EARTH SCIENCES
RESEARCH JOURNAL
Earth Sci. Res. J. Vol. 28, No. 4 (December, 2024): 409 - 419

UNIVERSIDAD NACIONAL DE COLOMBIA

GEOPHYSICS

Classification of Remote Sensing Datasets with Different Deep Learning Architectures

Maryam Mehmood[1,3*], Farhan Hussain[1], Ahsan Shahzad[1], Nouman Ali[2]
1. Department of Computer and Software Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan
2. Department of Software Engineering, Mirpur University of Science and Technology.
3. Department of software engineering, National University of Modern Languages, NUML, Islamabad
*Corresponding author: Maryam.mehmood@numl.edu.pk

ABSTRACT

Remote sensing image classification has great advantages in the areas of environmental monitoring, urban planning, disaster management and many others. Unmanned Aerial Vehicles (UAVs) have revolutionized remote sensing by providing high-resolution imagery. In this context, effective image classification is crucial for extracting meaningful information from UAV-captured images. This study presents a comparison of different deep learning-based approach for supervised image classification of UAV images. We have experimented on four different CNN models like VGG 16, Alex net, Resnet50 and a deep neural network Efficient-Net-B0 on different remote sensing datasets; AID and AIDER. Multiple combinations were tried to find out which model performs better on which type of datasets. We have used pre-trained initial layers of four CNN models (AlexNet, VGG 16, Resnet50 and Efficient-Net-Bo) then last three layers of each of the selected models are removed and new layers have been added with better tuned parameters. Two different schemes were analyzed. In Scheme-1 the original AlexNet, VGG 16, Resnet50 and Efficient-Net-B0 were experimented without changing and tuning their number of parameters, while in Scheme-2 transfer learning was applied on the pre-trained models and after removing last three layers new layers were added with better tuned hyper-parameters. The evaluation of above schemes was ensured through comprehensive metrics across diverse land cover classes, four different performance evaluation matrices namely; F1 score, precision, accuracy and recall. The main focus of this research is towards transfer learning and adding new layers into pre-trained models to get better classification accuracy.

Keywords: Convolutional Neural Networks (CNN), AlexNet, Resnet-50, VGG16, Efficient-Net-B0, Remote Sensing (RS) Image Classification, AID, AIDER, Unmanned Aerial Vehicles (UAVs).

Clasificación de grupos de datos de detección remota con diferentes arquitecturas de aprendizaje profundo

RESUMEN

La clasificación de imágenes de detección remota tiene grandes ventajas en las áreas de monitoreo ambiental, planeación urbana, manejo de desastres y muchos otros. Los vehículos aéreos no tripulados han revolucionizado la detección remota al proveer imágenes de alta resolución. En este contexto, la clasificación efectiva de imágenes es crucial para extraer información significativa de las imágenes capturadas por vehículos aéreos no tripulados. Este estudio presenta una comparación de diferentes técnicas de aprendizaje profundo para la clasificación supervisada de imágenes capturadas por vehículos aéreos no tripulados. Los autores experimentaron con diferentes grupos de datos AID y AIDER en cuatro modelos diferentes de Redes Neuronales Convolucionales (CNN), VGG 16, Alex net, Resnet50 y en la red neuronal profunda Efficient-Net-B0. Se intentaron múltiples combinaciones para encontrar el modelo con mayor desempeño en cada grupo de datos. Los autores usaron capas iniciales de preentrenamiento de los modelos CNN y luego se retiraron las tres últimas capas de cada uno de los modelos seleccionados para añadir luego capas con parámetros más acordes. Se analizaron dos esquemas diferentes. En el Esquema 1 se experimentaron los modelos CNN originales sin cambiar y sin adecuar el número de parámetros, mientras que en el Esquema 2 se aplicó la transferencia de aprendizaje en los modelos pre-entrenados y después de remover las tres últimas capas se añadieron nuevas capas con hiperparámetros más adecuados. La evaluación de estos esquemas fue asegurada a través de métricas completas para diversas clases de cobertura del suelo y con cuatro matrices de evaluación de desempeño llamadas puntuación F1, precisión, exactitud y exhaustividad. El foco principal de esta investigación se basa en la transferencia de aprendizaje y en la adición de nuevas capas en modelos pre-entrenados para tener una clasificación más precisa.

Palabras clave: Redes Neuronales Convolucionales (CNN); AlexNet; Resnet-50; VGG16; Efficient-Net-B0; clasificación de imágenes de detección remota; grupo de datos AID; grupo de datos AIDER; vehículos aéreos no tripulados.

Record
Manuscript received: 16/03/2024
Accepted for publication: 22/10/2024

How to cite item:
Mehmood, M., Hussain, F., Shahzad, A., & Ali, N. (2024). Classification of remote sensing datasets with different deep learning architectures. *Earth Sciences Research Journal, 28*(4), 409-419. https://doi.org/10.15446/esrj.v28n4.113518

ISSN 1794-6190 e-ISSN 2339-3459
https://doi.org/10.15446/esrj.v28n4.113518

1. Introduction

Remote sensing image classifications has multiple applications e.g., environment monitoring, land use land cover, disaster management, urban planning and many more (S. Zhang et al., 2023). Due to improvement in technology, it has become very easy to get satellite images of high resolution. Getting useful information from these remote sensing images is a tedious and difficult task that need high resource computation and time. With the passage of time this task has improved a lot and now there are quite many convolutional neural networks (CNN) available that can perform image interpretation task efficiently (Pronina & Piatykop, 2023).

According to the literature, from 2014 there has been a drastic evolution in remote sensing image classification (Kuang & He, 2014), the renowned CNN architectures include; Residual Network (Res Net), VGG network, Google Net and, Alex Net (Song, Gao, Zhu, & Ma, 2019). Latest Deep Neural Networks includes Efficient Net that ranges from b0-b7 Efficient-Net-B0 has 290 layers and Efficient-Net-B7 had 813 layers in the base models (Huang, Liu, Van Der Maaten, & Weinberger, 2017).

The main focus of this research on remote sensing image classification is to achieve a CNN model that has improved accuracy. So that we can accurately predict the scenes, especially when there is a disaster and accurate and timely prediction is required. There exists a huge gap to improve performance of these CNN by improving their architecture (Jeddi, Shafieezadeh, & Nateghi, 2023). Also, to understand the architecture and improve the performance by network tweaking. Another issue that still holds in efficient interpretation of remote images is that most of the existing CNN models are pre-trained on Image Net which is quite different from remote images in terms spectral bands, spatial resolution, and radiometric resolution so the process of transfer learning doesn't give very accurate results (Lilay & Taye, 2023).

Another issue with RS image classification is that, most of the existing CNN model giving very good accuracy is very bulky and require high computational systems and storage requirement (S. Song et al., 2019). There is a need to focus towards tweaking and tuning of existing small networks to get better accuracy (X. Cheng et al., 2022). By adjusting hyper parameters and network tweaking, where we change the existing base models by adding or removing the layers, in our case by network tweaking it is meant that we have removed the last three layers and added three new layers.

