IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Novel Framework For Image Classification Using Hybrid Cnn-Gan-Based Masking

[1]KADAPALA ANJAIAH and [2]Dr.K SAGAR

[1]Research Scholar in University College of Engineering Osmania University Hyderabad, Telangana, India

[2]Principal & Professor of CSE Geethanjali College of Engineering and Technology Hyderabad, Telangana, India

Abstract: This research evaluates object identification systems using machine learning techniques for processing satellite photos. It uses transfer learning and multi-object identification deep learning algorithms on remotely sensed satellite data. Five object identification techniques based on YOLO R-CNN architectures were tested, with YOLOv8 showing the fastest detection speed and highest accuracy. Support vector machines were found to be more useful than k-means clustering for object detection on satellite photos. The study also proposes an efficient 3-D CNN model using overlapping 3-D patches and a kernel function, achieving higher accuracy and faster convergence than existing 2-D/3-D CNN methods.

Index Terms **- Satellite Images, Object Recognition, k-Means Clustering, Support Vector Machines,3DCNN, Hyperspectral Images [HSIs], classification, kernel function**

I. INTRODUCTION

Basically, Generative Adversarial Networks (GANs) consist of two neural networks: a generator and a discriminator. GANs have the ability to reproduce samples. A Generative Adversarial Network (GAN) is used to do Semi-Supervised Learning (SSL) based on the Hilbert-Schmidt Independence Criterion (HSIC), using spectral features.

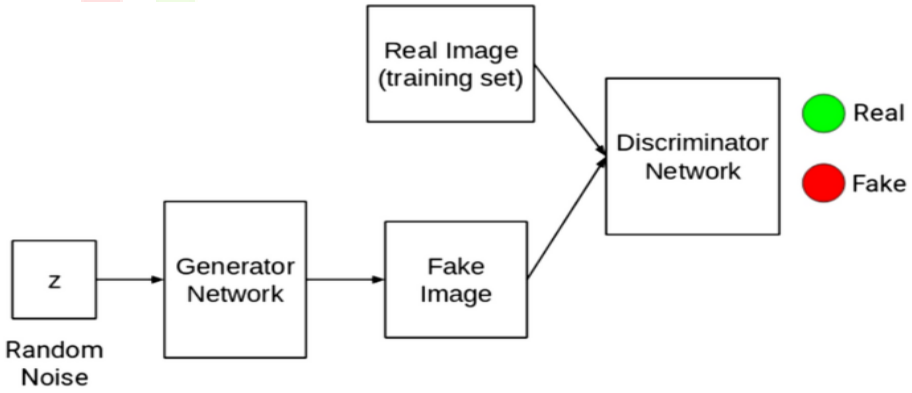

Fig 1 A general architecture GAN

Generative Adversarial Networks (GANs) are neural networks that reproduce samples using spectral and spatial information. The framework combining GANs, Convolutional Neural Networks, and a majority voting technique was developed.in this approach combines high-level contextual information and uses an auxiliary classifier to generate more coherent virtual training samples. In this the solution addresses the issue of unbalanced data distribution across different classes.

Transfer learning

The ability of a model to complete a secondary task by drawing on knowledge from a related primary job. In other words, data from one another to acquire new information that has never been labeled or observed before. Hence, transfer learning may be efficiently used in fields were depending on the availability of labelled training examples. Typically, it is believed that the source and target domains are correlated but not identical. However, it is possible for them to have distinct distributions, as is the case with HSIC.

The model learns features hierarchically in hierarchical sparse information coding (HSIC) based on deep neural networks (DNN). When trained on a variety of images, the model's bottom layers typically extract generic characteristics. As a result, a novel classifier for the intended dataset may be trained using the knowledge that these layers have acquired. As an illustration, we employed that was extensively trained using data from other HSIs. To accurately classify the target HSI, we then applied the lowest layers of this pre-trained model to the target network. The target network's upper layers are randomly initialized to gather information about its distinctive features. A small number of labelled examples from the target HSI dataset are then used to adjust the entire network. Furthermore, a workable method for pre-training and optimizing a CNN network was proposed in order to efficiently categorize new HSIs. The research combined transfer learning and data augmentation techniques in an effort to improve HSIC performance. Data in the source and target domains may vary in a number of ways, as was previously mentioned. When it comes to hyperspectral images (HSIs), for instance, two HSIs may have different sizes since they were taken using different sensors. Heterogeneous transfer learning refers to the process of managing variances across different domains and transferring information across them.

