GMR Institute of Technology
An Autonomous Institute Affiliated to JNTU-GV Vizianagaram

TERRAIN CLASSIFICATION FOR ENHANCED AUTONOMUS SYSTEMS

A project report submitted in partial fulfilment of the requirement

for the award of degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted by

CH. Bhavya Sai	**(20341A0525)**
V. Sai Surya Teja	**(20341A05I2)**
G. Kuswanth	**(20341A0559)**
B.Bikash	**(20341A0527)**
A.Prudhvi	**(20341A0515)**

Under the esteemed guidance of

Dr.K.Lakshmana Rao

Associate Professor Dept. of CSE

GMR Institute of Technology

An Autonomous Institute Affiliated to JNTU-GV, Vizianagaram

(Accredited by NBA, NAAC with 'A' Grade & ISO 9001:2015 Certified Institution)

GMR Nagar, Rajam – 532127,
Andhra Pradesh, India

March 2024

GMR Institute of Technology
An Autonomous Institute Affiliated to JNTU-GV Vizianagaram

GMRIT
Training Tomorrow's Engineers Today

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the thesis entitled **TERRAIN CLASSIFICATION FOR ENHANCED AUTONOMUS SYSTEMS** submitted by **CH.Bhavya Sai (20341A0525), V.Sai Surya Teja (20341A05I2), G.Kuswanth (20341A0559), B.Bikash (20341A0527), A.Prudhvi (20341A0515)** has been carried out in partial fulfilment of the requirement for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **GMRIT, Rajam** affiliated to **JNTUGV, Vizianagaram** is a record of Bonafide work carried out by them under my guidance & supervision. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree.

Signature of Supervisor
Dr.K.Lakshmana Rao
Associate Professor
Department of CSE
GMRIT, Rajam.

Signature of HOD
Dr. A. V. Ramana
Professor & Head
Department of CSE
GMRIT, Rajam.

The report is submitted for the viva-voce examination held on ………………..

Signature of Internal Examiner

Signature of External Examiner

ACKNOWLEDGEMENT

It gives us an immense pleasure to express deep sense of gratitude to my guide Dr.K.Lakshmana Rao , Associate Professor of Computer Science and Engineering for his whole hearted and invaluable guidance throughout the project work. Without her sustained and sincere effort, this project work would not have taken this shape. She encouraged and helped us to overcome various difficulties that we have faced at various stages of our project work.

We would like to sincerely thank our Head of the department Dr. A. V. Ramana, for providing all the necessary facilities that led to the successful completion of our project work.We would like to take this opportunity to thank our beloved Principal Dr.C.L.V.R.S.V.Prasad, for providing all the necessary facilities and a great support to us in completing the project work. We would like to thank all the faculty members and the non-teaching staff of the Department of Computer Science and Engineering for their direct or indirect support for helping us in completion of this project work.

Finally, we would like to thank all our friends and family members for their continuoushelp and encouragement.

Ch.Bhavya Sai	**(20341A0525)**
V.Sai Surya Teja	**(20341A05I2)**
G.Kuswanth	**(20341A0559)**
B.Bikash	**(20341A0527)**
A.Prudhvi	**(20341A0515)**

ABSTRACT

Terrain classification is an essential aspect of many applications, including robotics, autonomous vehicles, and military operations. It involves the categorization of different types of terrains based on their physical characteristics, such as texture, elevation, and surface composition. This process helps machines and systems understand and adapt to various landscapes, enabling them to make informed decisions about how to navigate and interact with the environment. To achieve accurate terrain classification, various techniques are used, such as deep learning algorithms, transfer learning techniques, Auto Encoders and Vision transformers. These methods analyze data from sensors like LiDAR, cameras, and radar to make informed decisions about the nature of the ground. By distinguishing between categories like flat surfaces, slopes, vegetation, water bodies, and obstacles, these systems can enhance their navigation and decision-making capabilities. Accurate terrain classification is crucial for enhancing the navigation and decision-making capabilities of autonomous systems. It plays a vital role in path planning, obstacle avoidance, and overall situational awareness. Autonomous vehicles require precise terrain classification for route planning and obstacle avoidance, while robots engaged in disaster relief missions rely on terrain classification to safely navigate through complex environments, On-going research in terrain classification aims to improve the robustness and efficiency of these systems across diverse environmental conditions. As technology advances, there is a growing need for more reliable and efficient terrain classification systems that can operate in various environmental conditions.

Keywords: Terrain Classification, Autonomous Systems, Navigation, LiDAR, surface composition.

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVATIONS

DEM: Data Elevation Model

ViT: Vision Transformers

BP: Back Propogation

SVM: Support Vector Machine

CNN: Convolutional Neural Networks

SIH: Smart India Hackathon

DRDO: Defence Research and Development Organisation

PolSAR: Polarimetric synthetic aperture radar

Chapter – 1

INTRODUCTION

1. INTRODUCTION

Terrain classification is a fundamental task with widespread applications in various fields such as robotics, autonomous vehicles, and military operations. The ability to accurately classify terrains plays a crucial role in enabling machines to navigate diverse landscapes effectively. By leveraging sensor data, including camera data and acoustic information, terrain classification algorithms empower machines to understand and interpret different terrain characteristics, which are essential for making informed decisions. Terrain classification is pivotal across various domains, notably in robotics, autonomous vehicles, and military operations, where effective navigation relies on understanding diverse landscapes. Leveraging sensor data, camera data and acoustic information, terrain classification enables machines to discern terrain characteristics crucial for informed decision-making.

The project described in this study focuses on leveraging deep learning techniques for terrain classification, with a specific emphasis on military applications and decision-making processes. By employing transfer learning models and vision transformers, augmented with preprocessing techniques, the study aims to develop robust terrain classification algorithms capable of accurately distinguishing various terrain features.

Furthermore, the study conducts a comparative analysis among different deep learning models, including transfer learning models and vision transformers, to identify the most accurate and efficient model for terrain classification tasks. By evaluating the performance of each model based on metrics such as accuracy, computational efficiency, and generalization capability, the study aims to provide insights into the strengths and limitations of different approaches.

Notably, terrain classification emerges as a critical challenge, underscored by its inclusion in the Ministry of Defence's problem statement during the Smart India Hackathon 2023, highlighting its significance in addressing real-world needs. Additionally, the dataset used in this study is recommended by the Ministry of Defence, ensuring relevance and applicability to military applications.

In summary, the project addresses the critical challenge of terrain classification, with a focus on its practical applications in military operations and autonomous systems. By leveraging state-of-the-art deep learning techniques and conducting a comprehensive evaluation of different models, the study aims to contribute to the development of robust and effective terrain classification algorithms with real-world relevance and applicability.

Chapter – 2

LITERATURE SURVEY

2. LITERATURE SURVEY

Terrain Classification for Robotic applications:

Reference [12] by Manduchi et al. (2015) introduces novel sensor processing algorithms tailored for autonomous navigation in challenging cross-country environments. The study focuses on obstacle detection techniques, color-based classification systems, and ladar data analysis to ensure safe navigation by identifying obstacles and distinguishing between terrain types. While successful in highlighting the significance of autonomous navigation in challenging environments, the study acknowledges potential limitations in accuracy due to lighting variations and complex scenes, suggesting refinements in color-based classification and the integration of additional sensors for enhanced perception.

In [4] by Christie & Kottege (2016), a real-time terrain classification system using acoustic data is developed. The study employs audio data from interactions on various terrains and applies Support Vector Machines (SVM) for classification. While achieving sensitivity improvements with noise removal techniques, the study identifies challenges in distinguishing certain terrains. Future scope suggests testing the system in diverse environments to evaluate its practical applicability and performance under varied conditions.

Wellhausen et al. (2019) in [36] propose predicting terrain properties from images via self-supervised learning. The study utilizes the ERFNet + skip connection regression network for this purpose. While focusing on predicting terrain properties, the study lacks direct comparison metrics. Nonetheless, it highlights self-supervised learning as a promising avenue for improved accuracy in terrain classification.

Pokonieczny (2018) in [44] addresses terrain passability classification using Self-Organizing Maps. The study successfully utilizes SOM to generate an Index of Passability (IOP) and categorizes terrain passability based on it, albeit with a limited focus on specific regions in northeastern Poland.

In [2] by Bai et al. (2019), a 3-D vibration-based terrain classification for mobile robots is proposed. The study employs Fast Fourier Transformation for BP neural network training, achieving 98% accuracy across five terrain types. However, limitations in precise rover data and neural network algorithms are identified, suggesting avenues for improvement.