The main research contributions are as under;

- Four different deep learning models were tested by implying transfer learning. The last three layers of pre-trained model was removed and three new layers were added with tuned parameters and hyper parameters.
- Experimentation was performed on two different Remote Sensing (RS) data sets; AID and AIDER on four different CNN models.
- Two types of comparative analysis have been performed; one with the original deep learning models and the other one with the recent published research.

2. Related work

RS scene classification with CNN models has advanced significantly since 2014. Even while CNN-based techniques have been remarkably successful at classifying RS scenes, designing network architectures—which is essential to the CNN models' performance—remains a very difficult task (Rodriguez-Conde, Campos, & Fernandez-Riverola, 2022). The most advanced Designing a CNN architecture calls for proficiency in both domain knowledge and deep learning methodology. Aside from the fact that creating handcrafted architecture takes time and is prone to mistakes (Velayudhan, Hassan, Ahmed, Damiani, & Werghi, 2022), In practice, it is common to find a dearth of experts with substantial understanding in both the examined topic and deep learning technology. These spatial cues can improve the trained model's capacity for learning and classification accuracy in the context of remote sensing picture classification (X. Cheng et al., 2022). An image format based on rotation invariant feature vectors that use orthogonal vector histograms to compute spatial cues (Pittaras, Giannakopoulos, Stamatopoulos, & Karkaletsis, 2023). Publicly assessable satellite image benchmarks (RSSCN, AID and Siri-WHU) are used in the computation of the results. Using a fresh data set with a varied number of classes, fine-tuning is a technique used in conjunction with transfer learning to modify the parameters of a pre-trained CNN model (Ansith & Bini, 2022). If we reduce epochs and train a model on smaller learning rate the process would be beneficial (Kurakin et al., 2022).

According to Sinaga, Yudistira, & Santoso (2023), the usage of a CNN that has already been trained is the latest focus of image classification research. These retrieved features were then used to train a CNN for feature extraction. The authors used transfer learning to fine-tune using CNNs that had already been trained (Subramanian, Shanmugavadivel, & Nandhini, 2022).

AlexNet, Caffenet, Vggnet, Googlenet (Inception), and Resnet were among the pre-trained networks used in the majority of CNN-based RS scene classification experiments (Mehmood, Shahzad, Zafar, Shabbir, & Ali, 2022). Using transfer learning approach, we can fine tune networks and can use them for feature extraction. To depart from pretrained networks (Thribhuvan & Elayidom, 2022). CNN is trained from scratch using a weakly supervised triplet network with weakly single-label input. AlexNet is a deep learning model that was originally developed for image classification on natural images (Krizhevsky, Sutskever, & Hinton, 2012) and recently used in many other researches. Chaudhari et al., (2021) is one of those researchers. However, its architecture can also be used for remote sensing image analysis. Remote sensing images typically have different characteristics than natural images, such as different spectral bands and higher spatial resolution (Q. Cheng et al., 2022).

To use AlexNet for remote sensing image analysis, the model can be adapted by modifying the input layer to match the number of spectral bands in the image, and adjusting the number of neurons in the output layer to match the number of classes in the specific classification task (S. Song et al., 2019).

The pre-processing steps may also be different, such as normalizing the pixel values to a specific range and performing data augmentation (Tellez et al., 2019). There have been several studies that have used AlexNet for remote sensing image analysis, including land cover classification (Reddy, 2023), crop type classification (Wu et al., 2022), and object detection in high-resolution satellite images (Alganci, Soydas, & Sertel, 2020). AlexNet is the state-of-the-art model for remote sensing image analysis (Neris, Guerra, López, & Sarmiento, 2021); it is useful and has effective tools for certain other classification tasks as well. The VGG16 model consists of 16 layers, primarily composed of convolutional layers and followed by fully connected layers (Simonyan & Zisserman, 2014) and recently used by Adegun, Viriri, & Tapamo (2023). The key idea behind VGG16 is to use a series of small 3x3 convolutional filters repeatedly, which leads to a network with a larger number of trainable parameters but with a smaller number of layers compared to other models like AlexNet, accepts the input image with a specified size (usually 224x224 pixels) followed by convolutional layers, the network consists of 13 convolutional layers various features at different spatial scales. After some of the convolutional layers, max pooling layers are applied to reduce the spatial dimensions of the feature maps while retaining the most prominent features. Towards the end of the network, there are three fully connected layers. These layers combine the features learned by the convolutional layers to make predictions (Mehmood et al., 2022). The last layer is SoftMax layer, the function of this layer is to produce the final probability distribution over the classes. This allows the model to assign a probability to each class label. Many computer vision tasks, such as feature extraction, object identification, and image classification, have made extensive use of the VGG16 model as a benchmark (Saini, Khosla, Chand, Chouhan, & Prakash, 2023). Despite being outperformed by more recent architectures such as Resnet and Efficient-Net in terms of accuracy, VGG16 is still widely used because of its clarity and simplicity.

A deep learning model called Resnet50 has shown promise in a variety of computer vision applications, such as remote sensing image processing (K. He, Zhang, Ren, & Sun, 2015). The architecture of Resnet50 is based on residual blocks (Mandal, Okeukwu, & Theis, 2021). Each block in the Resnet50 design has multiple convolutional layers, followed by batch normalization and ReLU activation (Luo, Pearson, Xu, & Rich, 2022). The architecture has 50 layers total (Mandal et al., 2021). Each block's residual connections aid in preventing the vanishing gradient issue and enable more efficient model training (Basodi, Ji, Zhang, & Pan, 2020).

Large datasets of labelled images, such as satellite photos, can be used to train the Resnet50 model for use in remote sensing image analysis (Reedha, Dericquebourg, Canals, & Hafiane, 2022). The input data can be pre-processed to meet the specific criteria of the model, such as normalizing the pixel values and performing data augmentation to improve the diversity of the training data, in order to execute tasks like land cover classification, object detection, and change detection (Li, Chen, Chen, & Shi, 2021).

Resnet50 has been used in a number of research to analyze remote sensing images for the purposes of classifying vegetation, mapping land cover, and detection of objects in aerial images (X. Zhang, Han, Han, & Zhu, 2020). Resnet50 is considered to be one of the state-of-the-art models for many computer vision tasks, and its use in remote sensing image analysis can help to achieve highly accurate and efficient results (Müller & Hutter, 2021).

Efficient-Net is a family of deep convolutional neural networks (CNNs) that was introduced in 2019 (Yeşilmen & Tatar, 2022). Efficient-Net uses a novel compound scaling method that balances the depth, width, and resolution of the network to achieve better performance with fewer parameters than other Models (Hao, Han, Yang, Hao, & Li, 2021). The architecture of Efficient-Net consists of multiple blocks, each containing several layers, including convolutional layers, batch normalization, and ReLU activation (Hoang & Jo, 2021). The blocks are organized hierarchically, with higher-level blocks having fewer channels and lower spatial resolution, and lower-level blocks having more channels and higher spatial resolution (X. He, Zhao, & Chu, 2021).