Active learning

Using an unlabeled sample pool, Active Learning (AL) actively increases the classifier's predicted accuracy. Active learning (AL) selects the data during each iteration, thereby improving the training dataset. Assigned to these instances by an oracle, which could be a computer or a person.

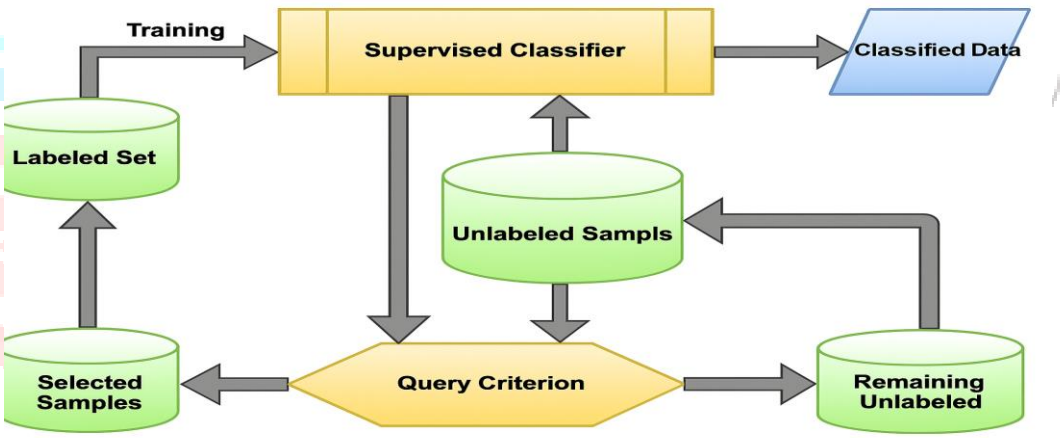

Fig 2: A general overview of Active Learning

Improve accuracy. Stream-based and pool-based AL frameworks can be distinguished based on how they incorporate new cases into the training set. Ranking scores that are determined by various criteria in order to evaluate the utility of the sample.

When using pool-based selection, it is crucial to provide a variety of samples in the pool to prevent duplication. When choosing or querying useful samples, the attention is often on three aspects: the heterogeneous behaviour, the performance of the model, and the representativeness of the samples. Below is a concise overview of several sampling methodologies. One possible approach to address this issue might include examining the incorporation of different learning processes outlined in section 3.10 in order to capitalise on the collective advantages. Another method involves using few-shot or K-shot learning techniques, which may make precise predictions of class labels using just a small number of labelled data. Additionally, it is necessary to prioritise the combined use of spectral-spatial characteristics of hyperspectral imagery (HSI)

Therefore, including the aforementioned factors into the creation of a new HSIC framework involves effectively using the limited training samples, taking into account the combined spectral-spatial properties of HSI, and ensuring minimal computing load.

II. TYPE STYLE AND FONTS

The precise identification of objects in remotely sensed pictures is essential for mapping and monitoring socioeconomic and biophysical aspects [18-21]. A technique for identifying buildings from images was presented by Wang et al. [22] and included combining a CNN with an LSTM network. The pictures' features were extracted using CNN, and the spatial connections between the features were modeled by LSTM. Furthermore, a strategy based on deep learning was put out to identify ships from synthetic aperture radar (SAR) pictures [23]. The suggested technique beat various deep learning systems that were already in use, according to the findings, which were obtained by using a region proposal network (RPN) and CNN to extract features from the photos [24-28]. In a factory, the model was used to identify objects in real time. An Auto-T-YOLO version of YOLOv4 was suggested by Sun et al. [29] to recognize objects in pictures.

Faster R-CNN achieved 86.0% accuracy at 12 fps on the AGs-GF1 dataset but showed lower classification accuracy.

YOLOv3 improved performance with 90.4% accuracy and 73 fps on AGs-GF1, though its classification rate was slower.

YOLOv4, tested on the MS-COCO dataset, obtained 44.5% accuracy and 64.7 fps but required higher computational power.