Shi et al. (2020) in [3] investigate vibration-based terrain classification with a modified Laplacian SVM. The study integrates terrain images and axle-mounted vibration data, showcasing significant improvement with the modified Laplacian SVM. However, limitations in computational complexity and the lack of comparative analysis are noted, suggesting areas for further research.

Abete et al. (2021) in [10] propose terrain a pioneering approach to terrain identification for humanoid robots, leveraging Convolutional Neural Networks (CNNs) and utilizing raw sensor data. Their study attains remarkable success, achieving an accuracy exceeding 98% across six distinct terrain types. However, the researchers highlight a notable challenge regarding the potential impact of variations in walking velocities on classification accuracy. This underscores the necessity for deeper exploration and investigation into the robustness of their method, pointing towards avenues for future research and development in this field.

In [11] by Abete et al. (2021), introduced an innovative terrain classification algorithm employing monocular video data. This comprehensive study amalgamates multiple techniques to yield a commendable accuracy level, nearing 80%. Despite this achievement, the researchers identify significant hurdles related to the fluctuating walking speeds and diverse ground types encountered. These findings underscore the pressing need for enhancements in the algorithm's adaptability, presenting valuable avenues for further research and refinement in this domain.

Ahmadi et al. (2021) in [16] employ semi-supervised gated recurrent neural networks (RNNs) for robotic terrain classification. Their research marks a notable advancement in the field, showcasing a mean accuracy of 83.68%. This achievement represents a substantial improvement over existing literature, demonstrating the efficacy of their approach. However, the study highlights initial limitations in the effectiveness of RNN architectures, pinpointing specific areas ripe for optimization and further refinement. These insights provide valuable directions for future research endeavours aimed at enhancing the performance and versatility of RNN-based terrain classification systems.

Grezmak et al. (2021) in [19] showcase a groundbreaking application of compliance sensing for terrain classification utilizing a hexapod robot. Their study presents compelling evidence of success by employing Support Vector Machine (SVM) classifiers. The findings underscore the significant potential of compliance sensing technology in enhancing the locomotion capabilities of legged robots. This research represents a crucial step forward in the field, offering valuable insights into the practical implementation and efficacy of compliance sensing for terrain classification tasks, thereby paving the way for advancements in legged robot mobility and navigation.

Li et al. (2020) in [20] introduce C-Terrain, a pioneering approach for quadruped robot terrain prediction. The study places particular emphasis on the importance of stability and superiority in the context of terrain prediction algorithms. Notably, their methodology yields exceptional results, attaining the highest accuracy levels through the implementation of the KNN regression method. This research signifies a significant advancement in the field of robotic terrain prediction, showcasing the efficacy of their approach and its potential to enhance the performance and reliability of quadruped robots in various real-world scenarios.

In [22] by Wu et al. (2016), present an innovative framework focused on ground reaction force sensing and terrain classification tailored for small legged robots. Their study impressively achieves accuracy levels surpassing 90% within a single stride. However, the researchers conscientiously recognize potential limitations stemming from dataset diversity. This acknowledgment underscores the importance of addressing dataset biases and expanding diversity in future research endeavours, thereby ensuring the robustness and generalizability of terrain classification algorithms for small legged robots.

Li et al. (2023) in [21] introduce an innovative wearable computer vision system enhanced with a gimbal mechanism, specifically designed for terrain classification tasks. Their study showcases commendable success in achieving high accuracy levels, particularly in outdoor environments. However, the researchers astutely acknowledge the imperative for enhancing real-time accuracy. This recognition of the need for ongoing refinement and optimization underscores their commitment to advancing the practical utility and effectiveness of wearable computer vision systems for terrain classification applications, thereby paving the way for enhanced performance and broader adoption in real-world scenarios.

Zhao et al. (2021) in [23] propose a novel terrain classification and adaptive locomotion strategy for hexapod robots, achieving an impressive 96.67% accuracy. Their approach leverages joint torques and IMU data to enhance stability and mobility in outdoor settings. By employing Support Vector Machine (SVM) algorithms and dynamic alternating tripod trotting gait, they showcase significant advancements in terrain classification and adaptable locomotion. This research signifies a pivotal step towards enhancing robotic navigation capabilities, with implications for various real-world applications.

Finally, **Lv et al. (2020) in [27]** present FT-S2ELM, a novel method for robotic terrain classification, demonstrating an accuracy rate of 89%. Despite the notable achievement, the research notably omits a comparison with established methods, indicating potential avenues for future investigation and evaluation. This observation underscores the importance of comprehensive benchmarking and comparative analysis in advancing the understanding and efficacy of terrain classification techniques for robotics applications.

DEM Data Papers:

Pingel et al. (2013) [5]: Pingel et al. developed the SMRF algorithm for terrain classification using LIDAR data, aiming for high accuracy in ground filtering and terrain modeling for immersive virtual environments. SMRF utilizes a progressive morphological filter to identify ground/object points and achieved competitive results compared to other methods.

Li et al. (2020) [6]: Li et al. implemented a deep learning-based approach for landform classification using integrated data sources of digital elevation models and imagery. Their method outperformed the random forest method, showcasing higher accuracy and more distinct landform boundaries. However, the dependency on specific datasets for network training and ambiguous boundaries of certain landform entities were noted limitations.

Na et al. (2021) [24]: Na et al. proposed a novel method for large-scale terrain classification in China, combining an object-based framework with the random forest algorithm. They achieved an overall accuracy of 80.53% by refining segmentation outcomes and selecting features from DEMs, surpassing existing methods. The study demonstrated the method's transferability in provincial-scale mapping with a different classification system.

Yang et al. (2023) [1]: Yang et al. demonstrated the effectiveness of deep learning in automatically classifying landforms using high-resolution DEM data. They trained a semantic segmentation model using DEM data and achieved stable performance with precision accuracy at 80% and Mean Intersection over Union (MIoU) at 68%. The study highlighted the significance of deep learning in geomorphological research.

Iwahashi et al. (2018) [25]: Iwahashi et al. advanced global terrain classification using 280m DEMs, segmenting areas based on slope characteristics and clustering similar polygons using surface texture information. While enhancing classification for smaller landform elements, challenges arose in accurately describing specific geometric signatures.

Dai et al. (2020) [29]: Dai et al. introduced an integrated approach for mapping terrace risers on the Loess Plateau of China using remote sensing and elevation data. They achieved high precision (90.81% to 97.57%) in mapping terrace risers, demonstrating the method's accuracy for agricultural terrace delineation.

Na et al. (2021) [24]: Na et al. proposed a novel method for large-scale terrain classification in China, integrating an object-based framework with the random forest algorithm. The approach refined segmentation outcomes and selected features from digital elevation models (DEMs) to achieve an overall accuracy of 80.53%, surpassing existing methods.

Yin et al. (2019) [43]: Yin et al. addressed the challenge of high computational costs in polarimetric SAR (PolSAR) image classification by optimizing feature selection. They achieved high classification accuracies using optimal combinations of polarimetric features for vegetation classification in PolSAR images.

Barrett et al. (2023) [30]: Barrett et al. classified terrain on Mars using deep learning, achieving mean IoU of 74.15% for descriptive classes and 92.33% for interpretive groups. They generated useful maps aiding in understanding the aeolian history and geology of the landing site, surpassing manual mapping capabilities.

These papers focus on terrain classification using DEM data, showcasing advancements in object-based frameworks, random forest algorithms, and feature selection techniques. These methods have improved accuracy in large-scale terrain classification and mapping, addressing challenges in accurately describing geometric signatures and reducing computational costs.

PolSAR Data Papers:

Shi et al. (2020) [26]: proposed a fast multi-feature joint sparse learning method for PolSAR image classification. By integrating three types of features and employing joint multi-feature sparse representation, their approach outperformed alternative methods. Utilizing SLIC for super pixel extraction and a topic model for high-level feature learning, the method achieved enhanced classification accuracy and computational efficiency. However, limitations such as misclassification in identifying buildings underscore the need for further refinement in feature extraction for PolSAR image classification.

Ai et al. (2021) [34]: proposed the TFF-ICAE algorithm for fine PolSAR terrain classification, merging texture features and deep learning. Evaluation on Flevoland and Oberpfaffenhofen datasets showcased its superior performance, achieving a classification accuracy of 97.61%. The study highlights the algorithm's potential in overcoming limitations of traditional methods, suggesting future exploration of deep CNNs for further enhancements in classification.

