International Journal of Advanced Trends in Computer Applications
www.ijatca.com

Satellite Image Processing Using Azure Databricks and Residual Neural Network

[1]Dinesh Kalla, [2]Nathan Smith, [3]Fnu Samaah
[1,2]Department of Doctoral Studies, Colorado Technical University, Colorado Springs, CO, USA
[3]Department of Computer Science, Harrisburg University of Science and Technology, Harrisburg, PA, USA
[1]dinesh.kalla@student.ctuonline.edu, [2]nathan.smith246@student.ctuonline.edu,
[3]fsamaah@alumni.harrisburgu.edu

Abstract: *This paper aims to perform satellite image processing using Machine Learning models and evaluate their prediction scores. This research will classify satellite images into four distinct categories, namely "green area," "desert," "water," and "cloudy, and will train and evaluateseveral criteria to highlight the model's exceptional performance, including precision, recall, F1-scores, and total accuracy. The model exhibits exceptional accuracy in correctly predicting and identifying positive instances, as seen by the near-perfect scores achieved for precision and recall across most classes. The F1 scores demonstrate a cohesive equilibrium across various measures, indicating the approach's efficacy. Significantly, the model has a remarkable overall accuracy rate of 99%, emphasizing its proficiency in precise picture categorization. The use of macro and weighted averages highlights the resilience and uniformity of its performance, irrespective of variations in class distribution. The findings presented in this study provide evidence supporting the appropriateness of the model for a range of applications, with a particular emphasis on computer vision and machine learning. Evaluation measures like accuracy, recall, and F1-score provide a detailed analysis of the model's capabilities, rendering them essential for assessing classification models.*

Keywords: Artificial Intelligence, Azure Databricks, Machine Learning, Deep Learning, Convolution Neural Network, Residual Network, Artificial Neural Network, ResNet 50, Image Processing, Vikram lander , Pragyan Lunar Rover, Perseverance Mars Rover, LRV, NASA, Remote Sensing and Satellite Image Processing**.**

I. Introduction

In the ever-changing world of technology, Artificial Intelligence (AI) and its offshoot, Machine Learning (ML), have emerged as front-and-center players, revolutionizing our relationship to and use of data. In the field of Machine Learning, Deep Learning has quickly become a formidable force because of its ability to teach computers to learn and make sound judgments by emulating the neural network architecture of the human brain [1].

Convolutional Neural Networks (CNNs) are crucial in image processing, one of the many novel applications spawned by this revolutionary advancement. Residual Networks (ResNets) are an innovative design that has contributed significantly to deep learning's development. Using these networks, models with hundreds of layers may be built without sacrificing efficiency or accuracy, revolutionizing the training of extremely deep neural networks. RES 50, often known as ResNet-50, is an architecture that has proven to be a popular solution for many different computer vision applications due to its excellent balance between computational complexity and speed. AI and deep learning continue to push their limits, their applications reaching new heights and beyond typical fields. Artificial intelligence (AI) has recently met satellite image processing [2]. Combining satellite data with ANNs and other advanced machine-learning methods has opened up many new applications, from environmental monitoring and disaster management to city planning and farming. This research paper delves into the interplay between cutting-edge technologies and their impact on a wide range of sectors to learn how AI, ML, DL, RNNs, ANNs, RES 50, and sat-image processing are changing the world [3]. This exploration of cutting-edge discoveries will shed light on how these tools have altered our view of the world and how they are used to expand human potential in the present day.

II. Literature Review

2.1 Convolution Neural Network

Convolutional Neural Networks (CNNs) have significantly transformed the domains of computer vision, image processing, and pattern recognition, representing a pivotal element within deep learning. Convolutional neural networks (CNNs) are specifically developed to autonomously and flexibly acquire spatial hierarchies of features from input. As a result, they have become fundamental in a wide range of applications, particularly in image analysis and beyond [4]. The inception and evolution of Convolutional Neural Networks (CNNs) may be traced back to the late 1980s, signifying their extensive historical background [5]. Nevertheless, it was only in the mid-1990s that Yann LeCun and his colleagues successfully devised LeNet-5, an innovative convolutional neural network (CNN) framework that laid the foundation for further advancements in this field. The LeNet-5 architecture showcased the efficacy of using convolutional layers and subsampling layers to recognize handwritten digits, hence serving as a forerunner to contemporary advanced image recognition systems.