The compound scaling method used in Efficient-Net involves scaling the depth, width, and resolution of the network in a balanced way (Tan & Le, 2019). The depth is scaled by increasing the number of layers, the width is scaled by increasing the number of channels, and the resolution is scaled by increasing the input image size. Efficient-Net requires fewer processing resources than other models, yet it delivers state-of-the-art performance on several computer vision benchmarks by striking a balance between these aspects.

Efficient-Net makes use of a method known as “swish” activation, which is a ReLU activation function variation that has been demonstrated to enhance model performance. Also, Efficient-Net employs a method known as “stochastic depth” to drop several layers at random during training in an effort to reduce over-fitting and enhance model performance. Efficient-Net is a deep learning network that is incredibly effective and efficient. It has demonstrated state-of-the-art performance on many computer vision tasks, such as semantic segmentation, object detection, and image classification.

The smallest variant in the Efficient Net family is the Efficient-Net-B0. It is intended to be a lightweight and effective model. The Efficient-Net-B0 architecture is made up of numerous blocks, each of which has a number of layers, including batch normalization, convolutional layers, and the Swish activation function. Because of the hierarchical organization of the blocks, lower-level blocks have more channels and higher spatial resolution, whereas higher-level blocks have fewer channels and lower spatial resolution.

Compound scaling is used by Efficient-Net-b0 to maintain network balance in terms of depth, width, and resolution. The number of channels increases the breadth, the number of layers increases the depth, and the size of the input image increases the resolution. Numerous computer vision tasks, such as semantic segmentation, object detection, and image classification, have demonstrated the efficacy of Efficient-Net-b0. Numerous real-world applications have made use of it, including pedestrian detection in surveillance systems and mobile facial recognition (Arif et al., 2022).

It addresses challenges in farmland semantic segmentation, such as variations in scale and similar visual features among agriculture patterns. It has two key modules: a feature fusion module to capture multi-scale variations and a global contextual module using pyramid pooling to aggregate contextual information from different image regions. Experimental results on the Agriculture-Vision benchmark dataset show that the framework outperforms state-of-the-art methods, achieving superior performance in identifying complex agricultural patterns (Khan, Alarabi, & Basalamah, 2023b)

This article addresses the problem of land scene classification in satellite imagery faces challenges due to complex textures, cluttered backgrounds, small object sizes, and scale variations in objects. It focuses on multi-branch deep learning framework addresses these issues by combining global contextual features and multi-scale local features through two branches: one for global information extraction and the other using a fully convolutional network (FCN) for local features. The results on the UC-Merced, SIRI-WHU, and EuroSAT datasets show that the proposed framework outperforms other models in classifying complex land scenes in high-resolution satellite images (Khan & Basalamah, 2023).

The article discussed the problem of road segmentation in high-resolution satellite images that is crucial for traffic monitoring and intelligent transportation, as it faces challenges such as cluttered backgrounds, inter-class correlation, and occlusions. DSMSA-Net, an encoder-decoder network with attention units, efficiently addresses these challenges by extracting multi-scale features and improving spatial representation using two attention modules: Scale Attention Unit (SaAU) and Spatial Attention Unit (SpAU).The framework is evaluated on two benchmark datasets, DeepGlobe and Massachusetts Road Dataset, and the results demonstrate that it achieves superior performance in road extraction compared to other reference methods (Khan, Alarabi, & Basalamah, 2023a).

3 Methodology

The proposed research aims to find out the appropriate CNN models best classifies the remote sensing datasets after fine turning of hyper parameters and network surgery. To get the desire results four base CNN models (Alex Net, VGG16, Resnet50 and Efficient-Net-b0) were used.

Different models have been selected starting from the ones that are published earliest to the models published in latest researches. Alex net and VGG16 are the models having lesser parameters and require less resources for computation. Resnet50 is a type of model that works on skip connections and relatively deeper model as compared to Alex net and VGG 16. The final model used for experimentation is Efficient-Net-B0, which has 290 layers and considered to be a deeper model.

Two problems are mainly in focus:

1. The available datasets for remote imaging are relatively smaller in size as compared to other image datasets.
2. High performance systems are required to process deeper models.

To find the best suited CNN Models for Remote sensing image classification that works well for smaller data sets using lesser resources. To overcome these problems our focus is to find the model that has lesser parameters and can perform well with lesser data with epochs not more than 30.

Remote sensing is also very useful for disaster management. One data set that contains images from disasters was selected and other data set contains normal aerial images. The general methodology was kept same for all the four models, starting from transfer learning of models to performing network surgery of the last three layers and tuning of hyper parameter. The Figure 2 shows the process.

Base Alex Net is the basic CNN model having 8 layers and has 11x11, 5x5, 3x3, convolutions, max pooling, dropout, data augmentation, ReLU activation, SGD with momentum. ReLU activation function is attached after every convolutional and fully-connected layer. Pre-trained Alex-net model was used that was initially trained on ImageNet dataset having 1.2 million images. Features and weights are transferred from that pre-trained to get better accuracy using less epochs. Same process is followed for VGG16 for our desired problem by performing network surgery on the last three layers. Fc1000, SoftMax layer and classification layer are removed from both base models and three new layers: fully connected layer where bias and weight values are adjusted as 20, SoftMax layer and classification layers are attached to the base model, for AlexNet the layers are attached at 2nd last fully connected layer whereas, for VGG 16 the three new layers are attached at last dropout layer. Learning rate is the most important hyper parameter. Here in this case piece-wise learning rate scheduler is used where initial learning rate is set to 0.0001, with drop factor of 0.1. The third model used for the experimentation is Resnet 50. Resnet 50 is the short form of residual network that has 50 deep layers. There is a concept in literature that says “deeper the network better the accuracy”. To check this, we performed our experimentation on four different categories of deep network to overcome the issue of deep neural networks like vanishing gradient decent that occurs when the learning rate drops to zero while moving deeper down the layers, we have decided to perform fine tuning of pre-trained networks. Same process is followed on Resnet 50, we have taken pre-trained model and by fine tuning of hyper parameters we have achieved better accuracy. Resnet 50 works on the principal of skip connections which was kept same during experimentation. The Figure 1 shows the whole process of transfer learning of initial layers, moving towards the fine tuning of fully connected layers and adjusting weight and bias learning factor as 20 that is attached to average pooling layer. The Resnet 50 was pre-trained on image net that has 1.2 million images.

