Focal-Hinge Loss based Deep Hybrid Framework for imbalanced Remote Sensing Scene Classification

NEHA KUMARI[1]
SONAJHARIA MINZ[2]

Spatial and Data Mining Lab, School of Computer and Systems Sciences,
Jawaharlal Nehru University, New Delhi - 110067, India
[1]neha742@gmail.com
[2] sona.minz@gmail.com

Abstract. Deep learning models have received a significant breakthrough in remote sensing scene classification due to their discriminative, hierarchical feature extraction ability. Nevertheless, CNN-based methods deliver accurate classification results only with sufficient annotated training samples. The computational bottleneck of CNNs with numerous parameters in case of inadequate training samples and the inherent class imbalance problem in high-resolution satellite scene classification question the classifier's performance. Existing deep CNNs with conventional Cross-Entropy loss function neglected the significance of gradient contribution from minority classes in handling imbalanced LULC class distribution. In this context, we propose the hybrid probabilistic gradient-based deep learning framework CNN-FHSVM with regularized novel Focal-Hinge loss cost function optimization for alleviating misclassifications in imbalanced datasets. The empirical experimentation with Sentinel 2 EuroSat Dataset benchmarked for deep learning algorithms demonstrated that the proposed model is superior in mitigating classification errors in imbalanced class distribution contrasted to the cutting-edge deep learning frameworks. The proposed loss function adaptively updates the gradient of the minority classes, drifting the focus to misclassified scenes. Focal-hinge loss is the first endeavor adapted to remote sensing LULC multiclass classification to reduce misclassifications. The model demonstrates higher accuracy with reduced misclassifications and training time and can benefit other remote sensing applications like early deforestation urban planning, where LULC maps are imbalanced.

Keywords: Class Imbalance, Convolutional Networks, Focal HingeLoss, Gradient Learning, Scene Classification.

(Received July 19th, 2023 / Accepted July 1st, 2024)

1 Introduction

Scene Classification has been a burgeoning domain of research investigation in the realm of computer vision for decades, with extensive real-world applications. Nevertheless, the classification of scenes is noticeably more complex due to the semantic similarity of the ground objects between the scenes and the diversity in scales, distribution, and illumination effect of objects within the class, leading to misclassifications[3, 13, 24, 27, 33, 25]. With an emerging paradigm shift in high-resolution data availability in Earth Observation, scene classification in Remote Sensing (RS) has gained momentum with real-world applications in deforestation, urban planning, vegetation mapping, etc. It encompasses the task of assigning semantic labels to patches cropped from imagery acquired by sensors[3]. In applications like urban planning early stages of deforestation, LULC maps are generally imbalanced, and the misclassifications from minority classes are higher[23].

Capturing contextual and spatial intricacies of minority classes is also very significant in LULC maps in RS applications. The heterogeneity of land cover classes with the complex spatial distribution of inherent ground objects makes scene classification arduous[3, 27]. In the Machine Learning Community, it is the prevalent imbalanced learning problem having an adverse impact on classification accuracy[12] Deep Learning (DL) algorithms in RS image classification have achieved significant breakthroughs, offering novel approaches in recent years with the advent of extraordinary computational GPUs[2]. The deep network's multiple stacked feature extraction stages are imperative for hierarchical feature representations of objects in scene classification[2, 10, 29]. Based on the author's insights from the literature, Deep CNNS has yielded remarkable outcomes in RS scene classification[3, 10, 13, 29, 33, 27, 24, 6, 5]. Nonetheless, the imbalanced distribution of land cover classes with the co-existence of ground objects in scenes leads to misclassification errors from minority classes[24] Researchers have endeavored to reduce misclassifications through the feature extraction stages of the scene classification pipeline[3]. Still, gradient learning in comprehending the essential aspects of minority classes is neglected in reducing misclassification.