The ImageNet Large Scale Visual Recognition Challenge was a significant turning point for Convolutional Neural Networks (CNNs), ushering in a profound revolution in deep learning. In 2012, Alex Krizhevsky and his research team created a deep Convolutional Neural Network (CNN) architecture called AlexNet. This design demonstrated a noteworthy decrease in error rates for the image categorization task [6]. The triumph in question catalyzed the subsequent deep learning revolution, thereby demonstrating the efficacy of convolutional neural networks (CNNs) in effectively managing and processing extensive collections of picture information. The emergence of AlexNet as a successful model in Convolutional Neural Networks (CNNs) has subsequently led to the development of other architectural advancements. Researchers undertook the introduction of deeper networks with more intricate structures. Prominent instances include the VGGNet, GoogLeNet, and ResNet designs, which have effectively expanded network depth and precision limits. Residual Networks (ResNets) were crucial in advancing deep network training by introducing skip connections. These skip connections effectively mitigated the vanishing gradient issue and significantly improved the performance of Convolutional Neural Networks (CNNs). Figure 1 represents the architecture of a convolution neural network with several stages, including convolution and pooling.

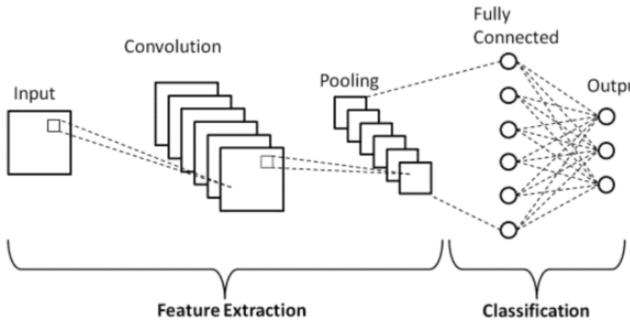

Figure 1: Convolution Neural Network Architecture

Transfer learning has emerged as a significant advancement in the field. It entails using pre-trained convolutional neural network (CNN) models trained on extensive datasets to tackle novel problems [6]. This method significantly reduced the need for extensive computing resources and annotated data to train new models. Deep Learning model exhibits better accuracy and prediction than conventional machine learning model [7][8]. Transfer learning has facilitated the rapid implementation of convolutional neural networks (CNNs) in various domains, including medical image analysis and natural language processing [9]. Convolutional Neural Networks (CNNs) have expanded their use beyond image categorization and object identification domains. Convolutional neural networks have been used in several domains, such as image segmentation, artistic creation, video analysis, autonomous cars, and even the comprehension of the human genome. Convolutional neural networks (CNNs) have shown their versatility in a wide range of scientific and commercial fields due to their capacity to acquire knowledge of intricate spatial hierarchies and patterns. Challenges and Ongoing Research: Notwithstanding their remarkable capabilities, Convolutional Neural Networks (CNNs) are not exempt from encountering obstacles. Many concerns arise concerning their decision-making process, including the issue of interpretability, the ability to withstand adversarial assaults, and the need for extensive labeled datasets [10]. Continuing research efforts aim to tackle these challenges and extend the frontiers of Convolutional Neural Networks (CNNs). With thehelp of satellite image processing using deep learning techniques we can track rover and landers like Vikram, Pragyan and Perseverance.

2.2 ResNet 50

ResNetdeep learning models are better in prediction when usually trained with group images. The network is used to process the data that has been entered. Adding blank rows and columns to the supplied data is known as zero padding [11]. The spatial dimensions of the feature maps are preserved by using convolutional layers. The convolutional layer is what this symbol represents. It is a cornerstone of Convolutional Neural Networks. When applied to the input data, convolution

procedures create feature maps that capture various patterns and characteristics [12]. Batch Normalization is a method for improving the consistency and speed with which deep neural networks are trained. It speeds up and stabilizes training by normalizing the layer's input such that its mean and variance are both one. Rectified Linear Unit (ReLU) is an activation function that may be used at a layer's output. By setting all negative values to zero, it creates a non-linear function.
By selecting the most significant value in each pooling window, Max Pooling is a downsampling procedure that preserves the most relevant details while decreasing the spatial dimensions of the feature maps. Network stages are labeled with numbers (1–5) to indicate their relative complexity [12]. Convolutional and identification (ID) blocks are the standard building components for each ResNet level. It is common practice for a Convolutional Block to begin with a set of convolutional layers, implement batch normalization, and finish with a ReLU activation function. An Identity Block, or ID Block, is a residual block with a skip link that allows it to avoid one or more convolutional layers. It reduces the impact of the vanishing gradient issue during the training of extremely deep networks. The average value inside a pooling window is calculated using another downsampling procedure called "Average Pooling." The 2D feature maps are "flattened" or transformed into a 1D vector [13]. This is a typical procedure before establishing a fully connected (FC) layer connection in the network. The acronym FC refers to "Fully Connected" layers. Predictions and classifications using the retrieved features are often performed in these layers, generally reserved for the network's final stage. The network's output is produced in the last layer, called the output layer, and it often represents class probabilities in a classification job.