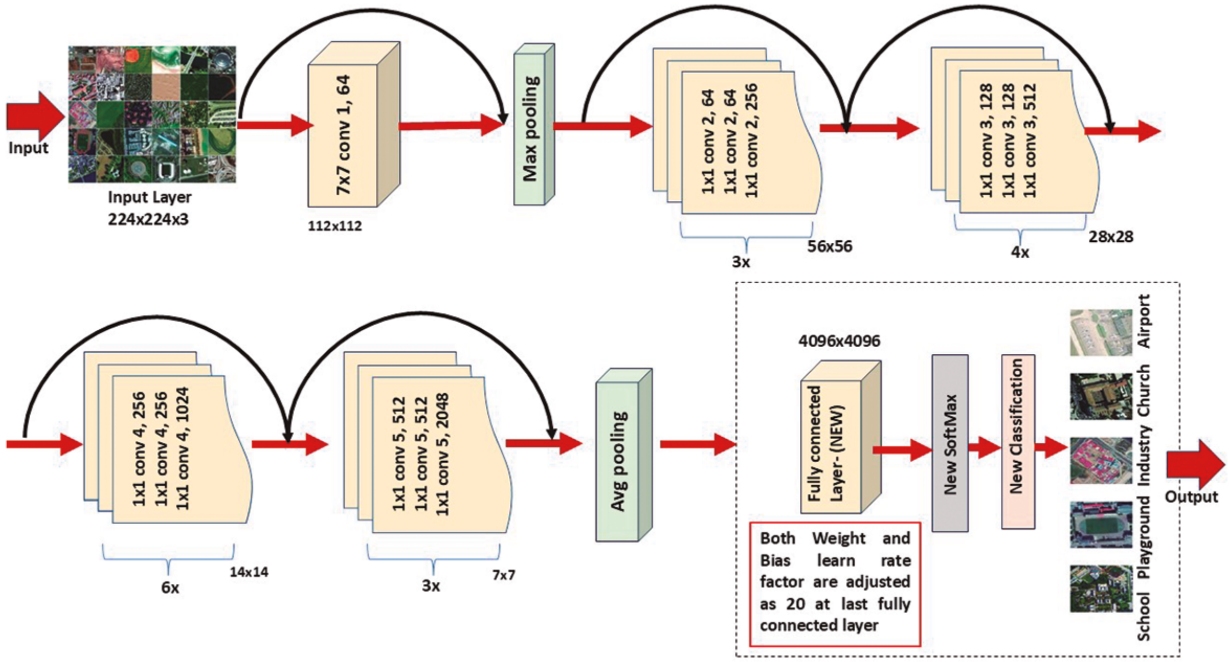

Figure 1. Proposed Resnet50 model used for experimentation after performing fine tuning and network surgery

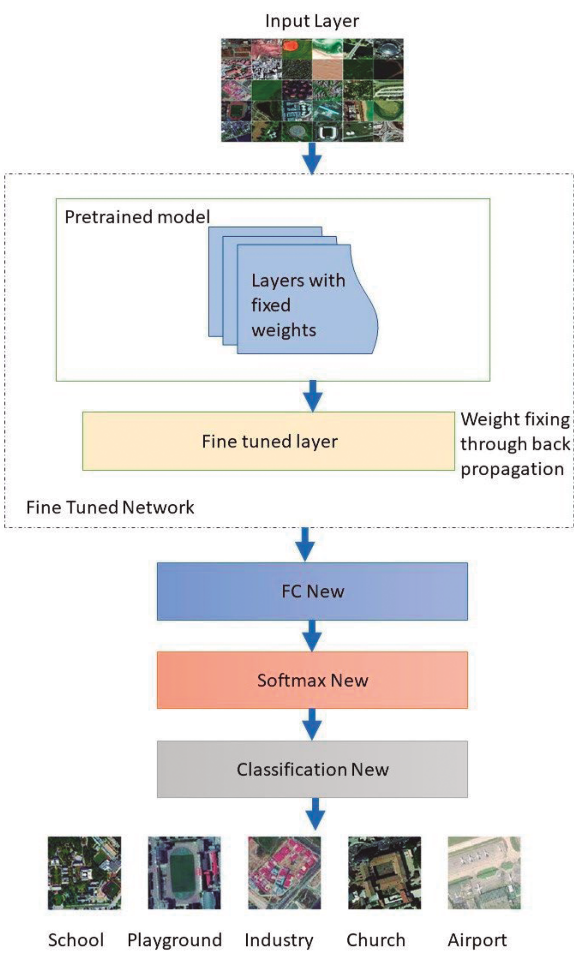

Figure 2. Methodology

The fourth model used for experimentation is Efficient-Net Bo that is 290 layers deeper. Network surgery is performed on pre-trained Efficient-Net b0 model on last three layers and three new layers fully connected layer with weight and bias values assigned as 20, SoftMax layer and classification layers are added at the end. Same network surgery is performed at this model as well where we removed the last three layers and added new layers with tuned hyper parameters. Same process is followed for all other models Alex Net, VGG16 and Resnet 50.

3.1 Experimentation

Experimentation was performed on four models Alex Net, VGG 16, Res Net50 and Efficient-Net Bo after performing the desired tweaking in the layers and tuning parameters and hyper parameters. The training testing ratio also kept same for each proposed model that is 70:30 Same parameters were used for each proposed model. The details are as follows in Table 1.

Also, the hyper parameters were kept same for all the three proposed models. The details are shown in Table 2.

Table 1. Parameters Values

Names	Values
Epochs	20 and 30
Optimizer	SGDM
Scheduler	Piece wise Learning rate

Table 2. Hyper Parameters Value

Names	Values
Weight Learning Rate Factor	20
Bias Learning Rate Factor	20
Initial Learning Rate	0.0001

3.2 Data set Details

Two Remote sensing data sets were used for experimentation. The details of data sets are given below;

3.2.1 AID

The AID data set has 17 different scene classes and about 200 to 400 samples of size 600x600 in each class. AID is a new large-scale aerial image data set, by collecting sample images from Google Earth imagery. The new data set is made up of the following aerial scene types: Airport Bare Land, Baseball Field, Beach, Bridge, Center, Church, Commercial, Dense Residential, Desert, Farm Land, Forest, Industrial, Meadow, Medium Residential, Mountains, Parking and Pond. The spatial resolution of images are 0.3-0.8m/pixel as shown in Figure 3.

Figure 3. Photo gallery of randomly selected images from AID data set

3.2.2 AIDER (Aerial Image Data set for Emergency Response Applications)

The data set construction involved manually collecting all images for four disaster events, namely Fire/Smoke, Flood, Collapsed Building/Rubble, and Traffic Accidents, as well as one class for the Normal case. As shown in Figure 4 The aerial images for the disaster events were collected through various online sources (e.g., google images, Bing images, YouTube, news agencies web sites, etc.) using the keywords Aerial View or UAV or Drone and an event such as Fire, Earthquake, Highway accident, etc. This subset includes around 500 images for each disaster class and over 4000 images for the normal class. Spatial resolution of images is 1m/pixel. The details of the data sets used during experimentation are shown in Table 3.

