Invasive Alien Plant Identification Based on Hyperspectral Images. Agronomy. 2022;12:2825.

[24] Trojovska E, Dehghani M, Trojovsky P. Fennec Fox Optimization: A New Nature-Inspired Optimization Algorithm. IEEE Access. 2022;10:84417-84443.

[25] Xu Y, Wu Z, Chanussot J, Comon P, Wei Z. Nonlocal Coupled Tensor Cp Decomposition for Hyperspectral and Multispectral Image Fusion. IEEE Trans Geosci Remote Sens. 2020;58:348-362.

[26] Huan W, Li S, Qian Z, Zhang X. Exploring Stable Coefficients on Joint Subbands for Robust Video Watermarking in DT CWT Domain. IEEE Trans Circuits Syst Video Technol. 2022;32:1955-1965.

[27] `https://www.geeksforgeeks.org/machine-learning/inception-v4-and-inception-resnets/`

[28] Alruwaili M, Shehab A, Abd El-Ghany S. COVID-19 Diagnosis Using an Enhanced Inception-ResNetV2 Deep Learning Model in CXR Images. J Healthc Eng. 2021;2021:6658058.