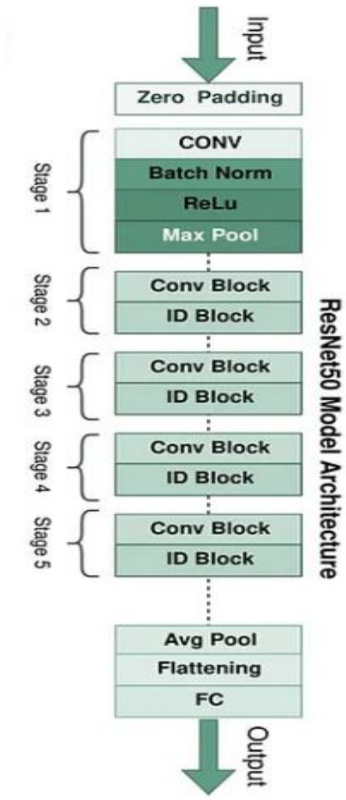

Figure 2 : ResNet 50 Architecture

III.Methodology and Architecture of Residual Network

Specific actions and considerations must be taken to enable the effective implementation of a deep learning model such as RESNET-50. The methods and procedures for creating a RESNET-50 model for a computer vision problem are outlined in this section.

3.1 Data Collection and Storing in Azure Data Lake

The data gathering section compiles a labeled dataset relevant to the objective. This dataset should include sets for training, validation, and testing [14]. The dataset is collected and stored in Azure data lake, which is further moved to the preprocessing stage.

3.2 Data Preprocessing and Splitting the Dataset

This stage includes resizing photos, standardizing pixel values, and enriching data to boost variety. Import a pre-trained RESNET-50 model into TensorFlow or PyTorch using a deep learning framework. Using Azure Databricks python script, the dataset is divided into training, validation, and testing. The weights of the additional custom layers are adjusted to fine-tune the pre-trained RESNET-50 model on the dataset [15]. At this stage, the model defines a loss function appropriate for the purpose (for example, categorical cross-entropy for classification). Training needs to be monitored to avoid overfitting using measures like accuracy and loss and approaches such as early stopping.

3.3 Testing and training using ResNet50 in Azure Databricks

This stage will use the Azure Databricks tool to train and test the ResNet 50 Model.To improve model performance, an experiment will be conducted with hyperparameters such as learning rate, batch size, and the number of layers in the custom section of the network is decided in this model. It is done using Databricks ML computes [16].