Liu et al. (2018) [35]: introduced the Spatial Multi-attribute Graph (SMAG) method for terrain classification using polarimetric SAR data. SMAG optimizes classification accuracy by fusing multiple features and considering physical data attributes. Achieving an average error rate of only 0.46%, SMAG outperforms other semi-supervised algorithms like SSA, SGS, and Hyper, showcasing its effectiveness in terrain classification. However, further research is warranted to evaluate SMAG's generalizability to diverse terrains and its computational scalability for large-scale classification applications.

Liu et al. (2019) [37] proposed the DSMR model, designed for feature extraction and classification of PolSAR (Polarimetric Synthetic Aperture Radar) data. This model demonstrated superior accuracy when compared to alternative algorithms. The DSMR model effectively extracts features and performs classification tasks, showcasing its efficacy in handling PolSAR data. Liu et al.'s research signifies significant advancements in the domain, offering a robust solution for PolSAR data analysis with improved accuracy.

Yang et al. (2019) [42]: introduced a CNN-based polarimetric decomposition feature selection method for PolSAR image interpretation. Their objective was to enhance terrain classification accuracy by leveraging a 1-D CNN algorithm that efficiently selects high-performance feature subsets using Kullback-Leibler distance. Employing polarimetric target decomposition algorithms and a 1-D CNN model with feature selection, their approach demonstrated superior performance in real PolSAR datasets. However, challenges such as potential computational overhead during CNN training and the black-box nature of CNNs were acknowledged, highlighting areas for further research to address model interpretability and hyperparameter sensitivity.

These papers emphasize PolSAR data analysis, presenting various algorithms and models for terrain classification. These methods leverage deep learning, feature selection, and multi-feature representation to enhance classification accuracy.

Terrain Classification Methods with Various Data Sources:

Zou et al. In [31] introduced a reservoir-based spiking neural network (r-SNN) for terrain classification, achieving remarkable test accuracy exceeding 95%. Their approach outperformed traditional methods like Support Vector Machine (SVM) and logistic regression models. The study utilized synthesized and real PolSAR datasets from different SAR systems.

Lv et al. In [32] proposed a multi-feature joint learning method for swift polarimetric SAR terrain classification. This method not only enhanced classification accuracy but also improved computational efficiency, offering advantages over conventional techniques. The study utilized polarimetric SAR data and explored techniques like joint multi-feature sparse representation and superpixel extraction.

Wellhausen et al. In [33] investigated ensemble methods and hybrid models to boost classification accuracy across diverse terrain types. Their study emphasized the importance of model selection in achieving accurate terrain classification results. The research didn't focus on specific datasets but rather on methodological advancements in terrain classification.

Zhao et al. in [34] discussed recent advancements in convolutional neural networks (CNNs) and their efficacy in terrain classification tasks. They emphasized the significance of model selection in achieving accurate classification results. The study didn't mention specific datasets but highlighted the importance of leveraging CNN architectures for terrain analysis.

Yang et al. In [35] proposed a CNN-based feature selection algorithm for PolSAR image classification, aiming to improve terrain classification accuracy. Their approach selected high-performance feature subsets efficiently, enhancing classification performance on real PolSAR datasets. The study utilized polarimetric SAR data and focused on refining feature selection techniques for terrain analysis.

These papers showcase diverse approaches to terrain classification, leveraging various data sources such as SAR images and neural networks. Each study contributes to advancing classification accuracy and computational efficiency, addressing challenges in terrain analysis and remote sensing applications.

Chapter – 3

REQUIREMENT SPECIFICATION

3. REQUIREMENT SPECIFICATION

3.1 Software Requirements

The software requirements are description of features and functionalities of the target system. Requirements convey the expectations of users from the software product. The requirements can be obvious or hidden, known, or unknown, expected or unexpected from client's point of view.

3.1.1 Operating System

Windows is a graphical operating system developed by Microsoft. It allows users to view and store files, run the software, play games, watch videos, and provides a way to connect to the internet. It was released for both home computing and professional works.

MacOS is the computer operating system (OS) for Apple desktops and laptops. It is a proprietary graphical OS that powers every Mac. OSes interact with a computer's hardware, allocating the resources necessary to complete tasks given to it, for example, running an application. OSes allocate
resources including memory, processing power and file storage.

Linux is an operating system. In fact, one of the most popular platforms on the planet, Android, is powered by the Linux operating system. An operating system is software that manages all of the hardware resources associated with your desktop or laptop. To put it simply, the operating system manages the communication between your software and your hardware. Without the operating system (OS), the software would not function.

3.1.2 SDK

SDK stands for software development kit or devkit for short. It's a set of software tools and programs used by developers to create applications for specific platforms. SDK tools will include a range of things, including libraries, documentation, code samples, processes, and guides those developers can use and integrate into their own apps. SDKs are designed to be used for specific platforms or programming languages.

Some of SDK's used in this project are stated below:

1. TensorFlow
2. Flask
3. Numpy
4. Scikit-Learn etc.

3.2 Hardware Requirements:

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware, A hardware requirements list is often accompanied by a hardware compatibility list (HCL), especially in case of operating systems. An HCL lists tested, compatible, and sometimes incompatible hardware devices for a particular operating system or application. The following sub-sections discuss the various aspects of hardware requirements. The Hardware Interfaces Required are:

1. Ram: Minimum 8GB or higher
2. GPU: 4GB dedicated
3. SSD: 128GB
4. Processor: Intel i5 10th Gen or Ryzen 5 with Octa core.

3.3 Non-Functional Requirements:

Non-Functional Requirements are the constraints or the requirements imposed on the system. They specify the quality attribute of the software. Non- Functional Requirements deal with issues like scalability, maintainability, performance, portability, security, reliability, and many more. NonFunctional Requirements address vital issues of quality for software system. It includes below things:
Capacity, Availability and Performance etc.

3.4 Python Libraries To be installed:

The following libraries with the specific versions type must be installed for this project to function. These can be installed using command “pip install library_name==version"

- numPy==1.23.3
- Pillow==9.2.0

- scikit-learn==1.1.2
- scipy==1.9.2
- sklearn==0.0
- tensorboard==2.10.0
- tensorflow==2.10.0
- tensorflow-estimator==2.10.0
- tensorflow-gpu==2.10.0
- tensorflow-hub==0.12.0

Chapter – 4

METHODOLOGY

4. METHODOLOGY

DEEP LEARNING

Deep learning is a subset of artificial intelligence (AI) that has revolutionized various fields such as computer vision, natural language processing, and speech recognition. It is inspired by the structure and function of the human brain, particularly the way neurons communicate with each other through synapses. At the core of deep learning are neural networks, which are computational models consisting of interconnected layers of nodes, or neurons. These networks are designed to learn from large amounts of data to perform tasks such as classification, regression, clustering, and generation.

One of the key features of deep learning is its ability to automatically discover and learn intricate patterns and features from raw data. This is achieved through a process called training, where the network is presented with input data along with corresponding labels or desired outputs. During training, the network adjusts its internal parameters, known as weights and biases, to minimize the difference between its predictions and the true labels. This optimization process is typically carried out using algorithms such as stochastic gradient descent.

Deep learning architectures can vary in complexity and depth, with some consisting of only a few layers (shallow networks) while others may have many layers (deep networks). Deep neural networks (DNNs) with multiple hidden layers are capable of learning hierarchical representations of data, which allows them to capture intricate patterns and relationships in the input.

Convolutional Neural Networks (CNNs) are a type of deep neural network commonly used in computer vision tasks such as image classification and object detection. CNNs leverage convolutional layers to automatically learn spatial hierarchies of features from raw pixel values, enabling them to effectively recognize objects in images. Recurrent Neural Networks (RNNs) are another important class of deep learning models frequently used in sequence modeling tasks like natural language processing and speech recognition. RNNs are designed to process sequences of data by maintaining internal memory, allowing them to capture temporal dependencies and long-range dependencies within the input.

In recent years, deep learning has seen significant advancements driven by improvements in computational power, availability of large-scale datasets, and algorithmic innovations. These advancements have led to breakthroughs in various domains, including healthcare (e.g., medical image analysis, drug discovery), finance (e.g., fraud detection, algorithmic trading), and autonomous vehicles.

Despite its remarkable success, deep learning still faces several challenges and limitations. One common issue is the need for large amounts of labeled data for training, which can be costly and time-consuming to acquire. Additionally, deep learning models are often considered "black boxes," making it difficult to interpret their decisions and understand the underlying reasoning process. This lack of interpretability raises concerns, particularly in safety-critical applications where transparency and accountability are crucial.