YOLOv5s achieved 5.9 ms inference time and 61.8 mAP on the SIMD dataset but had lower recognition speed and precision.

III. DATA SET AND FEATURES

Dataset Creation

The target detection process relies on the development of high-quality datasets for deep learning models. Roboflow, a vision platform, was used to organize, label, and compile a dataset of 92 satellite photos, categorized into training, testing, and validation sets. The dataset was divided into training, testing, and validation sets, with 61 photos and one annotation file for training, 21 photos and one annotation file for validation, and 10 pictures for testing. To improve the model's resilience and accuracy, geometric modifications and augmentations were used. The NSL-KDD dataset, an improved version of the original KDD99 dataset, was used to assess the model's performance. The dataset includes 42 attributes and five classes for multi-category attack prediction.

IV. METHODOLOGY

Praposed System

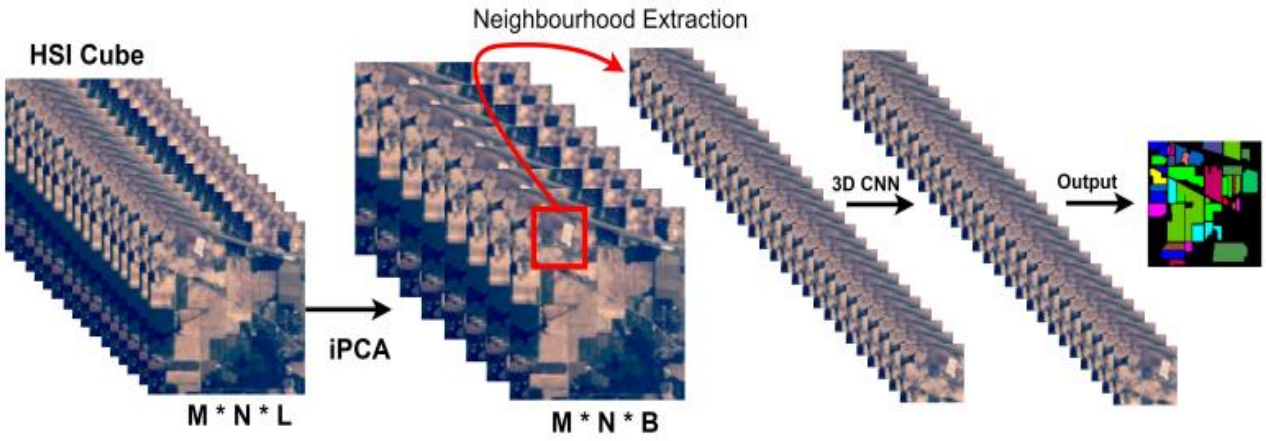

Fig 3. Process flow diagram.

Assuming a Hyperspectral Image HSI pixels demonstrate significant similarities across different classes, substantial variety within the same class. Additionally, there are instances when areas overlap and are nested inside each other. In order to address the aforementioned problems, the technique of independent principal component analysis (iPCA). To transfer the HSI cube to the model, it is necessary to partition it into tiny 3D

spatial patches that overlap. The ground labels are then generated based on the centre pixel of each patch. Subsequently, these acquired characteristics are subjected to an activation function, which creates nonlinearity.

Table 1: 3D CNN

Layer	Output Shape	Parameters
Input Layer	(11,11,20,1)	0
Conv3D-1 (Conv3D	(9,9,14,8)	512
Conv3D-2 (Conv3D	(7,7,10,8)	5778
Conv3D-3 (Conv3D	(5,5,8,32)	12365
Conv3D_4 (Conv3D	(3,3,6,64)	52369
Flatten_1 (Flatten)	(3456)	0
Dense_1 (Dense)	(256)	884992
Dropout_1 (Dropout)	(256)	0
Dense_2 (Dense)	128	32897
Dropout_2 (Dropout)	128	0
Dense_3 (Dense)	Classes	774

The proposed 3D CNN model uses four convolutional layers to improve spatial-spectral feature maps, ensuring accurate differentiation across different spectral bands. It has 994,166 tune-able weights and exhibits rapid convergence, but may result in overfitting. Regularization strategies and alternative targets have shown superior performance. Image processing methods can enhance information quality by analyzing satellite pictures at varying resolutions. The fusion-based classification process consists of five stages: image enhancement, fusion, segmentation, feature extraction, and classification. The evaluation is conducted using fused pictures from the Landsat collection.