The present research demonstrates the use of loss function with probability-based gradient learning to reduce misclassifications, specifically from minority classes. Before accentuating the significant contributions of the current work, it is essential to mathematically formulate the problem of misclassification in the context of the RS scene. Classification of scenes is a decision problem. The problem statement is described in section 1.1

1.1 Problem Statement

Given remote-sensing imagery, I, comprising patches ρ_i, ordered as a tensor, Assuming training Set s with K discrete classes ,$y = \{y_1, y_2, \ldots, y_k\}$, a set $D_{\text{train}} =$ patches$[i] = \{\rho_0, \rho_1, \rho_2, \ldots, \rho_n\}$, and with the vector, $[a_0, a_1, a_2, \ldots, a_n]$,as features $\forall P_i \in D$. Hence, the Classifier is mapped as $f : \rho_i \rightarrow Y$ here, and the Training Set is defined in (1)

$$(X_{\text{train}} = \{(\rho_i, y_i)\}_{i=1}^{N} \subset \mathbb{R}^n \times \{0,1\}^C,) \qquad (1)$$

$[\rho_{i_1}, \rho_{i_2}, \ldots, \rho_{i_n}]$ is i-th training sample n as features and c as classes. If S represents a finite set of samples, and X is a subset of S, $f(x)$ is the classifier on a set X. Mathematically, the misclassification rate is denoted as $P_e(f)$ and classifier $f(x)$, and $T = \{|x| \mid x \in S - X\}$ as Test Set for classifier performance evaluation. The misclassification rate is defined in (2)

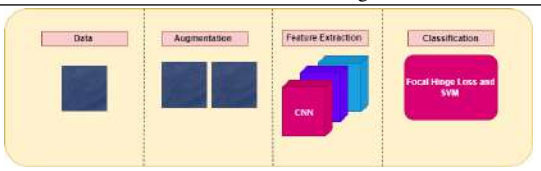

Figure 1: Processing Pipeline of the Proposed Model

$$([P_e(f) = \frac{|\{x \in T \mid f(x) \neq y(x)\}|}{|T|}) \qquad (2)$$

$F(x) \neq f(x)$ illustrates the inconsistency between the testing results and the class labels[30]. To the best of the author's awareness of the State-of-the-Art Deep CNN architectures, most of the research undermined the significance of gradient learning from training samples for underrepresented minority classes that can further contribute to classification accuracy. The proposed FH loss function postulates the concept of differential weight decay based on the difficulty of predicting the scenes. Hence, the motivation behind the FH loss function is making the model robust with Hinge loss maximizes the generalization ability of classifiers and accurate predictions, and the Focal Loss function down weights easy examples [11].

1illustrates the pipeline of the proposed model.

Thus, the significant contributions of this paper are outlined below:-

- A probabilistic gradient-based adaptive Focal-Hinge (FH) loss function is proposed for imbalanced Remote Sensing LULC maps. We upweighed the loss of misclassified instances in every class based on prediction difficulty and hence reduced misclassifications from minority classes.
- The proposed hybrid deep learning framework CNN-FHSVM further alleviated overfitting issues by introducing dropout and data augmentation techniques.
- Empirical evaluation of the proposed model on Sentinel 2 Euro SAT dataset's Red Green Blue (RGB) bands corroborated that CNN-FHSVM reduced LULC misclassifications with fewer computational parameters and training time compared to the recent advances in deep CNNs with relatively imbalanced datasets.

2 Literature Review

This section delves into a concise review of state-of-the-art Knowledge-oriented CNN architectures in RS

scene classification. Literature suggested some pioneering work of training a supervised model cross-domain with generalized pre-activation using a few labeled samples[32, 10]. Data Augmentation, Transfer Learning approaches and Active learning techniques are new learning paradigms discussed in the literature[2, 10, 17, 21, 32, 29]. In [9], the author illustrated a transfer learning algorithm trained with a top-2 smooth loss function for land classification problems for wide-scale TerraSAR-X images, addressing challenges of imbalanced classes and label noise. In [17], the author proposed two staged transfer learning frameworks to extract hierarchical feature vectors from the UC Merced Dataset. The results substantiated improved classification accuracy from 83.1% to 92.4%. To visualize the feature space in high dimensional stages, the author used t distributed stochastic neighbours embedded algorithm. In [19], the author fine-tuned Transfer Learning Knowledge ResNet and VGG 16 for LULC classification using the EuroSAT dataset. Training samples were augmented to enhance model performance and fine-tuned with techniques like gradient clipping. Further, in [28], the authors experimented with hyperspectral datasets and introduced the 3-D-LWNet model with sensors and cross-modal techniques for hyperspectral classification. In [21], the author validated similar investigations using Convolutional Networks for feature extraction. The results substantiated also delve into a systematic approach, including distortions and affine transformations in the model. In [11], the author proposed a fully differentiable Radial Basis Function Layer with CNN and manifested superior results compared to conventional models. In [14], the research presented a deep cube CNN model with Random Forests using spatial neighborhoods. The experiments were performed on high-dimensional hyperspectral datasets.