To address these challenges, researchers are exploring techniques such as transfer learning, which allows pre-trained models to be adapted to new tasks with limited labeled data, and explainable AI, which aims to provide insights into the decision-making process of deep learning models.

Deep learning represents a powerful paradigm for solving complex problems across various domains. Its ability to automatically learn representations from data has led to significant advancements in AI and continues to drive innovation in research and industry. However, addressing challenges related to data availability, interpretability, and robustness remains essential for realizing the full potential of deep learning in real-world applications.

IMAGE CLASSIFICATION

Image classification is a fundamental task in computer vision that involves categorizing images into predefined classes or categories. It typically utilizes machine learning algorithms, especially convolutional neural networks (CNNs), to automatically learn and extract features from input images. These features are then used to predict the most suitable class or label for each image. Image classification finds applications in various domains, including object recognition, medical diagnosis, autonomous driving, and security surveillance. It plays a crucial role in enabling machines to interpret and understand visual information, paving the way for advancements in artificial intelligence and automation.

DATASET COLLECTION AND PREPARATION:

The study commenced by accessing a dataset sourced from Kaggle, representing real-time terrain data compiled from diverse environments. This dataset encompassed various terrain classes such as grassy, marshy, rocky, and sandy terrains. To ensure comprehensive analysis, the dataset underwent meticulous processing, including division into an initial origin dataset and an augmented dataset. This division aimed to enhance the dataset's diversity and representativeness, enabling robust training and evaluation of terrain classification models. Description of the dataset is as follows.

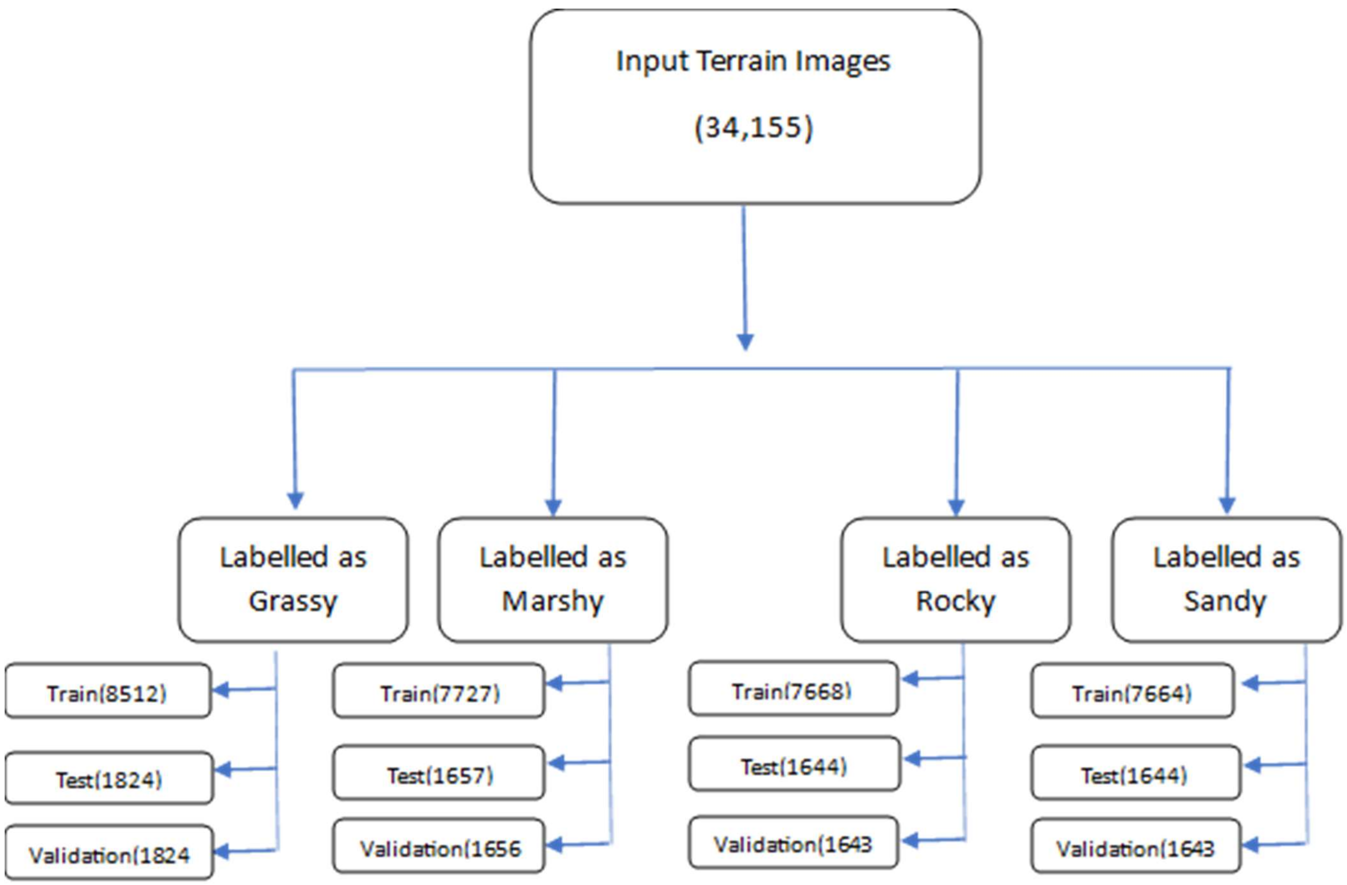

Fig 1: Label Distribution and Split Chart

Dataset Overview:

	Class	Train Count	Test Count	Validation Count
0	Marshy	7727	1657	1656
1	Rocky	7668	1644	1643
2	Sandy	7664	1644	1642
3	Grassy	8512	1824	1824

Fig 2: Label Split Summary Table

The dataset contributed by DRDO India for SIH2023, available on Kaggle, encompasses geo-stationary images collected in real-time, tailored for potential military applications. Consisting of a total of 34155 images, the dataset is segmented into training, testing, and validation subsets. However, specific geographical regions are not explicitly specified. Each image is categorized into one of four distinct labels: marshy, rocky, sandy, or grassy. Given its source and labeling, this dataset likely serves tasks such as image classification or remote sensing analysis, particularly focusing on land cover classification or environmental monitoring. With its potential implications in defense and military contexts, the dataset could find utility in surveillance, reconnaissance, or target identification endeavors.

Model Building:

Transfer learning models, such as VGG16, DenseNet-121, MobileNet, Inception V3, ResNet-101, DenseNet-201, and ResNet-50, are widely used in various computer vision tasks due to their pre-trained weights on large-scale datasets like ImageNet. VGG16 is characterized by its simplicity and consists of multiple convolutional and fully connected layers. DenseNet-121 employs densely connected blocks to promote feature reuse and parameter efficiency. MobileNet is designed for mobile and embedded applications, featuring lightweight depth-wise convolutions. Inception V3 utilizes inception modules with multiple filter sizes to capture diverse image features. ResNet-101 and ResNet-50 introduce residual connections to alleviate the vanishing gradient problem and facilitate training of deeper networks. These models offer varying trade-offs between accuracy, computational complexity, and memory footprint, making them suitable for different deployment scenarios.

ResNet-50 Model:

ResNet-50, a renowned convolutional neural network (CNN) architecture, comprises 50 layers, incorporating residual blocks with skip connections. These skip connections are pivotal in mitigating the vanishing gradient problem, facilitating the training of exceptionally deep networks. ResNet-50 stands out for its remarkable efficiency and effectiveness in diverse computer vision tasks, such as image classification and object detection. Its ability to handle deep networks effectively has made it a cornerstone in the field of deep learning, enabling breakthroughs in various visual recognition applications.