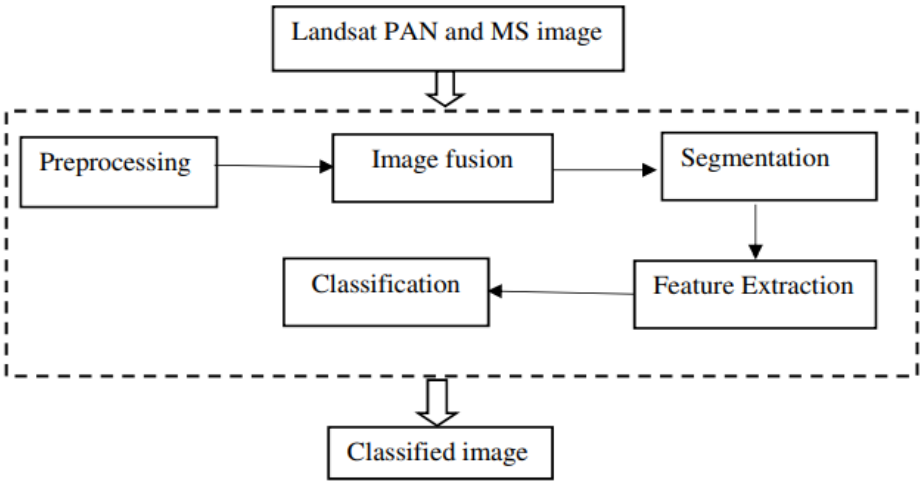

Fig 4 Block diagram of fusion based satellite image classification

Equations

An HSI is represented as X∈RL×(N×M)X \in \mathbb{R}^{L \times (N \times M)}X∈RL×(N×M), where each of the N×MN \times MN×M pixels has LLL spectral bands and a class label. To reduce redundancy, iPCA compresses bands from LLL to B≪LB \ll LB≪L without altering spatial dimensions. The HSI is divided into overlapping 3D patches P∈RS×S×BP \in \mathbb{R}^{S \times S \times B}P∈RS×S×B, centered on each pixel. Each patch is labeled by its center. The total number of patches is (M−S+1)×(N−S+1)(M-S+1)\times(N-S+1)(M−S+1)×(N−S+1).

$$v_{i,j}^{x,y,z} = \mathcal{F}\left(\sum_{\tau=1}^{d_{i-1}} \sum_{\lambda=-\nu}^{\nu} \sum_{\rho=-\gamma}^{\gamma} \sum_{\phi=-\delta}^{\delta} w_{i,j,\tau}^{\lambda,\rho,\phi} \times v_{(i-1),\tau}^{(x+\lambda),(y+\rho),(z+\phi)} + b_{i,j} \right)$$

Patches are passed through a 3D convolution layer, followed by activation to extract nonlinear features. The output at position (x,y,z)(x, y, z)(x,y,z) in layer iii, feature map jjj, is vi,jx,y,zv^{x,y,z}_{i,j}vi,jx,y,z.

EXPERIMENT MODEL EVALUATION WITH RESULTS

This section provides an overview of several benchmark actual hyperspectral imaging (HSI) datasets that were used in this research. Various airborne and satellite sensors are being used to collect these datasets. Additionally, there are the Pavia University and Pavia Centre datasets collected using the ROSIS sensor, the Botswana dataset acquired using the Satellite Hyperion sensor. The evaluation of information from ROSIS and Hyperion sensors poses a more difficult classification challenge compared to AVIRIS sensor datasets. This is primarily due to the presence of complex urban classifications and nested areas.

Table 2: Summary of the HSI datasets.

Dataset	**Dimension**	**Classes**	**Wave-length**
Botswana	512 X 217	13	400-2500
Landsat	1496 × 256	16	400-2500
Bhuvan	610 X 610	12	360-2600
Google Earth	1096 X 1096	9	430-860
QuickBird	145 X 145	9	430-860
Indian Pines	610 X 610	8	400-1260

The Training samples, which make up 70% of the total population, are then separated into a 50/50% ratio for the Training and Validation sets.