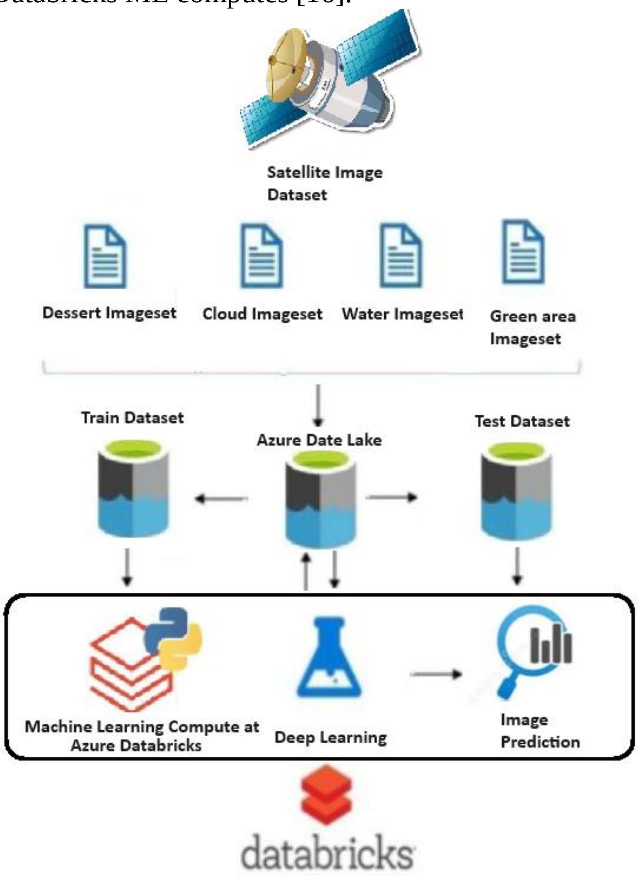

Figure 3: Implementation of ML Model

After training, the model will use relevant metrics to assess the model's performance on the testing set (e.g., accuracy, precision, recall, F1-score).
Based on the assessment findings, fine-tuning of the model is done by modifying hyperparameters and model architecture as appropriate [15]. Use regularization approaches such as dropout or L2 regularization to increase model generalization.
Many hyperparameters that shape deep learning models' behavior make them unique. These hyperparameters are examples of the learning rate, batch size, dropout rates, and the number of units inside each layer [17]. A detailed process of testing with several configurations to find the set of parameters that best fit the particular challenge is required to fine-tune these settings. The balance between the model's accuracy and flexibility is an art. Deep learning models are living, breathing things. They need ongoing maintenance and monitoring since their performance could deteriorate over time. Automated monitoring tools may spot problems, including model deterioration, idea drift, and changes in data distribution. The model must often be updated with new data to maintain its correctness.

IV. Data Analysis and Results

Exploratory data analysis (EDA) is an essential step in deep learning since it allows researchers to learn about a dataset's structure and properties. For my research we have used satellite image classification dataset RSI-CB256 and this dataset has 4 different classes from sensor and google map snapshot. About 5600 images are included in this collection, and they fall into four categories: vegetation, desert, water, and clouds. The following describes the breakdown of the dataset by category:

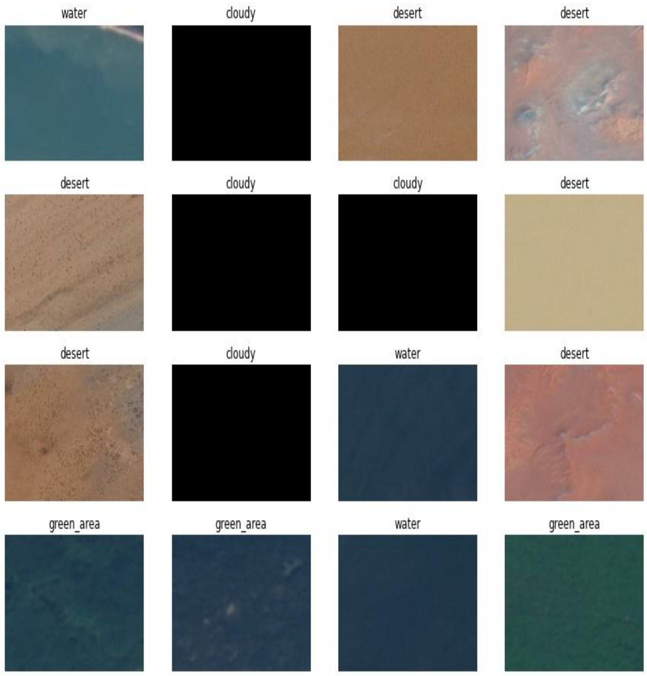

Figure 4: Image Categorization into Four Classes (Green area, Dessert, Water, Cloudy)

The four categories that illustrate the picture classification are green area, desert, water, and overcast. The distribution and make-up of the dataset can be grasped at a glance with the help of this visual representation, which is essential for developing and training deep learning models. The composition of the dataset displays the distribution of files across categories. The category "green area" has 1500 files, making it the biggest, followed by "water" with another 1500 files. There are 1,500 files in the "cloudy" category and 1,131 in the "desert" category. The class imbalance might affect the model's performance since it is more likely to categorize the overrepresented classes correctly but has more room for error with the underrepresented ones. When training a deep learning model, it is vital to consider the distribution of classes, as this will allow for class weighting and data augmentation to reduce bias and guarantee the model's capacity to generalize across classes. Table 1 represents loss accuracy during several 20 epochs.