Table 3. Dataset Details

Name	Classes	Total images	Image/ class	Image size	Spatial Resolution
AID	18	5400	300	600x600	0.5 to 8 meters per pixel
AIDER	5	6000	500/ class,4000 normal	281x288	1 meter per pixel

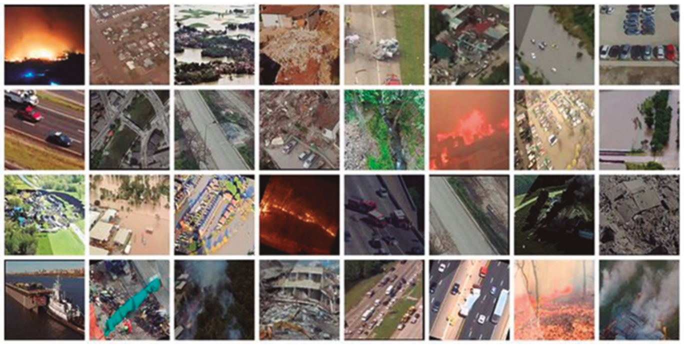

Figure 4. Photo gallery of randomly selected images of AIDER data

3.3 Performance Evaluation

The experimentation was performed on HP pavilion gaming Laptop the processor used in this system was AMD Ryzen 7 4800H with Radeon Graphics system. 64-bit Windows 10 were used as Operating system. 256 GB primary memory was used with 70:30 training testing ratio.

3.3.1 Time Performance Evaluation

Training time depends on many factors; hardware resources are one of them. Single NVIDIA GTX GPU was used for experimentation. The batch size was kept 8 through the work due to limited number of memory resources available.

While using the above-mentioned resource the training and prediction times are discussed in Table 4 and Table 5

Table 4. Training and Prediction Time for AIDER Dataset Using Proposed Resnet50 Model

Dataset	AIDER
Training time	115 min (6900 sec)
Total classes	5
Images per class	400, 2000 normal
Time for 1 epoch	230 sec
Time per class	46 sec
Prediction Time per image	0.0076 sec

Table 5. Training and Prediction Time for AID Dataset Using Proposed Resnet50 Model

Dataset	AID
Training time	119 min (6900 sec)
Total classes	18
Images per class	300
Time for 1 epoch	238 sec
Time per class	13.2 sec
Prediction Time per image	0.0044 sec

The matrices used for performance evaluation were accuracy, precision, F1 score and recall. Accuracy, precision, recall, and F1 score are commonly used performance metrics for evaluating the performance of CNN models. Accuracy is the most straightforward metric, which measures the percentage of correctly classified examples in the data set. Precision measures the proportion of true positives (correctly identified instances of a given class) among all positive predictions. Recall, also known as sensitivity, measures the proportion of true positives that are correctly identified by the model among all instances of that class in the data set. whereas, F1 score is the harmonic mean of precision and recall, and is often used as a single performance metric that combines both measures.

Accuracy is the most straightforward metric, which measures the percentage of correctly classified examples in the dataset. It is defined as:

Accuracy = (Number of correctly classified examples) / (Total number of examples)

Precision measures the proportion of true positives (correctly identified instances of a given class) among all positive predictions. It is defined as:

Precision = (True Positives) / (True Positives + False Positives)

Recall, also known as sensitivity, measures the proportion of true positives that are correctly identified by the model among all instances of that class in the dataset. It is defined as:

Recall = (True Positives) / (True Positives + False Negatives)

The F1 score is the harmonic mean of precision and recall, and is often used as a single performance metric that combines both measures. It is defined as:

F1 Score = 2 * ((Precision * Recall) / (Precision + Recall))

4. Results and Discussion

Initially the experimentation was performed on the base level Alex Net, VGG16, Resnet50 and Efficient-Net-B0 models and then with proposed changes the models were tested again using same datasets. it has noticed that with transfer learning and minor parameter tuning we can achieve higher accuracy. Then the same experimentation was performed on proposed models using two same datasets with 2 different epochs. Table 6 shows the accuracies of CNN models. It can clearly be seen that proposed Resnet50 performs best when the dataset is small and we are training our model with less Epochs, on AID dataset the accuracy of proposed Resnet50 is 97.67% when epochs were set to 30. The training and testing ratio was kept same for all the experimentation's that is 70:30. Efficient-net-b0 showed good results for AID dataset for 20 Epochs, the accuracy is observed as 93.73%, whereas when epochs were increased to 30. AIDER dataset showed better results, the accuracy observed was 95.29%.

The results of proposed four models were also compared with the results of all the four Base models. The Table 7shows the details.

Our results of AID dataset were then compared with the results of published data and have seen that with the above-mentioned network surgery and tweaking of parameters and hyper-parameters we can get better results with the help of conventional models. Our proposed model was compared with the different published researches, where they have also used AID dataset. In (Huang et al., 2017) google net model was used for experimentation and they claimed accuracy as 88.15%. The accuracy mentioned in(K. He, Zhang, Ren, & Sun, 2016) is 89.03% using inception net v3 model. Other models from different published researches like; resnet50, dense Net, mobile Net v2 were also validated with our proposed model.

As compared to published research where a new CNN model is proposed and AID dataset was used for experimentation. The accuracy of this proposed model is 96.10% (Broni-Bediako, Murata, Mormille, & Atsumi, 2021) whereas our model performs better as it gives accuracy as 97.61%. Table 8 shows the comparison of all the above-mentioned models already published in recent years.

A quantitative comparison is shown in Table 9, with recently published research in terms classification accuracy for AIDER image benchmark with 6000 images. Whereas, the published results are for AIDER dataset with 8540 total samples. Results of our proposed model were compared with different published Models like ER-Net, VGG16, Resnet50, Mobile Net, AISCC where accuracy mentioned were 90.16%, 91.25%, 89.61%, 88.55%, 95.24% respectively while the currently proposed model claims to have accuracy as 95.75% for AIDER dataset.

Table 6. Overall Accuracy of All the Four Models on Selected Aerial Image Datasets Where Training Testing Ratio Was Kept Same as 70:30

CNN models	Resnet50		Alex net		VGG16		Efficient net b0	
Epochs	20	30	20	30	20	30	20	30
AID	96.76%	97.61%	91.83%	91.88%	94.70%	94.72	93.73%	93.70%
AIDER	94.52%	95.75%	90.69%	90.72%	90.69%	90.84%	91.24%	95.29%

Table 7. Comparison of Accuracy of base models with proposed models

Scheme-1	Datasets		Scheme-2	Datasets	
CNN Model	AID	AIDER	CNN Models	AID	AIDER
Alex Net	85.60%	80.32%	Alex Net	91.88%	90.72%
VGG16	89.64%	82.20%	VGG16	94.72%	90.84%
Resnet50	94.03%	91.03%	Resnet50	97.61%	95.71%
Efficient net b0	93.43%	84.04%	Efficient net b0	93.70%	95.29%

Table 8. A Quantitative Comparison with Recently Published Research in Terms of Classification Accuracy for AID Image Benchmark.