Verifying the Precision of Class Categories

1: The Check Accuracy algorithm
2: Set up a counter variable to keep track of the number of correctly and incorrectly classified images.
3: Retrieve the first assigned classifications for the picture
4: Retrieve the anticipated label outcomes for the same picture acquired from the classifiers.
Iterate 5 times,
5: starting from 1 and ending at the maximum number of blocks for the picture.
6: For every class category.
7: if the original label and anticipated label are the same
8: Increase the counter for accurately classifying the related class.
9: otherwise
10: Increase the number for incorrectly classifying the appropriate class
11: Terminate the if statement.
12: Terminate the for loop.
13: Compute the % accuracy for each class category individually.

Compute the mean accuracy across all class categories.

The event concludes at 15:00. Based on these numbers, it can be inferred that the suggested model reached convergence at around 20 epochs.

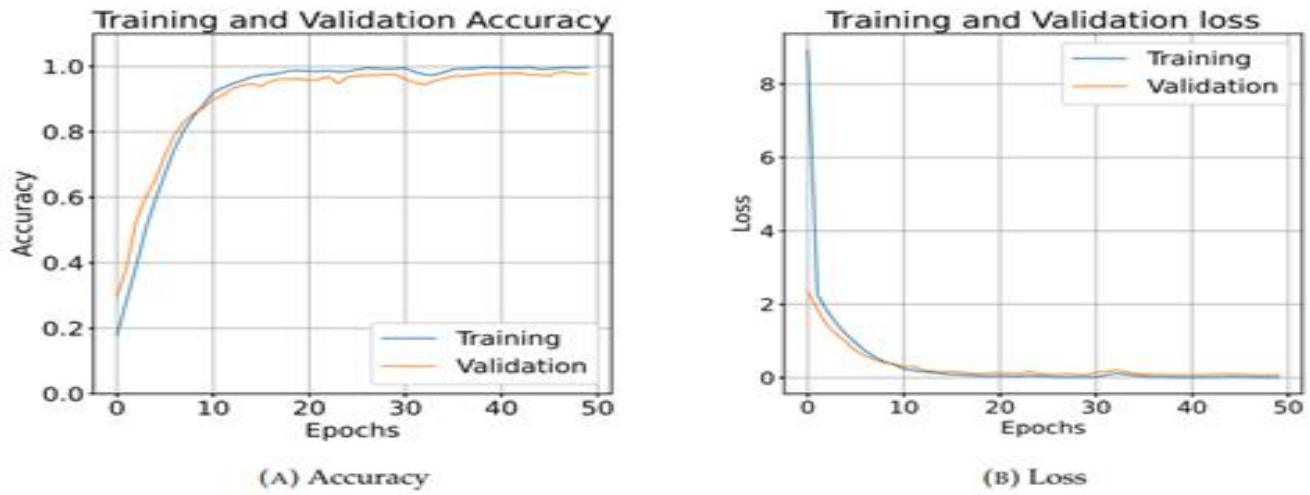

Fig 5: The accuracy and loss values over 50 epochs.

The experimental findings presented in this study were acquired using Google Colab, an online platform that allows for the execution of Python environments with high-speed internet connectivity. Google Colab offers the ability to run several versions of Python, access to a Graphical Processing Unit (GPU).

All models, including hybrid models, have a learning rate of 0.001 to guarantee a fair comparison. With the exception of the output layer, which uses Softmax as its activation function, all other layer's use Relu. In order to minimize computational load, the 15 most informative bands for each experiment were chosen using PCA, with the patch size set to 15. A 5-fold cross-validation procedure is used to select the training, validation, and test sets for all experimental results. Specifically, 25%, 25%, and 50% of the samples are assigned between the various classes in the samples if the patch size is too small. Because of this, the final result in both cases will have a higher percentage of misclassification, which will lead to subpar generalization performance. As a result, before the final experimental setup, a suitable patch size must be determined see in fig 6. Accuracy and Loss for 50 epochs on IP for Training and Validation sets.