2.1 Loss Functions and Class Imbalance

Class Imbalance issue is also inherent in LULC Scene Classification. Hence, the probability of misclassifications increases manifolds in minority classes. In [16], the author experimented with the CIFAR dataset and corroborated that entropies of SoftMax function of classified samples decrease with training. The network becomes more confident with its incorrect predictions. In [31], the author proposed the Class Focal Loss accentuating the concept of differential weight decay in the driving scenes dataset. Significant weight decay should be assigned to easily classified samples in highly imbalanced datasets. In [18], the author analyzed the factors for miscalibration and model bias in deep learning models. With insights into research and experiments, focal loss demonstrated excellent performance in imbalanced datasets with enhanced accuracy. The experiments using different deep network architectures proved that deep neural networks trained with focal loss are more calibrated than the Cross-Entropy loss function [10]. In [22], the author extended this concept by reconstructing Hinge Loss and formulated the new Focal-Hinge(FH) loss function for binary classification. FH loss accentuates more on misclassified instances for mitigating misclassifications and diminishing the influence of class imbalance. The idea is further reinforced for high-resolution segmentation in satellite images [1]. The author elucidated a confusion map for measuring classification ability with calibrated focal loss. [15] propounded a loss combined weighted binary CE loss in change detection in remote sensing.[26] incorporated spatial correlation in neighboring pixels and corroborated neighbor loss for the segmentation task in the satellite images. Inspired by the work in [22], this research extends the Focal-Hinge loss for multiclass classification of LULC scenes in imbalanced learning problem comprehended with deep neural networks.

3 Proposed Framework

This section discusses the proposed deep hybrid framework CNN-FHSVM with novel Focal-Hinge loss for multiclass classification in imbalanced RS LULC scene classification. We first elaborate on theoretical concepts of the Focal Loss Function, then the rationale behind the proposed Focal-Hinge Loss function with the objective function of Multiclass SVMs.

3.1 Focal Loss Function Theoretical Basis

CE loss is the substantial loss explored in DL algorithms [16, 31]. However, this loss function predominantly gives attention to the majority examples

Given a patch with label y, SoftMax CE Loss is formulated in (3)

$$CE(z, y) = -\log\left(\frac{e^{z_y}}{\sum_{j=1}^{c} e^{z_j}}\right) \tag{3}$$

CE loss, if improved with the modulating term $(1-\rho)^{\gamma}$,it can perform well in imbalanced datasets

3.2 Focal Hinge Loss

In [11], the author elucidated the concept of Focal Hinge loss relative to the imbalance ratio, as stated in (4)

$$\text{LFH}(\mathbf{Y}, f(x)) = \max\{0, \bar{\tau}(1-\tilde{p})^{\gamma}(1-\mathbf{Y}f(x))\} \tag{4}$$

$$\bar{\tau} = \begin{cases} \tau & \text{if } Y = 1 \\ 1 - \tau & \text{otherwise} \end{cases}, \quad \tau \in [0, 1]$$

$$\rho = \begin{cases} \rho & \text{if } Y = 1 \\ 1 - \rho & \text{otherwise} \end{cases}, \quad \rho = \sigma(f(\kappa)) = \frac{1}{1 + e^{-f(\kappa)}}$$

Hence, the new final modified FH loss is given in (5)

$$\text{LFH}(y, f(x)) = \max\left\{0, \frac{\bar{\tau}}{(1 + e^{yf(x)})^{\gamma}}(1 - yf(x))\right\} \tag{5}$$

3.3 Proposed FH loss for multiclass SVMs with hybrid DL-ML model

We extended this FH loss for multiclass SVMs, elucidating its application in RS LULC scene recognition. Here, the hyperparameter γ is calculated using sample quantization. Higher γ can contribute a higher gradient in loss function for lower values of probability and vanishing gradients for higher probability. This concept is first introduced in [18], where the author gleaned insights about focal loss and model calibration in deep learning algorithms through extensive experiments with this hyperparameter. This observation paves the way to select γ sample-wise, assigning more weight to instances with less correct class probability. Mathematical interpretation is given in (6)

The first page must contain, in the following sequence:

$$\frac{\partial L_{\text{Focal}}}{\partial W} = \frac{\partial L}{\partial W} g(\rho, \gamma) \tag{6}$$

$y \in \{0, \ldots, K-1\}$ K denotes the number of classes, and the ρ vector represents the calculated probability distribution on K classes. The accuracy of each class has been improved by reshaping Hinge Loss by a scaling factor γ obtained. The mathematical interpretation of this FH loss is given in (5).