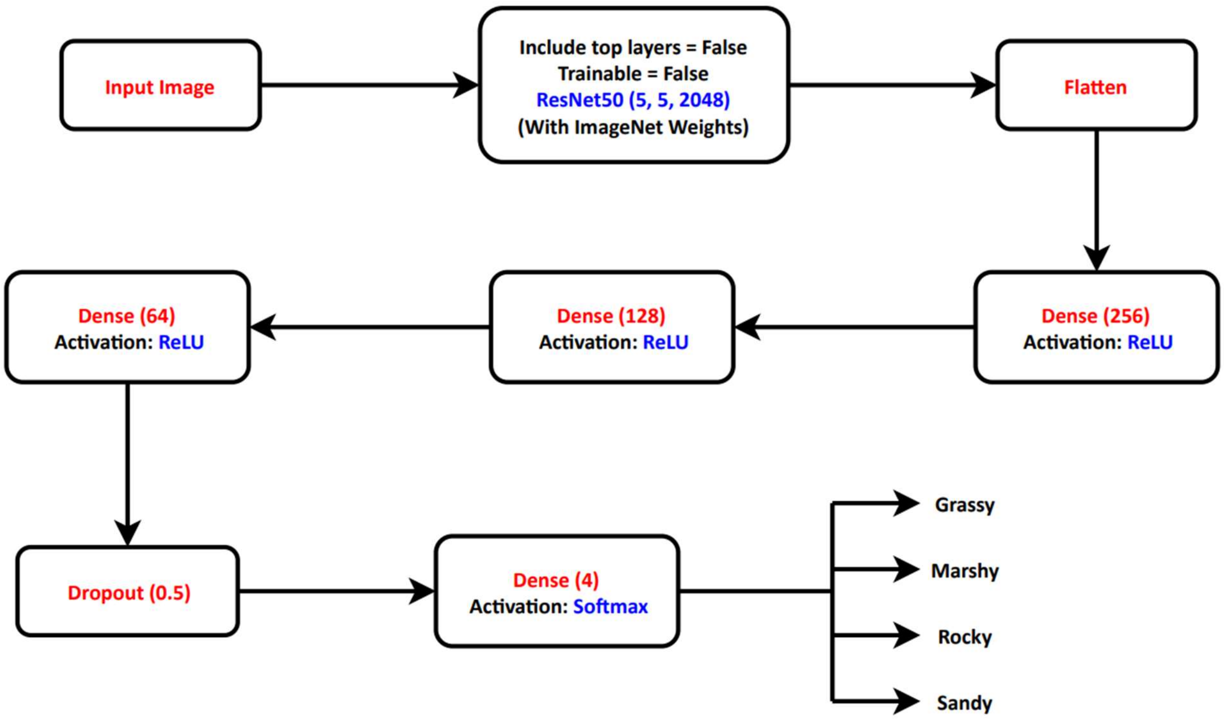

Fig 3: Architecture of ResNet50

Vision Transformer Model for ResNet-50:

The vision transformer model for ResNet-50 follows a similar architecture to that described previously. The pre-trained ResNet-50 base is augmented with additional layers for feature extraction and classification. After preprocessing the input image, it is passed through the ResNet-50 base with pre-trained weights. The top layers of the ResNet-50 model are excluded to prevent fine-tuning. Following the ResNet-50 base, a global average pooling layer and dense layers are added for feature extraction. These features are then processed using a multi-head attention mechanism, followed by layer normalization, dropout regularization, and a final dense layer for classification.

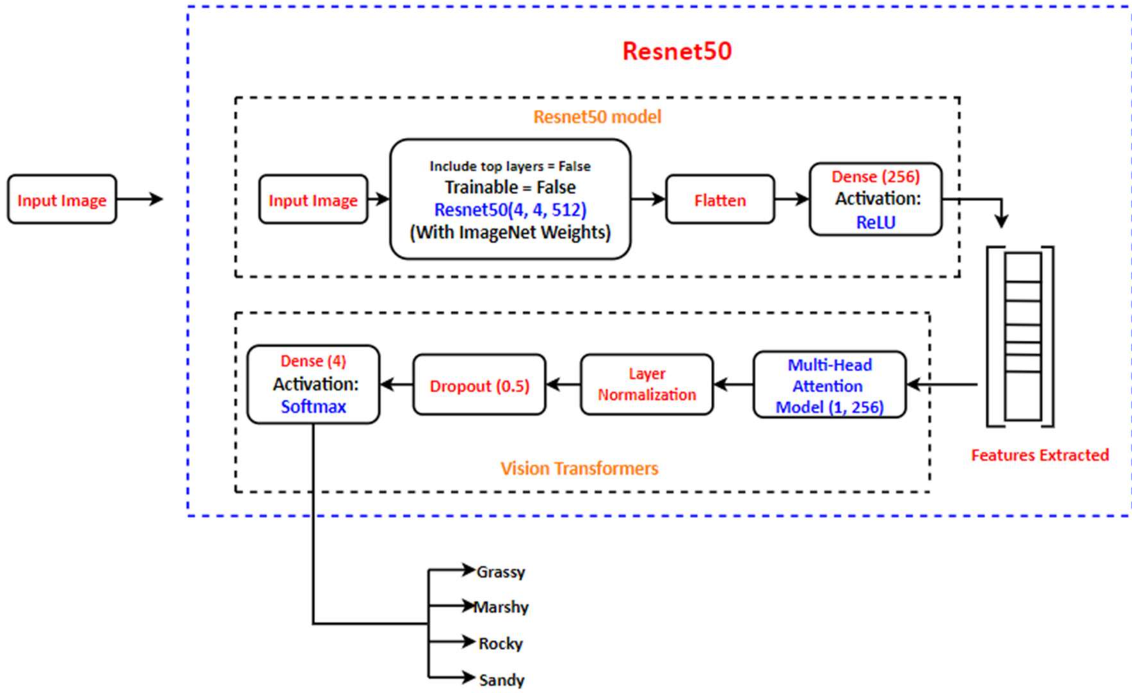

Fig 4: Architecture of ResNet50 with Vision Transformers

Image Input: The input image is pre-processed and augmented using techniques such as rotation, horizontal and vertical shifting, and flipping.

Transfer Learning: The pre-processed image is passed through the ResNet-50 base with pre-trained weights. The top layers of the ResNet-50 model are excluded to prevent fine-tuning.

Feature Extraction: The output features from ResNet-50 are obtained and further processed using additional layers for feature extraction.

Multi-Head Attention: The features are processed using a multi-head attention mechanism to capture relationships between different parts of the image.

Layer Normalization and Dropout: Layer normalization is applied to stabilize the training process, followed by dropout regularization to prevent overfitting.

Classification: Finally, the features are passed through a dense layer with a SoftMax activation function to perform classification into the desired classes.

ResNet-101 Model:

ResNet-101 stands as an advanced iteration of the ResNet architecture, offering a deeper network with 101 layers. It capitalizes on residual connections, which allow for the training of exceedingly deep neural networks while mitigating the challenge of vanishing gradients, thereby enabling more effective learning. Renowned for its versatility, ResNet-101 finds widespread application across various computer vision tasks, including image classification, object detection, and semantic segmentation. Its utilization extends to domains where intricate visual understanding is essential, owing to its ability to capture complex features and patterns across multiple layers. Furthermore, ResNet-101 often serves as a benchmark model due to its remarkable performance and widespread adoption in the research community.