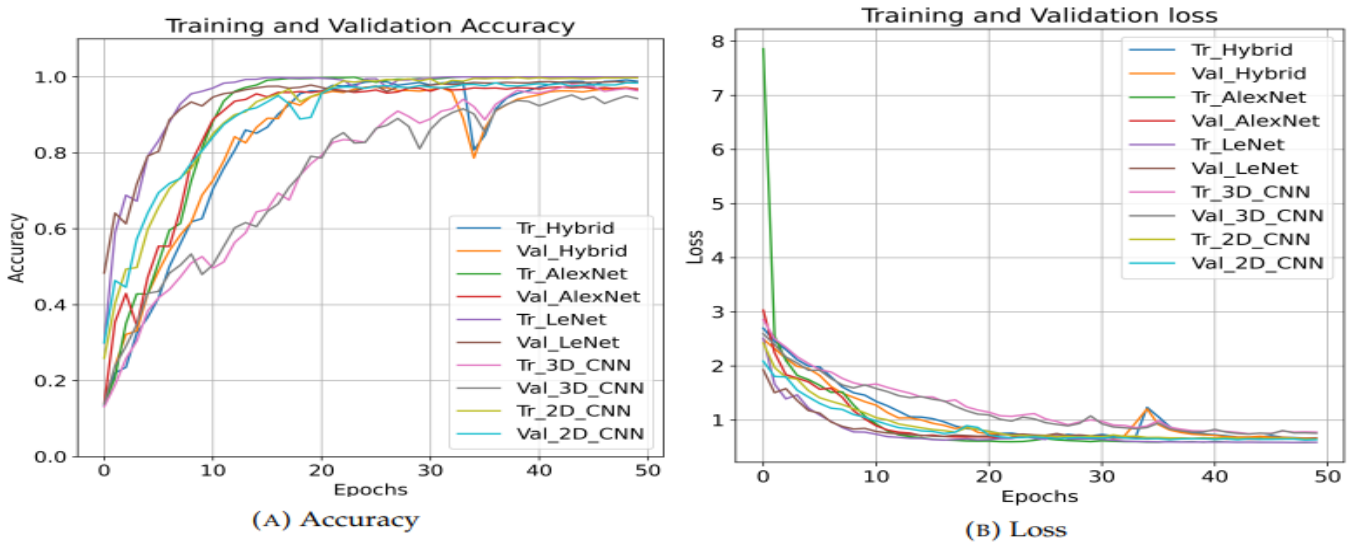

Fig 6: Accuracy and Loss for 50 epochs on IP for Training and Validation sets.

A set of criteria is used to determine the patch size that is used in these studies.

Precision =TP /(TP+ FP)

Recall =TP /(TP+ FN)

Accuracy =TP+ TN /(TP+ FP+ TN + FN)

F-measure = 2×Precision×Recall /(Precision+ Recall)