3.3.1 Objective Function of Multiclass Support Vector Machines and Loss Function

[hbt!] patches $\{\text{p}_0, p_1, \ldots, p_i\} each \{p_i\} is [64\times64\times3]$

Target conditional probability distribution $\text{P}(\langle Y_T \rangle | X_T)$

Pre-process the satellite images Feeding the augmented images to CNN-FHSVM Feature representation from alternate and last Convolutional Layers Computing the gradient from the class saving parameters θ

Training Classifier C with parameters θ

Focal loss optimization with factor (ρ, γ)

$\frac{\partial L_{\text{Focal}}}{\partial W} = \frac{\partial L}{\partial W} g(\rho, \gamma)$

$\theta \leftarrow \theta - \delta \nabla_{\theta} L(\theta)$;

Update parameter θ;

In Multiclass Classification, if training samples and their labels are $(x_n, y_n), n = 1 \ldots N, X_n \in \mathbb{R}$ [22], Then, the SVMs learning rate-constrained objective function is mathematically interpreted in (7)

$$(\min)_{(w,\xi^n)} \frac{1}{2} w^T w + C \sum_{n=1}^{N} \xi_n \tag{7}$$

$$w^T x_n t_n \geq 1 - \epsilon_n \quad \forall n, \quad \epsilon_n \geq 0 \quad \forall n$$

for optimization constraints given in (7) $_n$ Slack Variables that penalize data points violating margin requirements. Augmenting vectors xn, with scalar value 1, can also include bias. The corresponding optimization problem then becomes as in (8)

$$\min_{w} \frac{1}{2} W^T W + C \sum_{i=1}^{n} \max(1 - W^T(x_n, t_n), 0) \tag{8}$$

This is a primal form problem of $L1$ SVM. Optimization is performed in dual form for kernel SVMs.

3.3.2 Proposed Framework CNN-FHSVM

The proposed model, Convolutional Neural Networks with enhanced Focal Hinge loss-based SVM(CNN-FHSVM), improves Class accuracy and reduces misclassifications in imbalanced datasets. The rationale behind this approach is differential weight decay, giving more preference to instances where the model places less probability mass on the correct class. Focal loss regularizes weight norms in the network when the entire network achieves a certain level of prediction. When probability $\rho_i \to 1$, the modulating factor tends to 0, which downweights the loss value for well-classified examples. The hyperparameter $\gamma \geq 1$ rescales the modulating factor so that well-classified examples are down-weighted. The details of the layers of deep customized CNN are given in Fig. 3. Convolutions and Max Pooling are augmented to decrease the dimensionality of the feature matrix. Further, the dropout is added for the regularization of weights. Hence, it is beneficial for scenes where objects are at multiple scales. Four 2D convolutional layers with filters 32, 64, 128, and 256 are taken to capture spatial convolutions over the image. Two hundred fifty-six filters are used to learn the complex spatial hierarchies of the objects from the scenes. Scenes comprise a hierarchy of objects. Hence, filters are upsampled in every layer to capture the discriminative hierarchy of objects in complex scenes. The base

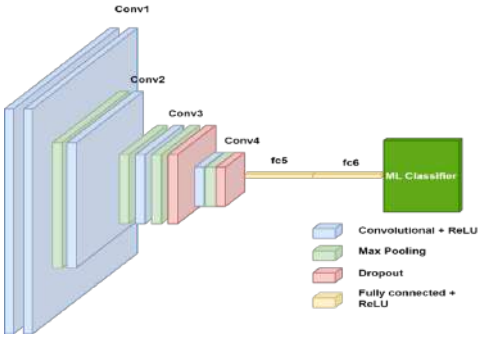

Figure 2: Base Architecture of CNN-FHSVM

architecture of the proposed framework with parameters is presented in Figure2 With the increase in confidence of the model, i.e., when the probability $\rho_i \to 1$, the modulating factor tends to 0. Hence, well-classified examples are down-weighted more than misclassified ones to reduce their impact on the loss function. Scaling diminishes the loss from easy examples. As an illustration, if we set $\gamma = 2$ and the probability associated with a well-classified example is $p = 0.9$, then the modulating factor will scale down its loss contribution by 100 times