Table 1: Epoch Vs Loss, Accuracy, Val_Loss, Val-Accuracy

Epoch	Loss	Accu	Val_Loss	Val_Acc
1	0.302	0.921	0.0323	1.000
2	0.081	0.982	0.0121	1.000
3	0.051	0.989	0.0065	1.000
4	0.044	0.988	0.0042	1.000
5	0.037	0.989	0.0030	1.000
6	0.032	0.990	0.0022	1.000
7	0.028	0.992	0.0017	1.000
8	0.025	0.993	0.0014	1.000
9	0.023	0.994	0.0011	1.000
10	0.020	0.993	9.3619e-04	1.000
11	0.022	0.993	8.3733e-04	1.000
12	0.020	0.994	7.1500e-04	1.000
13	0.017	0.994	5.8550e-04	1.000
14	0.017	0.995	5.1283e-04	1.000
15	0.015	0.996	4.2620e-04	1.000
16	0.016	0.995	4.0988e-04	1.000
17	0.015	0.995	3.5995e-04	1.000
18	0.014	0.996	3.4779e-04	1.000
19	0.015	0.995	3.2992e-04	1.000
20	0.015	0.995	3.0954e-04	1.000

The accuracy, recall, and F1-score are all lower (0.99), suggesting that class 3 (cloudy) is less reliable. Although the recall is significantly lower, the categorization performance is excellent. The model's overall performance in categorizing the photos is shown by the weighted average F1-score being 0.99 across all classes [18]. As measured by the average F1-score across all classes, the model's generalization ability is similarly excellent, at 0.99. The model shows a 99% accuracy on the test dataset, indicating that it is quite good at classifying data into the four categories. Figure 5 shows the decrease in training loss and val loss with increased epochs.

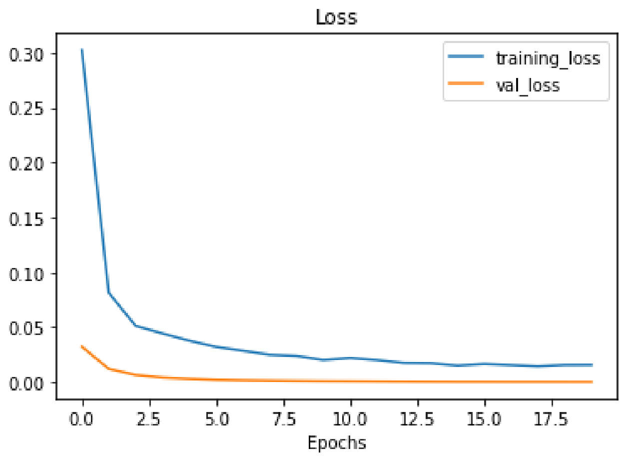

Figure 5: Epoch vs. Training loss and Val_Loss

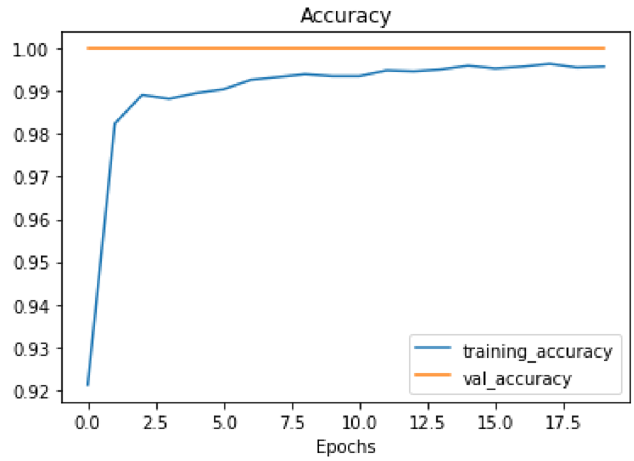

Figure 6: Epoch vs Training Accuracy and Val_Accuracy

Figure 6 shows a significant improvement in training and val accuracy with multiple epochs. The classification report provides a comprehensive analysis of the efficacy of a multi-class classification model. The goal of this model is to classify photos as either "green area," "desert," "water," or "cloudy." An essential parameter, precision, reflects how well the model can make correct positive predictions [19]. Figure 7 indicates the confusion metrics of the deep learning model (ResNet 50). For both "green area" and "desert," the model achieves a flawless score of 1.00, indicating that every one of its positive predictions was correct. With an accuracy of 0.97, the "water" class can properly label most photos depicting water. The "cloudy" classification likewise has a precision of 1.00, proving its superior accuracy.