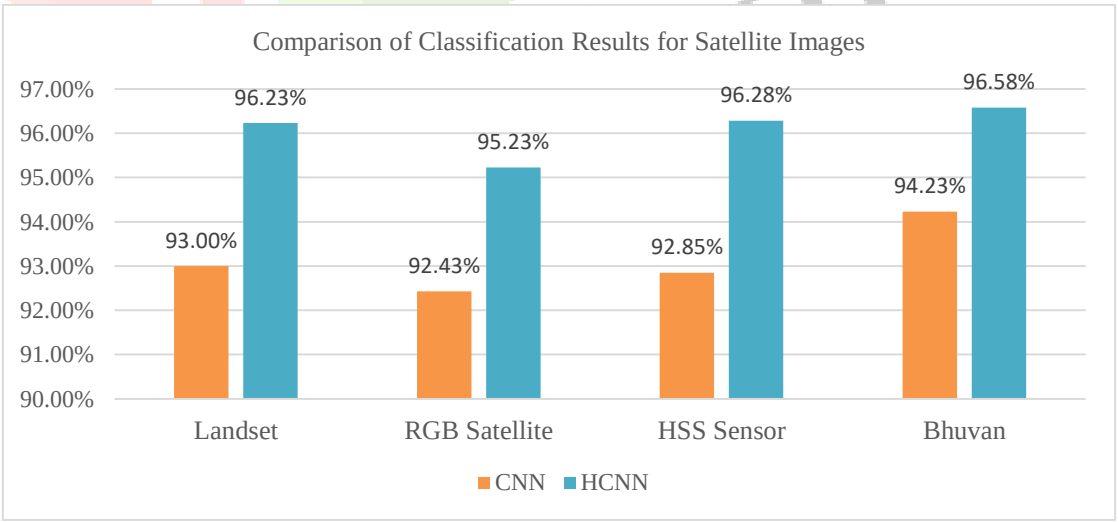

Fig .7: Accuracy for Multiple data set using CNN and HCNN

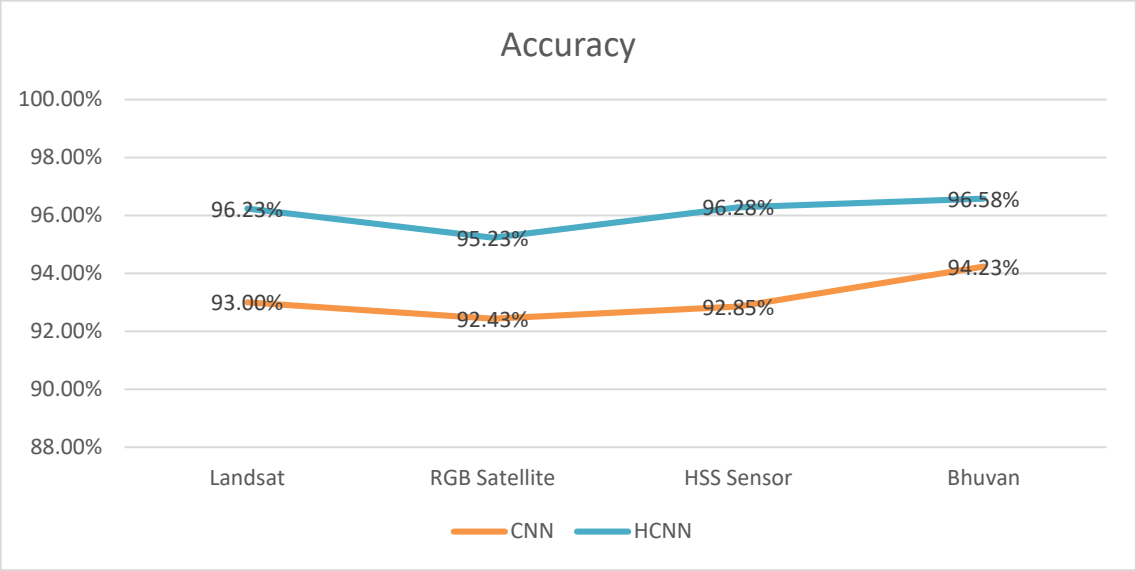

Fig .8.: Accuracy for Multiple data sets

The Gesture Recognition Project has developed a functional and interactive system that can instantly recognize and interpret hand gestures using computer vision and deep learning models. The system integrates the PyQt5 graphical user interface (GUI) for improved user experience. It can translate gestures like "hello," "thank you," and "I love you," making it an essential communication tool for people with hearing or speech impairments. However, issues like lighting, background noise, and hand placement still affect the system's accuracy. The technology has potential applications in healthcare, gaming, education, and augmented reality. The system is ideal for future assistive technology, security, and IoT devices. Improvements include expanding the model to recognize more descriptive information, solving environmental issues, exploring depth sensors or infrared cameras, and integrating with IoT devices and smart home devices. The project has the potential to change the way we interact with technology by addressing current limitations and exploring new applications.

CONCLUSION

The study demonstrates that classification accuracy in satellite image analysis can be significantly improved through a block-based approach and effective feature extraction techniques. The integration of PCA and GLCM enables the capture of both color and texture information, essential for handling the complexity of high-resolution images. Additionally, careful selection of block size and feature quantity plays a critical role in optimizing performance. Overall, the results validate that combining spatial segmentation with holistic feature extraction enhances the efficiency and accuracy of remote sensing classification tasks. Classification performance is evaluated using various classifiers on 8×8 pixel blocks, with accuracy influenced by feature type, number, computation time, and classifier settings. PCA (color) and GLCM (texture) improve results, especially when combined. Block-based methods enhance accuracy in satellite image classification, with block size and feature count playing key roles.

REFERENCES

[1] K. Vani, “Satellite Image Processing,” IEEE, Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), vol. 1, no. 4, p. 53,2017.