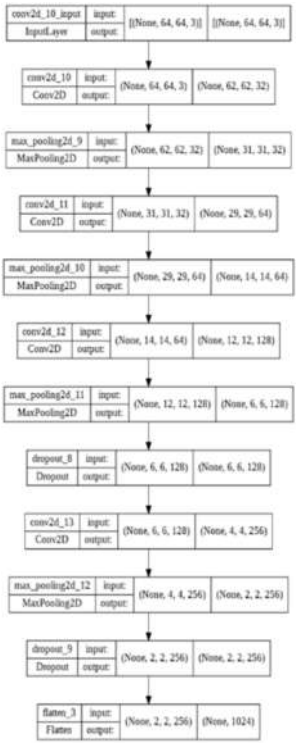

Figure 3: Model Summary

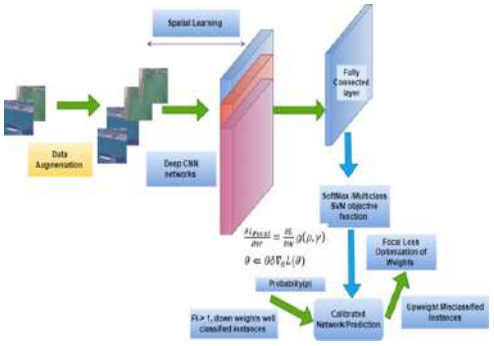

Figure 4: Proposed Framework for CNN- FHSVM model

4 Results

4.1 Experimental Configuration and Dataset Description

Experiments are executed using Python 3.8 on NVIDIA Tesla V100 GPU with 512GB RAM. The proposed Deep Learning Framework is evaluated with multiple experiments. The Sentinel-2 Eurosat comprises 27k 64×64 pixel patches categorized into ten LULC classes[7, 8]. It is multispectral, with thirteen spectral bands in the visible, near, and short-wave infrared [8]. The dataset description class-wise is given in Figure5 and Figure6.RGB bands are considered for experiments.

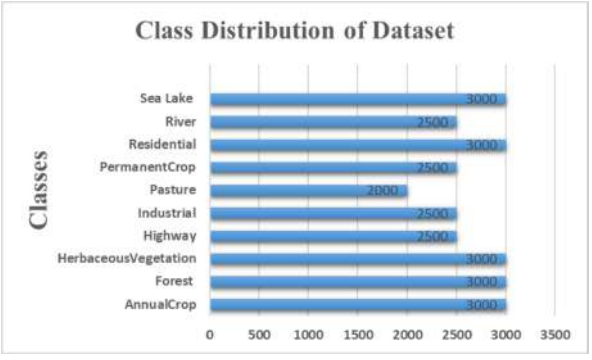

Figure 5: Class Distribution of Eurosat Dataset

4.2 Experimental Parameters

The proposed framework CNN-FHSVM is evaluated for assessment metrics training time, F1 score, and accuracy. Experiments were conducted using loss functions, such as categorical hinge loss and proposed FH loss. The focal loss hyperparameter γ has been calculated through gradient-based sample quantization. After extensive preliminary experiments, the best optimal value for hyperparameters is presented in Table1. The

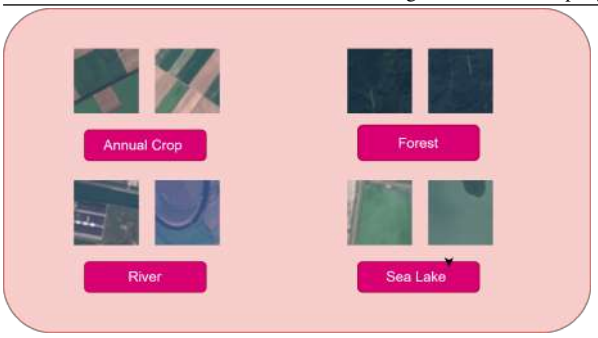

Figure 6: Sample images in EuroSAT [23]

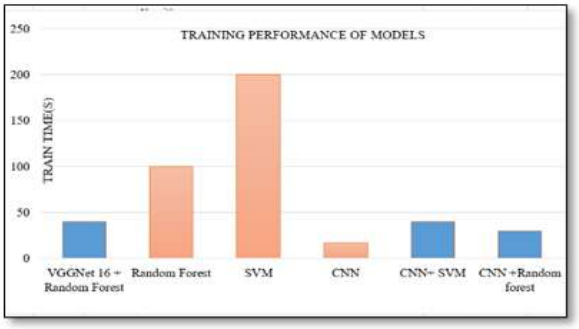

Figure 7: Comparison of Training Time in Seconds of Hybrid DL-ML Models

Table 1: Hyperparameters in Experiments

Hyperparameters	Experimental Values
Batch Size	64
Learning Rate Initial/Final	0.004
Dropout rate per layer	0.25
Patience	15
Activation Function	Relu
Loss Function	Hinge Loss/Focal Loss
Epochs	100
Data Augmentation	Rotate, Scale
γ	0.5(Optimal Value)