Model	Accuracy
Google net (Huang et al., 2017)	88.15%
Inception V3 (K. He et al., 2016)	89.03%
Resnet-50 (Szegedy et al., 2015)	88.10%
Dense Net (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016)	87.50%
Mobile net v2 (Sandler, Howard, Zhu, Zhmoginov, & Chen, 2018)	89.33%
Model proposed in (Broni-Bediako et al., 2021)	96.10%
Scheme-2 Resnet50	97.61% %

Table 9. A Quantitative Comparison with Published Research (Alrayes, Alotaibi Et Al. 2022) In Terms of Classification Accuracy for AIDER Image Benchmark With 6000 Images

Model	**Accuracy**
ERNet Model	90.16%
VGG16 Model	91.25%
Resnet50 Model	89.61%
Mobilenet Model	88.55%
AISCC-DE2MS model	95.24%
Scheme-2 Resnet50	95.75%

As Results of Resnet50 are relatively better than other experimented models. Table 12 provides the class vise comparison of recall, precision and F1 score using AID data set for Epochs 30, the average recall is calculated as97.61%, the average value of precision is 97.81% whereas average computed for F1 score is 97.69%. similarly, for Epochs 20 the average recall is calculated as 96.65%, the average value of precision is 96.61% whereas average computed for F1 score is 96.65%. Table 10 shows its confusion matrix for 30 epochs using proposed Resnet50 for AID dataset it that help to understand which class gives better predictions and Table 11 shows the confusion matrix of proposed resnet50 for 20 epochs, to see the right predictions per class.

Table 15 provides the class vise comparison of recall, precision and F1 score using AIDER dataset for Epochs 30, the average recall is calculated as96.18%, the average value of precision is 95.51% whereas average computed for F1 score is 95.40%. Similarly, For Epochs 20 the average recall is calculated as 95.91%, the average value of precision is 95.91% whereas average computed for F1 score is 95.38%.Table 13 and Table 14 shows the confusion matrix for proposed Resnet 50 using AIDER dataset for 30 and 20 epochs respectively.

Table 10. Confusion Matrix for AID Dataset using Proposed Resnet 50 for 30 Epochs

	Airport	**Bare Land**	**Baseball Field**	**Beach**	**Bridge**	**Center**	**Church**	**Commercial**	**Dense Residential**	**Desert**	**Farm Land**	**Forest**	**Industrial**	**Meadow**	**Medium Residential**	**Mountains**	**Parking**	**Pond**
Airport	104	0	0	0	0	0	0	0	3	0	0	0	0	0	0	1	0	0
Bare Land	1	91	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Baseball Field	0	0	118	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beach	0	0	0	108	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bridge	0	0	0	0	119	0	0	0	2	0	0	0	0	1	0	0	0	1
Center	0	2	0	0	0	88	0	0	0	0	0	0	0	0	0	0	0	0
Church	0	0	0	0	0	0	110	0	0	1	0	0	0	0	0	0	0	0
Commercial	0	0	0	0	0	0	0	75	0	0	0	0	0	0	0	0	0	0
Dense Residential	0	0	0	0	1	0	0	0	106	0	0	0	0	3	0	2	1	4
Desert	0	0	0	0	0	0	1	1	0	82	0	0	0	0	0	0	0	0
Farm Land	0	0	0	0	4	0	0	0	0	0	83	0	0	0	0	0	0	0
Forest	0	0	0	0	0	0	0	0	0	0	0	102	0	0	0	0	0	0
Industrial	0	0	0	0	0	0	0	0	1	0	0	0	116	0	0	0	0	0
Meadow	0	0	0	0	2	0	0	0	0	0	1	0	0	100	0	1	1	0
Medium Residential	0	0	0	0	0	0	0	0	0	0	0	0	0	0	66	0	0	0
Mountains	1	0	0	1	0	0	0	0	2	0	0	0	0	0	0	69	4	1
Parking	0	0	0	0	0	0	0	0	1	0	2	0	1	0	0	2	63	3
Pond	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	104

Table 11. Confusion Matrix for AID Dataset using Proposed Resnet 50 for 20 Epochs

	Airport	Bare Land	Baseball Field	Beach	Bridge	Center	Church	Commercial	Dense Residential	Desert	Farm Land	Forest	Industrial	Meadow	Medium Residential	Mountains	Parking	Pond
Airport	103	1	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	1
Bare Land	0	90	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
Baseball Field	0	0	116	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0
Beach	0	0	0	107	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Bridge	0	0	0	0	119	0	0	0	0	0	1	0	2	0	0	0	1	0
Center	0	0	0	0	0	90	0	0	0	0	0	0	0	0	0	0	0	0
Church	0	0	0	0	0	0	111	0	0	0	0	0	0	0	0	0	0	0
Commercial	0	0	0	0	0	0	0	75	0	0	0	0	0	0	0	0	0	0
Dense Residential	0	0	0	0	1	0	0	0	111	0	0	0	1	1	0	2	0	1
Desert	1	0	0	1	0	0	0	0	0	82	0	0	0	0	0	0	0	0
Farm Land	0	0	0	0	1	0	0	0	0	0	86	0	0	0	0	0	0	0
Forest	0	0	0	0	0	0	0	0	0	1	0	101	0	0	0	0	0	0
Industrial	0	0	0	0	0	0	0	0	1	0	0	0	116	0	0	0	0	0
Meadow	0	0	0	1	0	0	0	1	0	0	0	0	0	101	0	0	0	2
Medium Residential	0	0	0	0	0	0	0	0	0	0	0	0	0	0	66	0	0	0
Mountains	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	72	4	1
Parking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	70	1
Pond	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	103

Table 12. Class Vise Comparison of AID Dataset for Proposed Resnet50

Proposed Resnet50	Epochs30			Epochs20		
AID Labels	Precision	Recall	F1score	Precision	Recall	F1score
Airport	98.09%	99.03%	0.99	96.29%	98.11%	0.97
Bare Land	96.77%	98.90%	0.98	97.85%	97.84%	0.98
Baseball Field	96.67%	99.14%	0.98	98.33%	100%	0.99
Beach	99.07%	94.69%	0.97	100%	97.29%	0.99
Bridge	96.74%	98.34%	0.98	96.74%	94.44%	0.96
Center	100%	96.77%	0.98	97.77%	98.87%	0.98
Church	100%	98.23%	0.99	99.09%	99.09%	0.99
Commercial	100%	98.68%	0.99	100%	98.68%	0.99
Dense Residential	94.87%	98.23%	0.97	90.59%	92.98%	0.92
Desert	97.61%	98.79%	0.98	97.61%	98.79%	0.98
Farm Land	98.85%	98.85%	0.99	95.40%	95.34%	0.95
Forest	99.01%	100%	1.00	100%	100%	1.00
Industrial	99.14%	97.47%	0.98	99.14%	99.14%	0.99
Meadow	96.19%	99.01%	0.98	95.23%	96.15%	0.96
Medium Residential	100%	97.05%	0.99	100%	100%	1.00
Mountains	92.32%	97.29%	0.95	88.46%	90.78%	0.90
Parking	97.22%	92.10%	0.95	87.50%	91.30%	0.89
Pond	98.09%	94.49%	0.96	99.04%	92.03%	0.95

Table 13. Confusion Matrix for AIDER Dataset Using Proposed Resnet50 for Epochs 30