[2] Kadapala Anjaiah and K. Sagar, "Transfer learning in conjunction with multi-object detection using YOLO RCNN," The International Journal of Interdisciplinary Organizational Studies, vol. 19, no. 1, pp. 897–907, Jan.–Jun. 2024.

[3] Muhammad Ahmad, Adil Mehmood Khan, Manuel Mazzara, Salvatore Distefano, Mohsin Ali, and Muhammad Shahzad Sarfraz,
“A Fast and Compact 3-D CNN for Hyperspectral Image Classification,” IEEE Geoscience and Remote Sensing Letters, vol. 19, DOI: 10.1109/LGRS.2020.3043710, 2020

[4] Kadapala Anjaiah and K. Sagar, "Using Generative Adversarial Networks (GAN) to correct banding errors in satellite images," International Journal of Techno-Engineering (IJTE), vol. XVI, no. II, pp. 242–253, May 2024. [Online]. Available: http://ijte.uk/

[5] M. Kumar, "Digital image processing," Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, pp. 81–102, 2004.

[6] M. Kumar, "Digital Image Processing of Remotely Sensed Satellite Images for Information Extraction," Proceedings of the Conference on Advances in Communication and Control Systems-2013, pp. 406–410, 2013.

[7] N. Kumar, "Digital Image Processing for Image Enhancement and Information Extraction," Society of Photo-Optical Engineers, vol. 199, pp. 9–19, 1980.

[8] R. Aedla, "Satellite Image Contrast Enhancement Algorithm based on Plateau Histogram Equalization," IEEE Region 10 Symposium, pp. 213–218, 2014.

[9] S. S. Al-amri, N. V Kalyankar, and S. D. Khamitkar, "Contrast Stretching Enhancement in Remote Sensing Image," BIOINFO Sensor Networks, vol. 1, no. 1,pp. 6–9, 2011.

[10] P. Bidwai and D. J. Tuptewar, "Resolution and contrast enhancement techniques for grey level, color image and satellite image," Proceedings - IEEE International Conference on Information Processing, ICIP 2015, pp. 511–515, 2016.

[11] P. Kaushik and Y. Sharma, "Comparison of different image enhancement techniques based upon Psnr & Mse," International Journal of Applied Engineering Research, vol. 7, no. 11, pp. 2010–2014, 2012.

[12] T. Blaschke, "Object based image analysis for remote sensing," ISPRS Journal of Photogrammetry Remote Sensing, vol. 65, no. 1. pp. 2–16, 2010.

[13] W. Xiaoyun, Y. Weiqi, and C. V. Group, "Human Ear Recognition Based on Block Segmentation," IEEE, pp. 262–266, 2009.

[14] X. Sun, L. Zhang, H. Yang, T. Wu, Y. Cen, and Y. Guo, "Enhancement of Spectral Resolution for Remotely Sensed Multispectral Image," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 5, p

[15] I. Lee and T. Choi, "Accurate Registration using Adaptive Block Processing for Multi-spectral Images," IEEE Transactions On Circuits And Systems For Video Technology, no. c, pp. 1–11, 2013.

[16] A. G. Veenadevi.S.V, "Fixed Range Block Segmentation and Classification for Fractal Image Compression of Satellite Imageries," IEEE Fifth International Symposium on Electronic System Design, no. 1, pp. 1–4, 2014.

[17] R. R. S OudayaCoumar, R.Aravindraja, S.Arulambalam, Naaraayan and R. S. P.Prasad, "Contrast Enhancement Of Satellite Images Using Advanced Block Based DWT Technique," IEEE International Conference On Recent Trends In Information Technology, 2016.

[18] L. I. Smith, "A tutorial on Principal Components Analysis," Computer Science Technical Report No. OUCS-2002-12, 2002.

[19] T. Celik, "Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and k-Means Clustering," IEEE Geoscience And Remote Sensing Letters, vol. 6, no. 4, pp. 772–776, 2009.

[20] C. Munyati, "Use of Principal Component Analysis (PCA) of Remote Sensing Images in Wetland Change Detection on the Kafue Flats , Zambia," Geocarto International, vol. 19, no. 3, pp. 11–22, 2008.