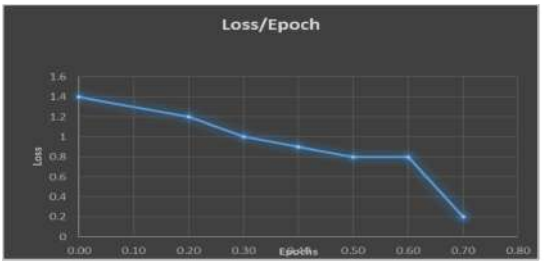

Figure 8: Convergence of FH loss function in CNN FHSVM model

dropout layer is added to regularize parameters and prevent overfitting. Data augmentation further increases the feature representation of training samples.

4.3 Results and Analysis

We performed experiments on RGB bands of Sentinel 2 EuroSat Dataset. The quantitative assessment of the introduced framework is contrasted with the leading-edge CNN architectures in Figure7, illustrating that deep architectures combined with ML techniques can outperform SVM and CNN models trained from scratch in training time and computational parameters with comparable accuracies[21, 11, 19, 13]. Matrix multiplications are maximum in the last neuron connections. Hence, mitigating dense layers and classifying with multiclass SVMs will enhance efficiency and decrease training time. The rationale behind choosing the hybrid CNN-FHSVM framework is the reduction of training time. Table2 compares our proposed deep learning framework with other hybrid frameworks. Further experiments with loss functions indicated that the proposed Loss function reduced misclassifications and increased class accuracy. Results are demonstrated in Tables 2-4. Table 2illustrates the experiments with a conventional loss function and deep networks. It is evident from 3 and 4 that the proposed loss function FH loss increased class accuracy and reduced misclassifications. In Table33, we have contrasted our model with variations in the loss function. From the results, it is imperative that the proposed loss function Focal-Hinge Loss outperforms conventional Hinge loss in terms of class accuracy and misclassifications. Further, Table4 indicates the accurate predictions of the proposed FH loss function. The experiments are also contrasted with deep learning architectures elucidated in the literature,our model enhances the accuracy with comparatively less training time. Figure8 shows that FH loss converges well with an increasing number of epochs. Since the Eurosat Dataset is not highly imbalanced, the re-weighting strategy of the loss function works well to enhance accuracy coupled with data augmentation techniques.

4.4 Discussion

Results corroborate that Focal Hinge loss qualifies as an alternative loss function with deep learning frameworks compared to conventional CE loss function in imbalanced datasets. The proposed loss function up weights misclassified instances based on prediction dif-

Table 2: Class-Wise Comparison of Accuracy of Hybrid Models

	Vggnet -Random Forest		CNN-SVM (Hinge Loss)		CNN-Random Forest	
CLASS	Class Accuracy	F1 Score	Class Accuracy	F1 score	Class Accuracy	F1 Score
Annual Crop	0.9045	0.8840	0.9025	0.9067	0.8923	0.8813
Forest	0.9980	0.9978	0.9413	0.9023	0.9589	0.9584
Herbaceous Vegetation	0.9650	0.9640	0.9124	0.9230	0.9450	0.9245
Highway	0.9650	0.9559	0.9410	0.9310	0.9456	0.9352
Industrial	0.9888	0.9780	0.9510	0.9712	0.9034	0.9445
Pasture	0.9440	0.9430	0.9034	0.9134	0.9245	0.9345
Permanent Crop	0.9528	0.9429	0.9312	0.9324	0.9034	0.8943
Residential	0.9986	0.9200	0.9489	0.9189	0.9556	0.9443
River	0.9476	0.9176	0.8981	0.8823	0.9267	0.9134
Sea Lake	0.9080	0.9180	0.9434	0.9134	0.9243	0.9223

Table 3: Comparative Class Accuracy of the proposed model CNN-SVM with variation in Loss Function