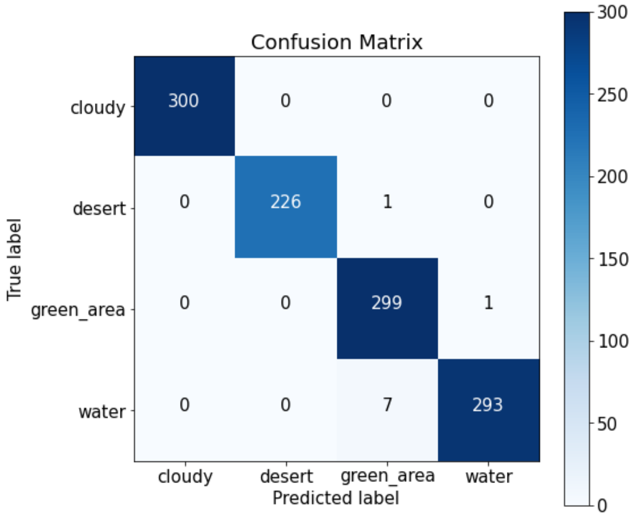

Figure 7: Confusion Matrix of DL Model

A critical parameter is recall, which measures how successfully the model identifies true positives. Again, the flawless recall score of 1.00 for "green area" and "desert" indicates that the model correctly identified all real-world occurrences of these classes. Class "water" receives a perfect score of 1.00, suggesting that all situations involving water can be recognized without error. "cloudy" achieves a good recall of 0.98, meaning that 98% of really cloudy occurrences were recognized correctly. The F1 score is often used to cope with unbalanced datasets since it compromises accuracy and recall. Here, "green area" and "desert" have F1 scores of 1.00, indicating a perfect equilibrium between accuracy and recall. The "water" category performs well, with an F1-score of 0.99. The F1-score of 0.99 for the word "cloudy" demonstrates an excellent balance between accuracy and recall. The model achieves an astonishing 99% overall accuracy, indicating its superior performance when categorizing photos into one of four predetermined categories. This degree of precision demonstrates the competence and predictive abilities of the model.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	300
1	1.00	1.00	1.00	227
2	0.97	1.00	0.99	300
3	1.00	0.98	0.99	300
accuracy			0.99	1127
macro avg	0.99	0.99	0.99	1127
weighted avg	0.99	0.99	0.99	1127

Figure 8: Accuracy, Precision, and F1 of ResNet 50

The above diagram shows the accuracy, precision, and f1 scores of the ResNet 50 deep learning model. A further affirmation of the model's stability comes from the macro and weighted averages of accuracy, recall, and F1-score. The model's macro and weighted averages are 0.99, demonstrating superiority across all classes and weighting schemes. The deep learning model showed 99 percent accuracy in predicting satellite images.

V. Discussions

The classification report's findings and insights give a thorough assessment of the deep learning model's efficacy in classifying photographs into four categories: "green area," "desert," "water," and "cloudy." This article explains the relevance of these results and their consequences for machine learning and computer vision. Each category's accuracy score exemplifies the model's propensity for reliable positive predictions. Its 100% accuracy for the classifications "green area" and "desert" is impressive, indicating that every positive prediction made for these categories was correct. In the instance of "water," it displays outstanding accuracy in detecting photos linked to water, with a precision of 0.97. The "cloudy" classification consistently displays an excellent precision of 1.00. This accuracy level is crucial when preventing false positives is essential.

A model's recall score measures how well it can recognize genuine positive examples. Again, the terms "green area" and "desert" perform very well, with a recall of 1.00, suggesting that the model correctly identifies every occurrence of these classifications. With a perfect recall of 1.00, class "water" demonstrates its superior capacity to recognize all genuine occurrences of water-related visuals. With a recall of 0.98 for "cloudy," the model performs well by properly labeling 98% of all category occurrences. Strong recall is essential when missing true positives might have profound implications. The F1 scores may be beneficial when working with unbalanced datasets due to their balance between accuracy and recall. In this report, "green area" and "desert" get perfect F1 scores of 1.00, demonstrating a perfect equilibrium between accuracy and recall. Keeping an F1 score of 0.99 indicates an excellent balance between accuracy and recall for the "water" class. Like "clear," "cloudy" achieves an F1-score of 0.99, demonstrating an exceptional equilibrium between these two dimensions. These impressive F1 scores show that the model can produce accurate predictions with a high level of recall, which is helpful in many practical settings.,= the model is competent in picture classification, as shown by its 99% accuracy. This high accuracy score indicates the model's predictive power and highlights its potential in automated content labeling and picture categorization. All classes are consistently represented by the model's high performance, as shown by the model's weighted and global averages for accuracy, recall, and F1 score. The model's versatility and capacity to accommodate different class imbalances are hinted at by its resilience.