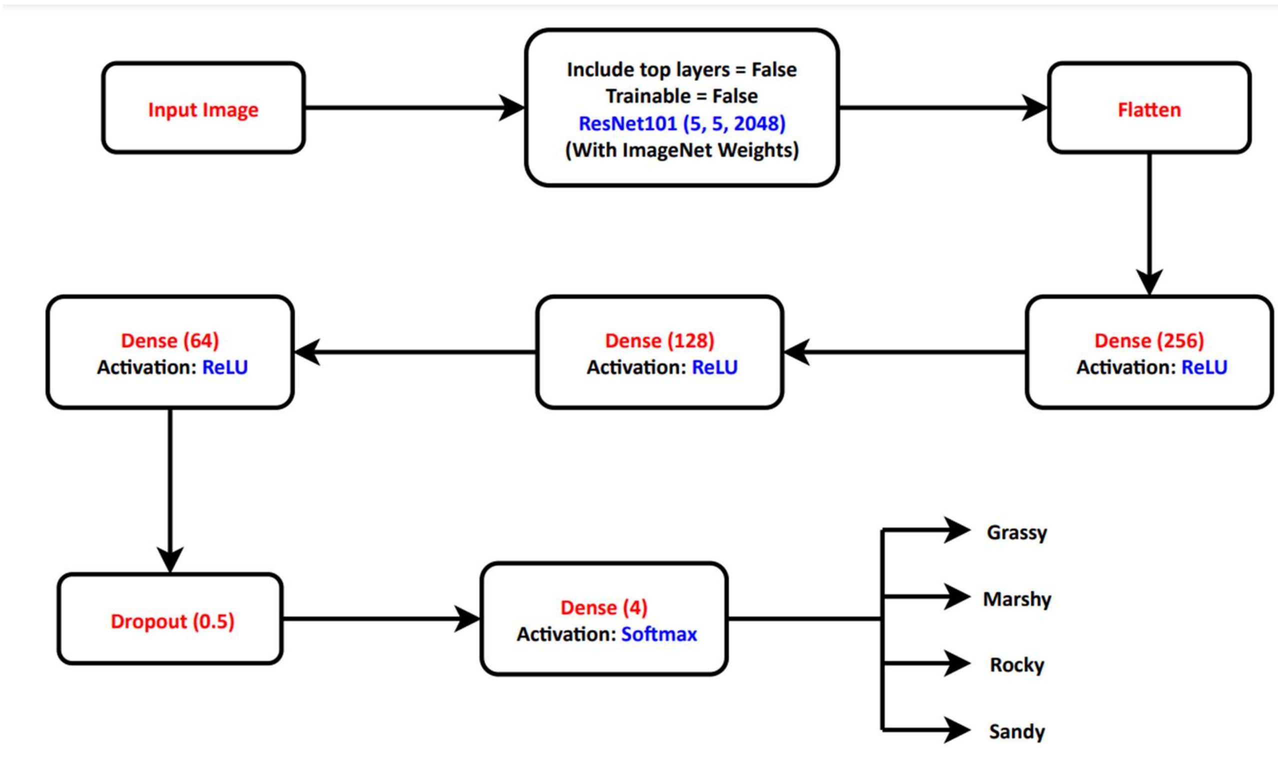

Fig 5: Architecture of ResNet101

Vision Transformer Model for ResNet-101:

Like the vision transformer model for ResNet-50, the vision transformer model for ResNet-101 augments the pre-trained ResNet-101 base with additional layers for feature extraction and classification. After preprocessing the input image, it is passed through the ResNet-101 base with pre-trained weights. The top layers of the ResNet-101 model are excluded to prevent fine-tuning.

Following the ResNet-101 base, a global average pooling layer and dense layers are added for feature extraction. These features are then processed using a multi-head attention mechanism, followed by layer normalization, dropout regularization, and a final dense layer for classification

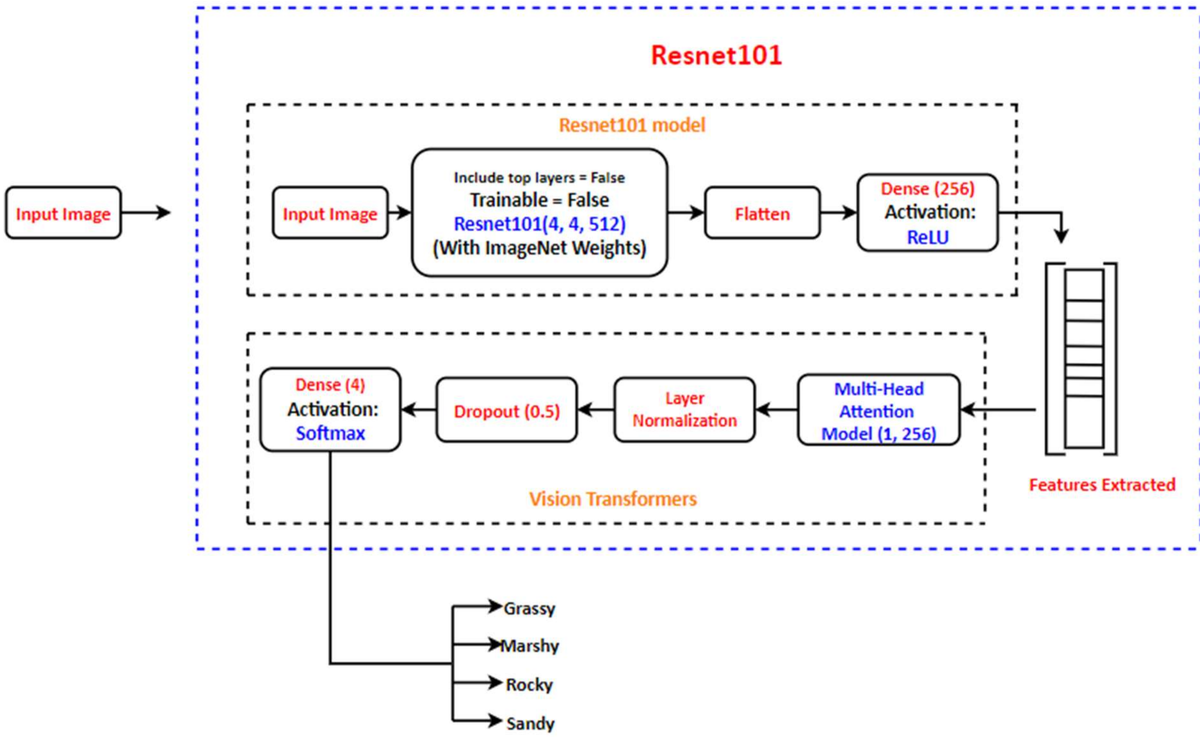

Fig 6: Architecture of ResNet101 with Vision Transformers

Image Input: The input image undergoes preprocessing and augmentation to enhance the diversity of training data.

Transfer Learning: The preprocessed image is fed into the ResNet-101 base with pre-trained weights, excluding the top layers to retain the original features.

Feature Extraction: Features are extracted from the ResNet-101 base and further processed to capture higher-level representations.

Multi-Head Attention: The features undergo multi-head attention to capture global dependencies and relationships between different parts of the image.

Normalization and Dropout: Layer normalization is applied to stabilize the training process, followed by dropout regularization to prevent overfitting.

Classification: The processed features are passed through a dense layer with a SoftMax activation function to perform classification into the desired classes.

The flow ensures that the ResNet-101 model's pre-trained weights are utilized for feature extraction, while the additional layers and multi-head attention mechanism enhance the model's ability to capture intricate patterns and relationships within the image.

DenseNet-201 Model:

DenseNet-201, an extension of the Dense Net architecture comprising 201 layers, is distinguished by its densely connected blocks. Unlike traditional architectures, where layers are connected in a sequential manner, DenseNet-201 incorporates dense connections, ensuring each layer receives input from all preceding layers. This design facilitates extensive feature reuse and parameter efficiency, enabling the network to capture intricate patterns and details effectively. Due to these characteristics, DenseNet-201 emerges as a versatile solution for a wide range of computer vision tasks, including image classification, object detection, and segmentation. Its dense connectivity fosters enhanced feature learning and enables the model to achieve impressive performance on various visual recognition tasks.

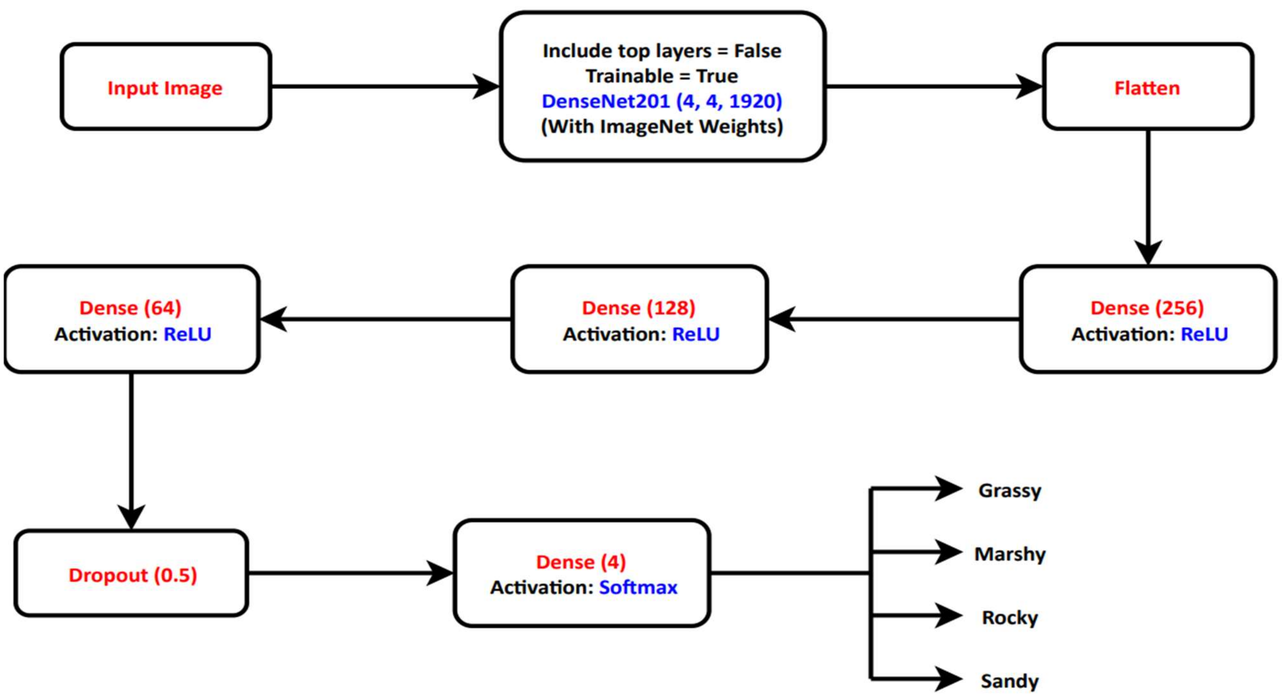

Fig 7: Architecture of DenseNet201

Vision Transformer Model for DenseNet-201:

The vision transformer model for DenseNet-201 follows a similar architecture to that described previously. The pre-trained DenseNet-201 base is augmented with additional layers for feature extraction and classification. After preprocessing the input image, it is passed through the DenseNet-201 base with pre-trained weights. The top layers of the DenseNet-201 model are excluded to prevent fine-tuning. Following the DenseNet-201 base, a global average pooling layer and dense layers are added for feature extraction. These features are then processed using a multi-head attention mechanism, followed by layer normalization, dropout regularization, and a final dense layer for classification.

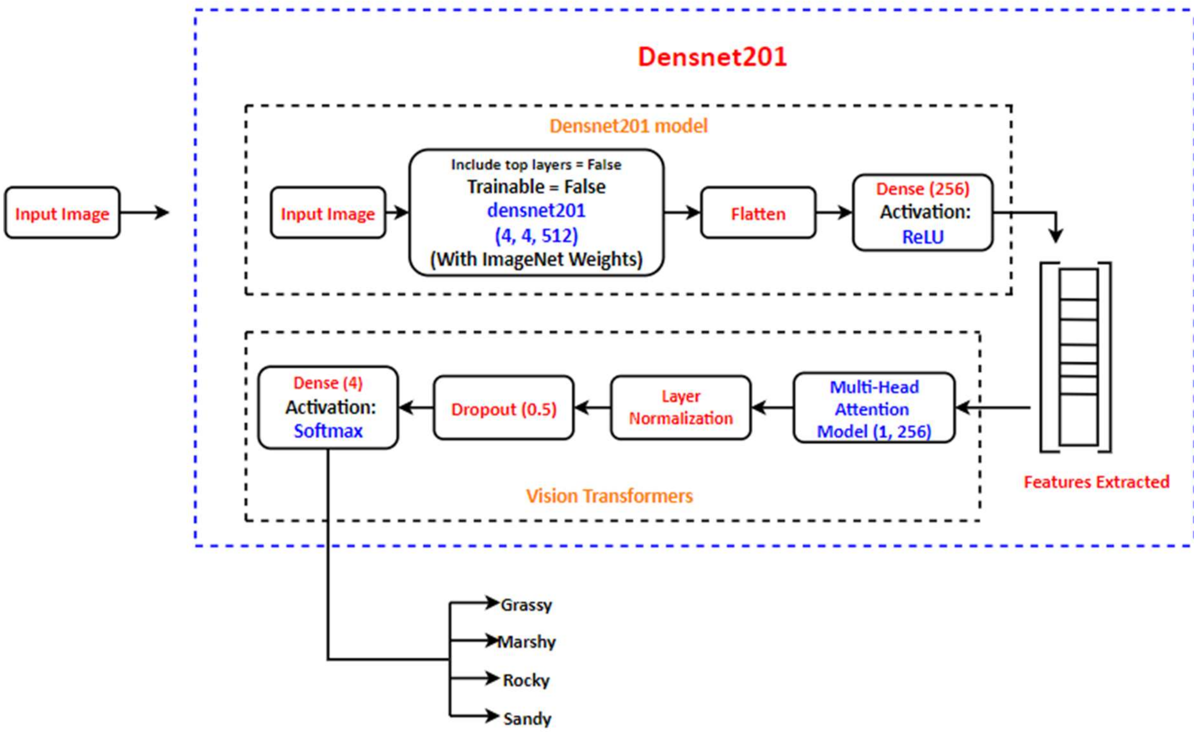

Fig 8: Architecture of DenseNet201 with Vision Transformers

Image Input: Preprocessed and augmented images are fed into the model to ensure dataset diversity.

Transfer Learning: The DenseNet-201 base, with pre-trained weights from ImageNet, is utilized for feature extraction. The top layers are excluded to retain the original features.

Feature Extraction: DenseNet-201 extracts features hierarchically, leveraging densely connected blocks for feature reuse and parameter efficiency.

Multi-Head Attention: Extracted features undergo multi-head attention to capture global dependencies and spatial relations.

Normalization and Dropout: Layer normalization and dropout are applied post-attention for stabilization and regularization.

Classification: Processed features are forwarded to a dense layer with softmax activation for classification.

The flow aligns with the DenseNet-201 model, with the addition of a multi-head attention mechanism post-feature extraction. Extracted features from DenseNet-201 are subjected to multi-head attention to capture global dependencies in the image. The processed features then undergo normalization, dropout, and classification.

DenseNet-121 Model:

DenseNet-121, a variant of the DenseNet architecture comprising 121 layers, shares the core principles of its counterpart, DenseNet-201. It utilizes densely connected blocks, allowing each layer to receive input from all preceding layers, thereby promoting feature reuse and parameter efficiency. Despite its more compact design compared to DenseNet-201, DenseNet-121 maintains effectiveness across a spectrum of computer vision tasks. Its dense connectivity fosters robust feature learning, enabling the model to excel in tasks such as image classification, object detection, and segmentation. This compact yet powerful architecture has become a popular choice in the computer vision community, owing to its balance between performance and computational efficiency.

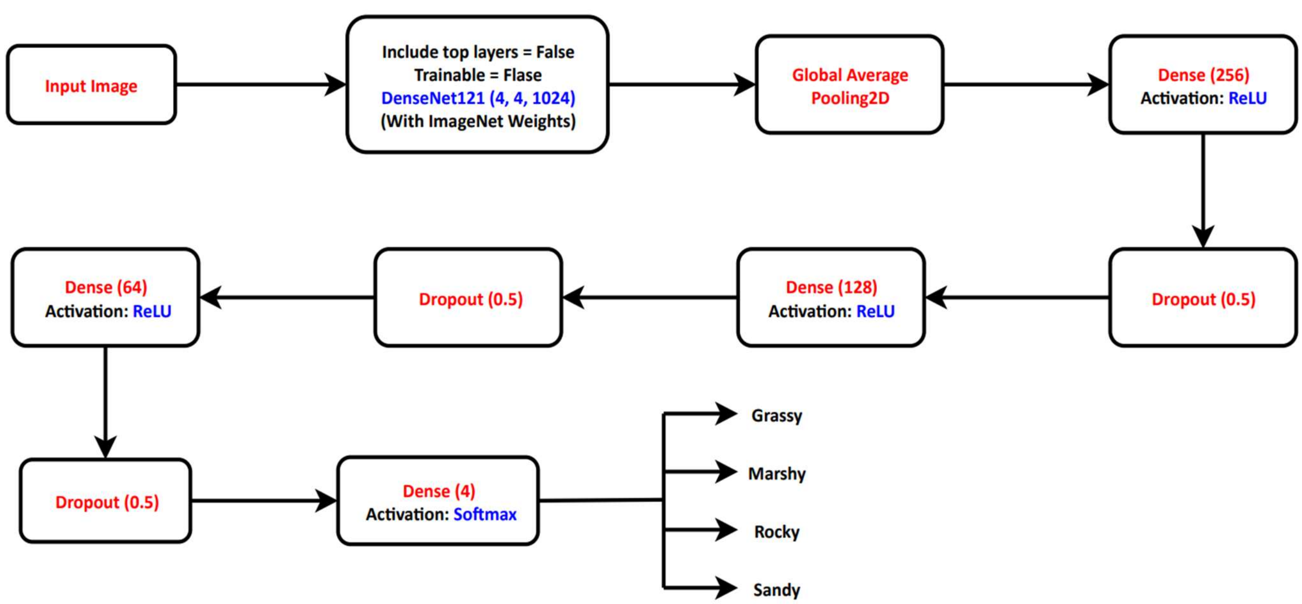

Fig 9: Architecture of DenseNet121

Vision Transformer Model for DenseNet-121:

The vision transformer model for DenseNet-121 follows a similar architecture to that described previously. The pre-trained DenseNet-121 base is augmented with additional layers for feature extraction and classification. After preprocessing the input image, it is passed through the DenseNet-121 base with pre-trained weights. The top layers of the DenseNet-121 model are excluded to prevent fine-tuning. Following the DenseNet-121 base, a global average pooling layer and dense layers are added for feature extraction. These features are then processed using a multi-head attention mechanism, followed by layer normalization, dropout regularization, and a final dense layer for classification.