	Traffic Incident	Normal	Flooded Areas	Fire	Collapsed Building
Traffic Incident	142	0	2	8	1
Normal	1	152	0	3	0
Flooded Areas	0	0	152	3	3
Fire	3	3	4	1300	7
Collapsed Building	1	0	1	9	134

Table 14. Confusion Matrix for AIDER Dataset Using Proposed Resnet50 for Epochs 20

	Traffic Incident	Normal	Flooded Areas	Fire	Collapsed Building
Traffic Incident	142	0	2	8	1
Normal	1	152	0	3	0
Flooded Areas	0	0	152	3	3
Fire	3	3	4	1300	7
Collapsed Building	1	0	1	9	134

Table 15. Class Vise Comparison of AIDER Dataset for Proposed Resnet50

Proposed Resnet50	Epochs20			Epochs30		
AIDER / Labels	Precision	Recall	F1score	Precision	Recall	F1score
Traffic Incident	92.82%	96.59%	0.94	93.33%	92.72%	0.93
Normal	97.43%	98.06%	0.97	96.77%	96.15%	0.96
Flooded Areas	96.20%	95.59%	0.96	98.08%	99.35%	0.99
Fire	98.70%	98.26%	0.98	98.62%	98.25%	0.98
Collapsed Building	92.41%	92.41%	0.92	87.66%	93.10%	0.90

5. Conclusion

The effectiveness of the deep learning-based model in accurately classifying UAV images, showcasing its potential for real-world applications. The scalability and adaptability of the proposed Resnet50 make it suitable for addressing challenges in monitoring and managing large-scale geographic areas especially for AID and AIDER datasets. The evaluations models such as; accuracy precision, recall, and F1 score have been used to assess the model's performance. This research contributes to the advancement of UAV image analysis by introducing a robust deep learning-based framework for supervised image classification. The outcomes underscore the significance of leveraging state-of-the-art technologies to enhance the precision and efficiency of land cover mapping using UAV imagery.

By performing different experimentation on different types of Scene image data sets we have come to the point where we can say that, for problems like disaster management where we have smaller data sets where we need efficient and urgent results deeper networks are not performing as good as moderate level networks like Resnet 50. We can get excellent results by altering few layers and by tweaking parameters and hyper parameters of base net Resnet50. Hence, it won't be wrong to claim that Resnet 50 with minor changes is the suitable network for Remote sensing images especially when your data set is small, we can get better accuracy by performing network surgery and by tuning hyper parameters. For AIDER data set we have achieved the accuracy of 95.75% using Resnet50 Model, which is better than latest published research Whereas for AID dataset we have achieved the accuracy of 97.61% using Resnet50 model which is also better than latest published researches. One of the limitations of transfer learning specifically for remote sensing imagery is that; the existing pretrained models are trained on image net dataset, although it is a very big dataset but most of its images are not related to remote sensing. which degraded the performance of CNN models. Future work may explore the integration of temporal information and the incorporation of additional modalities for a more comprehensive understanding of dynamic landscapes.

Code, Data, and Materials Availability

The data presented in this study are available on request from the corresponding author.

Conflict of Interest

The authors declare no conflict of interest.

References:

Adegun, A. A., Viriri, S., & Tapamo, J. R. (2023). Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis. *Journal of Big Data,* 10(1), 93. https://doi.org/10.1186/s40537-023-00772-x

Alganci, U., Soydas, M., & Sertel, E. (2020). Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images. *Remote sensing,* 12(3), 458. https://doi.org/10.3390/rs12030458

Ansith, S., & Bini, A. (2022). Land use classification of high resolution remote sensing images using an encoder based modified GAN architecture. *Displays,* 74, 102229. https://doi.org/10.1016/j.displa.2022.102229

Arif, E., Shahzad, S. K., Iqbal, M. W., Jaffar, M. A., Alshahrani, A. S., & Alghamdi, A. (2022). Automatic Detection of Weapons in Surveillance Cameras Using Efficient-Net. *Computers, Materials & Continua,* 72(3). DOI: 10.32604/cmc.2022.027571

Basodi, S., Ji, C., Zhang, H., & Pan, Y. (2020). Gradient amplification: An efficient way to train deep neural networks. *Big Data Mining and Analytics,* 3(3), 196-207.

Broni-Bediako, C., Murata, Y., Mormille, L. H., & Atsumi, M. (2021). Searching for CNN architectures for remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing,* 60, 1-13. https://doi.org/10.26599/BDMA.2020.9020004

Chaudhari, S., Sardar, V., Rahul, D., Chandan, M., Shivakale, M. S., & Harini, K. (2021). *Performance analysis of CNN, Alexnet and VGGNET models for drought prediction using satellite images.* Paper presented at the 2021 Asian Conference on Innovation in Technology (ASIANCON). https://doi.org/10.1109/ASIANCON51346.2021.9545068

Cheng, Q., Huang, H., Xu, Y., Zhou, Y., Li, H., & Wang, Z. (2022). NWPU-captions dataset and MLCA-net for remote sensing image captioning. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1-19. https://doi.org/10.1109/TGRS.2022.3201474

Cheng, X., He, X., Qiao, M., Li, P., Hu, S., Chang, P., & Tian, Z. (2022). Enhanced contextual representation with deep neural networks for land cover classification based on remote sensing images. *International Journal of Applied Earth Observation and Geoinformation, 107*, 102706. https://doi.org/10.1016/j.jag.2022.102706

Hao, W., Han, M., Yang, H., Hao, F., & Li, F. (2021). A novel Chinese herbal medicine classification approach based on EfficientNet. *Systems Science & Control Engineering,* 9(1), 304-313. https://doi.org/10.1080/21642583.2021.1901159

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

He, X., Zhao, K., & Chu, X. (2021). AutoML: A survey of the state-of-the-art. *Knowledge-based systems, 212*, 106622. https://doi.org/10.1016/j.knosys.2020.106622

Hoang, V. T., & Jo, K. H. (2021). *Practical analysis on architecture of EfficientNet.* Paper presented at the 2021 14th International Conference on Human System Interaction (HSI). https://doi.org/10.1109/HSI52170.2021.9538782

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely connected convolutional networks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Jeddi, A. B., Shafieezadeh, A., & Nateghi, R. (2023). PDP-CNN: A Deep Learning Model for Post-Hurricane Reconnaissance of Electricity Infrastructure on Resource-Constrained Embedded Systems at the Edge. *IEEE Transactions on Instrumentation and Measurement,* 72, 1-9. https://doi.org/10.1109/TIM.2023.3236321

Khan, S. D., Alarabi, L., & Basalamah, S. (2023a). DSMSA-Net: Deep spatial and multi-scale attention network for road extraction in high spatial resolution satellite images. *Arabian Journal for Science and Engineering, 48*(2), 1907-1920. https://doi.org/10.1007/s13369-022-07082-z

Khan, S. D., Alarabi, L., & Basalamah, S. (2023b). Segmentation of farmlands in aerial images by deep learning framework with feature fusion and context aggregation modules. *Multimedia Tools and Applications, 82*(27), 42353-42372. https://doi.org/10.1007/s11042-023-14962-5

Khan, S. D., & Basalamah, S. (2023). Multi-branch deep learning framework for land scene classification in satellite imagery. *Remote Sensing,* 15(13), 3408. https://doi.org/10.3390/rs15133408

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems,* 25.