VI. Conclusion

The primary goal of the model was to classify pictures into four categories: "green area," "desert," "water," and "cloudy." Overall accuracy, F1 scores, precision, and recall all point to the outstanding competency of the model. High precision scores are routinely achieved, with perfect precision for "green area" and "desert." These scores indicate the reliability of the model's accurate positive predictions. Recall scores evaluate how well the model can spot genuine positive cases and perform admirably and consistently. The F1 scores show the model's equilibrium between accuracy and recall. The model's 99% accuracy highlights its proficiency in correctly categorizing photos. This level of precision indicates its potential usefulness in various contexts, such as automatic content labeling and picture categorization. The macro and weighted averages for accuracy, recall, and F1-score confirm the model's consistent and robust performance across all classes. This durability indicates the model's capability to deal with a wide variety of datasets in terms of complexity and class imbalance.In future with the help of satellite image processing we can track rovers and landers like Vikram, Pragyan and Perseverance. This kind of satellite image prediction using deep learning can help NASA and ISRO in future for tracking LRV and Chandrayaan like space programmes.

Limitation of Research

The deep learning model will be used in a real-world system in the last stages of the voyage to make predictions based on newly arriving data. This procedure requires careful consideration of model serialization, API interaction, and scalability to support real-time queries. The gap between the model's capabilities and its usefulness is closed by a successful deployment.

Deep learning models are living, breathing things. They need ongoing maintenance and monitoring since their performance could deteriorate over time. Automated monitoring tools may spot problems, including model deterioration, idea drift, and changes in data distribution. The model must often be updated with new data to maintain its correctness.

An optimization and fine-tuning phase will start if the model's performance is unsatisfactory. Changing the model architecture and hyperparameters and gathering new data is necessary. It is essential to address issues like data bias and class inequality. Dropout, batch normalization, and regularization techniques are used to improve the model's capacity for generalization to new data and to reduce the danger of overfitting.

References

[1] Esmaeilzehi, A., Ahmad, M. O., & Swamy, M. (2021, November). SRNHARB: A deep lightweight image super-resolution network using hybrid activation residual blocks. *Signal Processing: Image Communication*, *99*, 116509. https://doi.org/10.1016/j.image.2021.116509

[2] Lu, T., Wang, J., Zhang, Y., Wang, Z., & Jiang, J. (2019, July 4). Satellite Image Super-Resolution via Multi-Scale Residual Deep Neural Network. *Remote Sensing*, *11*(13), 1588. https://doi.org/10.3390/rs11131588

[3] Wang, L., & Sun, Y. (2022, March 10). Image classification using a convolutional neural network with wavelet domain inputs. *IET Image Processing*, *16*(8), 2037–2048. https://doi.org/10.1049/ipr2.12466

[4] Choi, S. P. (2017, March 15). Extraction of Protein-Protein Interactions based on Convolutional Neural Network (CNN). *KIISE Transactions on Computing Practices*, *23*(3), 194–198. https://doi.org/10.5626/ktcp.2017.23.3.194

[5] Nandini, B. (2021, July 10). Detection of Skin Cancer using Inception V3 And Inception V4 Convolutional Neural Network (CNN) For Accuracy Improvement. *RevistaGestãoInovação E Tecnologias*, *11*(4), 1138–1148. https://doi.org/10.47059/revistageintec.v11i4.2174

[6] Park, J., Kim, H., & Paik, J. (2021, April 20). CF-CNN: Coarse-to-Fine Convolutional Neural Network. *Applied Sciences*, *11*(8), 3722. https://doi.org/10.3390/app11083722

[7] CHANDRASEKARAN, A. and KALLA , D. (2023) Heart disease prediction using chi -square test and linear regression. Computer Science & Information Technology, 13, pp. 135-146.

[8] Kalla, D., Samaah, F., Kuraku, S. & Smith, N. Phishing Detection Implementation Using Databricks and Artificial Intelligence. SSRN Electronic Journal **185**, doi:10.2139/ssrn.4452780 (2023).