	CNN-SVM (Hinge Loss)		CNN-FHSVM (Proposed FH Loss Function)	
Class	Class Accuracy	F1 score	Class Accuracy	F1 Score
Annual Crop	0.9025	0.9067	0.9866	0.9867
Forest	0.9413	0.9023	0.9986	0.9780
Herbaceous Vegetation	0.9124	0.9230	0.9866	0.9867
Highway	0.9410	0.9310	0.9840	0.9812
Industrial	0.9510	0.9712	0.9900	0.9723
Pasture	0.9034	0.9134	0.9890	0.9723
Permanent Crop	0.9312	0.9324	0.9960	0.9760
Residential	0.9489	0.9189	0.9973	0.9812
River	0.9481	0.9323	0.9908	0.9817
Sea Lake	0.9434	0.9134	0.9866	0.9932

Table 4: True predictions with proposed FH loss function

Class	Samples	True predictions	Class Accuracy
Annual Crop	1500	1480	0.9866
Forest	1500	1496	0.9973
Herbaceous Vegetation	1500	1450	0.9666
Highway	1250	1230	0.9840
Industrial	1250	1225	0.9800
Pasture	1000	990	0.99
Permanent Crop	1250	1220	0.9760
Residential	1500	1496	0.9973
River	1250	1200	0.9600
Sea Lake	1500	1480	0.9866

ficulty and enhances class accuracy. Deep CNN architectures discussed in the literature essentially incorporated CE loss function with deep learning frameworks. The Focal-Hinge loss, as illustrated by the results, outperforms the contemporary literature studies in terms of training time and accuracy.

5 Conclusion and Future Investigations

This work proposed a Focal-Hinge loss-based deep learning framework for addressing misclassifications in relatively imbalanced RS scene classification. Insights from the literature suggested minimal efforts were directed toward the significance of gradients in loss func-

Table 5: Comparative performance summary with State of Art Proposed Models for LULC scene classification

Authors	Reference No.	Model	Accuracy
Chong	[4]	VGG 16	94.50
Helber	[8]	GoogleNet	98.18
Helber	[8]	ResNet 50	98.57
Sonune	[20]	Random Forest	61.46
Sonune	[20]	ResNet 50	94.25
Sonune	[20]	VGG 19	97.66
Naushad	[19]	VGG 16	98.14
Naushad	[19]	VGG16 (With Data Augmentation)	98.55
Naushad	[19]	Wide ResNet-50	99.04
Naushad	[19]	Wide ResNet-50 (With Data Augmentation)	99.17
Experiment	Experimental Model	Vggnet with Random Forest	97.5
Experiment	Proposed Model	CNN-FHSVM (With Data Augmentation)	99.6

tion for addressing misclassifications in LULC classification. The research propounded the idea of differential weight decay, giving more preference to instances where the model places less probability mass on the correct class. Results corroborate that the devised framework CNN-FHSVM has fewer misclassifications, specifically from minority classes, than state-of-the-art models. The work also demonstrates that FH loss qualifies as an alternative loss function with deep learning frameworks for the imbalanced RS multiclass scene classification. In the future, this novel loss function can be experimented on highly imbalanced satellite images.

6 References

References

[1] Bai, H., Cheng, J., Su, Y., Liu, S., and Liu, X. Calibrated focal loss for semantic labeling of high-resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:6531–6547, 2022.

[2] Ball, J. E., Anderson, D. T., and Chan, C. S. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of applied remote sensing*, 11(4):042609–042609, 2017.

[3] Cheng, G., Xie, X., Han, J., Guo, L., and Xia, G.-S. Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:3735–3756, 2020.

[4] Chong, E. Eurosat land use and land cover classification using deep learning, 2020.

[5] Farooqui, M. A. K. . M. R., N. A. Concatenated deep features with modified lstm for enhanced crop disease classification. *International Journal of Intelligent Robotics and Applications*, 7:510–534, 2023.

[6] Farooqui, N. A., Mishra, A. K., and Mehra, R. Iot based automated greenhouse using machine learning approach. *International Journal of Intelligent Robotics and Applications*, 7(3):226–231, 2022.

[7] Helber, P., Bischke, B., Dengel, A., and Borth, D. Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium*, pages 204–207. IEEE, 2018.

[8] Helber, P., Bischke, B., Dengel, A., and Borth, D. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019.

[9] Huang, Z., Dumitru, C. O., Pan, Z., Lei, B., and Datcu, M. Classification of large-scale high-resolution sar images with deep transfer learning. *IEEE Geoscience and Remote Sensing Letters*, 18(1):107–111, 2020.