Kuang, D., & He, L. (2014). *Classification on ADHD with deep learning.* Paper presented at the 2014 International Conference on Cloud Computing and Big Data. https://doi.org/10.1109/CCBD.2014.42

Kurakin, A., Song, S., Chien, S., Geambasu, R., Terzis, A., & Thakurta, A. (2022). Toward training at imagenet scale with differential privacy. *arXiv preprint arXiv:2201.12328*. https://doi.org/10.48550/arXiv.2201.12328

Li, W., Chen, K., Chen, H., & Shi, Z. (2021). Geographical knowledge-driven representation learning for remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing,* 60, 1-16. https://doi.org/10.1109/TGRS.2021.3115569

Lilay, M. Y., & Taye, G. D. (2023). Semantic segmentation model for land cover classification from satellite images in Gambella National Park, Ethiopia. *SN Applied Sciences,* 5(3), 76. https://doi.org/10.1007/s42452-023-05280-4

Luo, C.-Y., Pearson, P., Xu, G., & Rich, S. M. (2022). A computer vision-based approach for tick identification using deep learning models. *Insects,* 13(2), 116. https://doi.org/10.3390/insects13020116

Mandal, B., Okeukwu, A., & Theis, Y. (2021). Masked face recognition using resnet-50. *arXiv preprint arXiv:2104.08997*. https://doi.org/10.48550/arXiv.2104.08997

Mehmood, M., Shahzad, A., Zafar, B., Shabbir, A., & Ali, N. (2022). Remote sensing image classification: A comprehensive review and applications. *Mathematical Problems in Engineering,* 1-24. https://doi.org/10.1155/2022/5880959

Müller, S. G., & Hutter, F. (2021). *Trivialaugment: Tuning-free yet state-of-the-art data augmentation.* Paper presented at the Proceedings of the IEEE/CVF international conference on computer vision.

Neris, R., Guerra, R., López, S., & Sarmiento, R. (2021). *Performance evaluation of state-of-the-art CNN architectures for the on-board processing of remotely sensed images.* Paper presented at the 2021 XXXVI Conference on Design of Circuits and Integrated Systems (DCIS). https://doi.org/10.1109/DCIS53048.2021.9666179

Pittaras, N., Giannakopoulos, G., Stamatopoulos, P., & Karkaletsis, V. (2023). Content-based and knowledge-enriched representations for classification across modalities: a survey. *ACM Computing Surveys, 55*(14s), 1-40. https://doi.org/10.1145/3583682

Pronina, O., & Piatykop, O. (2023). *The recognition of speech defects using convolutional neural network.* Paper presented at the CTE Workshop Proceedings. https://doi.org/10.55056/cte.554

Reddy, K. V. V. K. (2023). Comparing Linear Discriminant Analysis to AlexNet as a Novel Approach for Better Remote Sensing Image Segmentation and Classification with Improved Accuracy. *Journal of Survey in Fisheries Sciences, 10*(1S), 2894-2903. https://doi.org/10.17762/sfs.v10i1S.522 https://doi.org/10.3390/rs14030592

Reedha, R., Dericquebourg, E., Canals, R., & Hafiane, A. (2022). Transformer neural network for weed and crop classification of high resolution UAV images. *Remote sensing,* 14(3), 592.

Rodriguez-Conde, I., Campos, C., & Fernandez-Riverola, F. (2022). Optimized convolutional neural network architectures for efficient on-device vision-based object detection. *Neural Computing and Applications, 34*(13), 10469-10501. https://doi.org/10.1007/s00521-021-06830-w

Saini, D., Khosla, A., Chand, T., Chouhan, D. K., & Prakash, M. (2023). Automated knee osteoarthritis severity classification using three-stage preprocessing method and VGG16 architecture. *International Journal of Imaging Systems and Technology, 33*(3), 1028-1047. https://doi.org/10.1002/ima.22845

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). *Mobilenetv2: Inverted residuals and linear bottlenecks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556

Sinaga, K. B., Yudistira, N., & Santoso, E. (2023). Efficient CNN for high-resolution remote sensing imagery understanding. *Multimedia Tools and Applications*, 1-23.

Song, J., Gao, S., Zhu, Y., & Ma, C. (2019). A survey of remote sensing image classification based on CNNs. *Big earth data, 3*(3), 232-254. https://doi.org/10.1080/20964471.2019.1657720

Song, S., Yu, H., Miao, Z., Zhang, Q., Lin, Y., & Wang, S. (2019). Domain adaptation for convolutional neural networks-based remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters,* 16(8), 1324-1328. https://doi.org/10.1109/LGRS.2019.2896411

Subramanian, M., Shanmugavadivel, K., & Nandhini, P. (2022). On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves. *Neural Computing and Applications, 34*(16), 13951-13968. https://doi.org/10.1007/s00521-022-07246-w

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). *Rethinking the inception architecture for computer vision.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Tan, M., & Le, Q. (2019). *Efficientnet: Rethinking model scaling for convolutional neural networks.* Paper presented at the International conference on machine learning.

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.-M., Ciompi, F., & Van Der Laak, J. (2019). Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical image analysis, 58*, 101544.

Thribhuvan, N., & Elayidom, S. (2022). Transfer Learning for Feature Dimensionality Reduction. *The International Arab Journal of Information Technology, 19*(5), 721-727. https://doi.org/10.34028/iajit/19/5/3

Velayudhan, D., Hassan, T., Ahmed, A. H., Damiani, E., & Werghi, N. (2022). *Baggage threat recognition using deep low-rank broad learning detector.* Paper presented at the 2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON). DOI: 10.1109/MELECON53508.2022.9842976

Wu, F., Maleki, R., Oubara, A., Gómez, D., Eftekhari, A., & Yang, G. (2022). *Machine Learning Approaches for Crop Identification from Remote Sensing Imagery: A Review.* Paper presented at the International Conference on Soft Computing and Pattern Recognition. https://doi.org/10.1007/978-3-031-27524-1_31

Yeşilmen, S., & Tatar, B. (2022). Efficiency of convolutional neural networks (CNN) based image classification for monitoring construction related activities: A case study on aggregate mining for concrete production. *Case Studies in Construction Materials, 17*, e01372. https://doi.org/10.1016/j.cscm.2022.e01372

Zhang, S., Yang, P., Xia, J., Wang, W., Cai, W., Chen, N., . . . Zhan, C. (2023). Remote sensing inversion and prediction of land use land cover in the middle reaches of the Yangtze River basin, China. *Environmental Science and Pollution Research, 30*(16), 46306-46320. https://doi.org/10.1007/s11356-023-25424-8

Zhang, X., Han, L., Han, L., & Zhu, L. (2020). How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? *Remote Sensing, 12*(3), 417. https://doi.org/10.3390/rs12030417