[9] Wang, W., Zhu, M., Chi, X., & Xu, H. (2021, June). S-CNN-ESystem: An end-to-end embedded CNN inference system with low hardware cost and hardware-software time-balancing. *Journal of Systems Architecture*, p. *116*, 102122. https://doi.org/10.1016/j.sysarc.2021.102122

[10] Masoudi, B., &Danishvar, S. (2022). Deep multi-modal schizophrenia disorder diagnosis via a GRU-CNN architecture. *Neural Network World*, *32*(3), 147–161. https://doi.org/10.14311/nnw.2022.32.009

[11] Fabricius, D. (2016, November 16). Architecture before architecture: Frei Otto’s ‘Deep History.’ *The Journal of Architecture*, *21*(8), 1253–1273. https://doi.org/10.1080/13602365.2016.1254667

[12] Kim, J. (2022, October 17). Making Architecture Relevant to Underserved Communities: Mapping Reconsidered. *Architecture*, *2*(4), 637–659. https://doi.org/10.3390/architecture2040034

[13] Gallaher, J. (2020). Division (Architecture 5), and: "Adopt" as in "a Program" (Architecture 11), The Second Idea of Absence (Architecture 12), Nest (Architecture 17), and Nest III (Architecture 102). *The Missouri Review*, *43*(4), 41–47. https://doi.org/10.1353/mis.2020.0048

[14] Rani, P. A. S., & Singh, N. (2022, August 8). Paddy Leaf Symptom-based Disease Classification Using Deep CNN with ResNet-50. *International Journal of Advanced Science Computing and Engineering*, *4*(2), 88–94. https://doi.org/10.30630/ijasce.4.2.83

[15] Loey, M., Manogaran, G., Taha, M. H. N., & Khalifa, N. E. M. (2021, February). Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable Cities and Society*, p. *65*, 102600. https://doi.org/10.1016/j.scs.2020.102600

[16] Kalla, D., & Chandrasekaran, A. (2023). HEART DISEASE PREDICTION USING MACHINE LEARNING AND DEEP LEARNING. *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, *13*(3). https://doi.org/10.5121/ijdkp.2023.13301.

[17] Tchórzewski, J., &Wielgo, A. (2021, December 23). A neural model of human gait and its implementation in MATLAB and Simulink Environment using Deep Learning Toolbox. *Studia Informatica*, *25*. https://doi.org/10.34739/si.2021.25.03

[18] McKenzie, A. L., Muñoz, C. X., & Armstrong, L. E. (2015, December 1). Accuracy of Urine Color to Detect Equal to or Greater Than 2% Body Mass Loss in Men. *Journal of Athletic Training*, *50*(12), 1306–1309. https://doi.org/10.4085/1062-6050-51.1.03

[19] Stamoulis, D. T., & Hauenstein, N. M. (1993, December). Rater training and rating accuracy: Training for dimensional accuracy versus training for rate differentiation. *Journal of Applied Psychology*, *78*(6), 994–1003. https://doi.org/10.1037/0021-9010.78.6.994

Author Profile

Dinesh Kalla works at Microsoft as a Big Data and Azure CloudEscalation Engineer and has eight years of industry experience as a .Net Developer, BIDeveloper, Data Engineer, and Azure Cloud Engineer. His main areas of expertise and research interestare Big Data Analytics, Data Science, Machine Learning, Artificial Intelligence, IoT, and Cybersecurity. He published several papers on Big data, Artificial Intelligence, and cybersecuritythreats in International Journals and conferences. He completed his Master of Science at the University of New Havenand is pursuing a professional doctorate in Computer Science with an emphasis on Big Data Analytics from Colorado Technical University.

Nathan Smith holds a Master's in Wireless Communications and an MBA in Technology Management from National University. For over 20 years, he has worked as a technical writer developing documentation for the aerospace and defense sectors. He is currently pursuing a professional doctorate in Computer Science with a focus on Big Data Analytics from Colorado Technical University. He leverages his expertise in XML, DITA, and Agile methodologies to manage large sets of unstructured data and transform them into structured technical manuals and guides. His research interests involve utilizing natural language processing and large language models to extract insights from technical documentation and to develop new documentation for highly complex systems. He seeks to apply his background in technical communication and data analytics to support emerging autonomous vehicle technologies.

Fnu Samaahcurrently working in U.S. bank as a Java Full Stack Developer and Team Lead.Samaah has six years of industry experience as a Java Full Stack Developer and Tech Lead. Her main areas of expertise and research interestare Python, Data Analysis, Machine Learning, Chatbotsand Cybersecurity. She published several papers on Chatbots, Artificial Intelligence and cybersecurityawareness in International Journals. She completed her Master ofComputer Science from Northeastern Illinois University and Master’s in data science from Harrisburg University of Science and Technology.