[10] Jia, S., Jiang, S., Lin, Z., Li, N., Xu, M., and Yu, S. A survey: Deep learning for hyperspectral image classification with few labeled samples. *Neurocomputing*, 448:179–204, 2021.

[11] Jiang, S., Hartley, R., and Fernando, B. Kernel support vector machines and convolutional neural networks. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–7. IEEE, 2018.

[12] Kumari, M. S., N. Land use land cover scene classification with focal loss optimization of convolutional networks using sentinel -2 eurosat dataset.

[13] Kumari, N. and Minz, S. Deep residual svm: A hybrid learning approach to obtain high discriminative feature for land use and land cover classification. *Procedia Computer Science*, 218:1454–1462, 2023.

[14] Li, T., Leng, J., Kong, L., Guo, S., Bai, G., and Wang, K. Dcnr: deep cube cnn with random forest for hyperspectral image classification. *Multimedia Tools and Applications*, 78:3411–3433, 2019.

[15] Li, X., He, M., Li, H., and Shen, H. A combined loss-based multiscale fully convolutional network for high-resolution remote sensing image change detection. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021.

[16] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollar, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[17] Marmanis, D., Datcu, M., Esch, T., and Stilla, U. Deep learning earth observation classification using imagenet pretrained networks. *IEEE Geoscience and Remote Sensing Letters*, 13(1):105–109, 2015.

[18] Mukhoti, J., Kulharia, V., Sanyal, A., Golodetz, S., Torr, P., and Dokania, P. Calibrating deep neural networks using focal loss. *Advances in Neural Information Processing Systems*, 33:15288–15299, 2020.

[19] Naushad, R., Kaur, T., and Ghaderpour, E. Deep transfer learning for land use and land cover classification: A comparative study. *Sensors*, 21(23):8083, 2021.

[20] Sonune, N. Land cover classification with eurosat dataset, 2020.

[21] Wagner, S. A. Sar atr by a combination of convolutional neural network and support vector machines. *IEEE transactions on Aerospace and Electronic Systems*, 52(6):2861–2872, 2016.

[22] Wang, Q., Tian, Y., and Liu, D. Adaptive fh-svm for imbalanced classification. *IEEE Access*, 7:130410–130422, 2019.

[23] Wang, X., Xu, M., Xiong, X., and Ning, C. Remote sensing scene classification using heterogeneous feature extraction and multi-level fusion. *IEEE Access*, 8:217628–217641, 2020.

[24] Yang, S., Song, F., Jeon, G., and Sun, R. Scene changes understanding framework based on graph convolutional networks and swin transformer blocks for monitoring lclu using high-resolution remote sensing images. *Remote Sensing*, 14(15):3709, 2022.

[25] Yu, C., Zhu, Y., Song, M., Wang, Y., and Zhang, Q. Unseen feature extraction: Spatial mapping expansion with spectral compression network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[26] Yuan, W. and Xu, W. Neighborloss: a loss function considering spatial correlation for semantic segmentation of remote sensing image. *IEEE Access*, 9:75641–75649, 2021.

[27] Zeng, D., Liao, M., Tavakolian, M., Guo, Y., Zhou, B., Hu, D., Pietikäinen, M., and Liu, L. Deep learning for scene classification: A survey. *arXiv preprint arXiv:2101.10531*, 2021.

[28] Zhang, H., Li, Y., Jiang, Y., Wang, P., Shen, Q., and Shen, C. Hyperspectral classification based on lightweight 3-d-cnn with transfer learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):5813–5828, 2019.

[29] Zhang, L., Zhang, L., and Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and remote sensing magazine*, 4(2):22–40, 2016.

[30] Zhou, X., Wang, X., Hu, C., and Wang, R. An analysis on the relationship between uncertainty and misclassification rate of classifiers. *Information Sciences*, 535:16–27, 2020.

[31] Zhu, X., Men, J., Yang, L., and Li, K. Imbalanced driving scene recognition with class focal loss and data augmentation. *International Journal of Machine Learning and Cybernetics*, 13(10):2957–2975, 2022.

[32] Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., and Fraundorfer, F. Deep learning in

remote sensing: A comprehensive review and list of resources. *IEEE geoscience and remote sensing magazine*, 5(4):8–36, 2017.

[33] Zou, Q., Ni, L., Zhang, T., and Wang, Q. Deep learning based feature selection for remote sensing scene classification. *IEEE Geoscience and remote sensing letters*, 12(11):2321–2325, 2015.