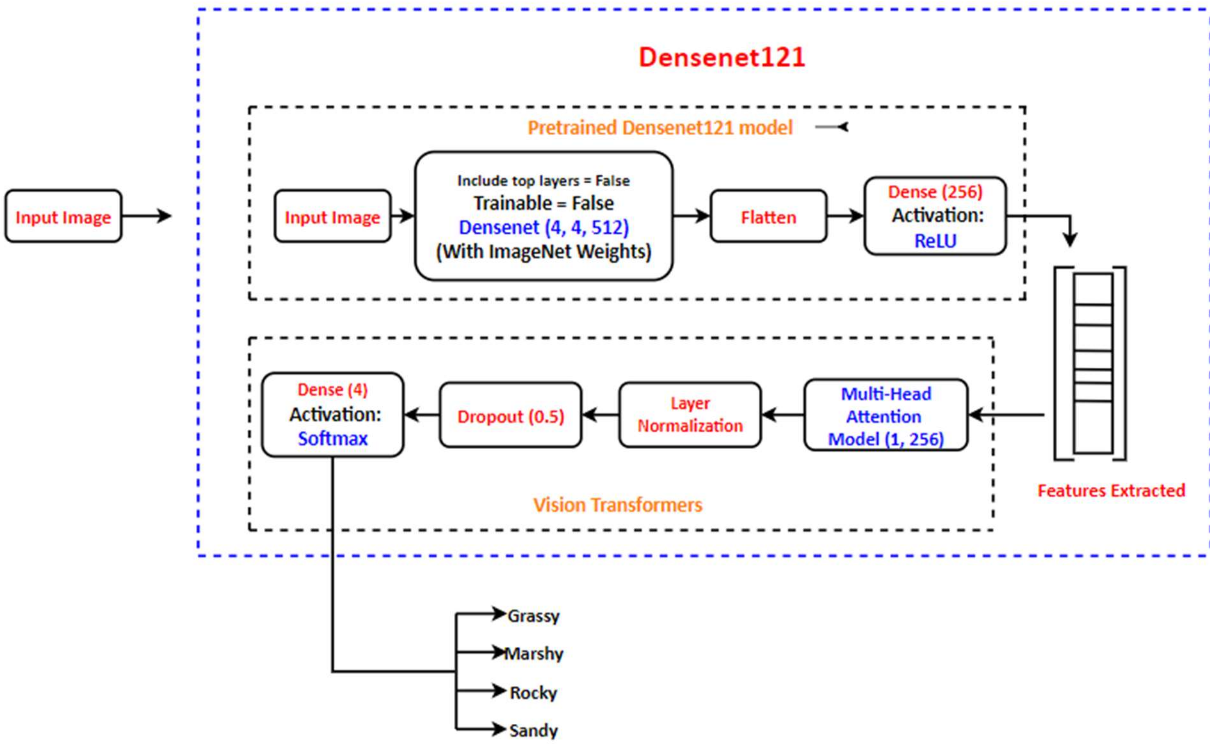

Fig 10: Architecture of DenseNet121 Vision Transformers

Image Input: Input images are preprocessed and augmented to increase the diversity of training data. Transfer Learning: Preprocessed images are passed through the DenseNet-121 base with pre-trained weights, excluding the top layers to retain original features.

Feature Extraction: Features are extracted from the DenseNet-121 base and further processed to capture more abstract representations.

Multi-Head Attention: The extracted features undergo multi-head attention to capture long-range dependencies and spatial relationships.

Normalization and Dropout: Layer normalization is applied to stabilize training, followed by dropout regularization to prevent overfitting.

Classification: Processed features are passed through a dense layer with SoftMax activation for classification into desired classes.

The flow is like the DenseNet-121 model, with the addition of a multi-head attention mechanism for feature processing. After feature extraction from DenseNet-121, features undergo multi-head attention to capture global dependencies in the image. The processed features are then normalized, subjected to dropout, and classified.

MobileNet Model:

MobileNet is a convolutional neural network architecture optimized for deployment on mobile and embedded devices with limited computational resources. It stands out for its innovative use of depth-wise separable convolutions, a technique that reduces computational complexity while preserving performance. This approach involves breaking down standard convolutions into two separate operations: depth-wise convolutions and point-wise convolutions. Depth-wise convolutions independently process each channel of the input data, while point-wise convolutions combine information across channels. By using this strategy, MobileNet achieves a significant reduction in the number of parameters and computations required, resulting in a lightweight model with a low memory footprint. This efficiency makes MobileNet suitable for resource-constrained environments, such as smartphones, IoT devices, and edge computing platforms. Despite its compact size, MobileNet maintains competitive accuracy across various computer vision tasks, including image classification, object detection, and semantic segmentation. Its deployment-friendly design enables real-time inference on mobile devices, facilitating applications such as image recognition, augmented reality, and autonomous navigation. Overall, MobileNet plays a pivotal role in enabling the widespread adoption of deep learning models in mobile applications by providing a balance between computational efficiency and performance.

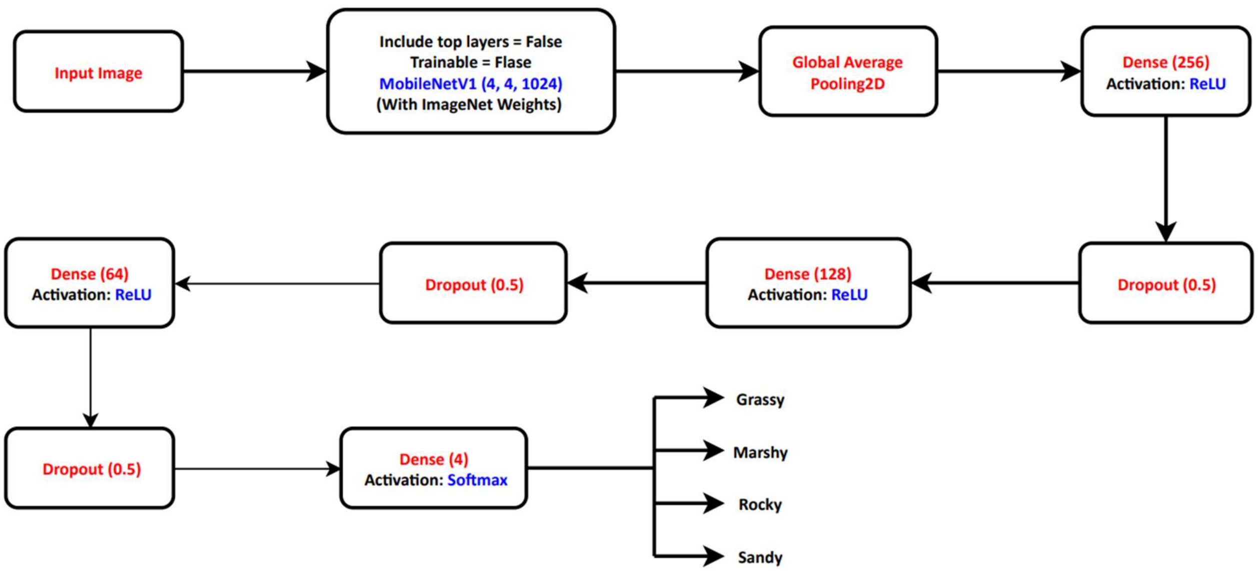

Fig 11: Architecture of MobileNet

Vision Transformer Model for MobileNet:

The vision transformer model for MobileNet follows a similar architecture to that described previously. The pre-trained MobileNet base is augmented with additional layers for feature extraction and classification. After preprocessing the input image, it is passed through the MobileNet base with pre-trained weights. The top layers of the MobileNet model are excluded to prevent fine-tuning. Following the MobileNet base, a global average pooling layer and dense layers are added for feature extraction. These features are then processed using a multi-head attention mechanism, followed by layer normalization, dropout regularization, and a final dense layer for classification.

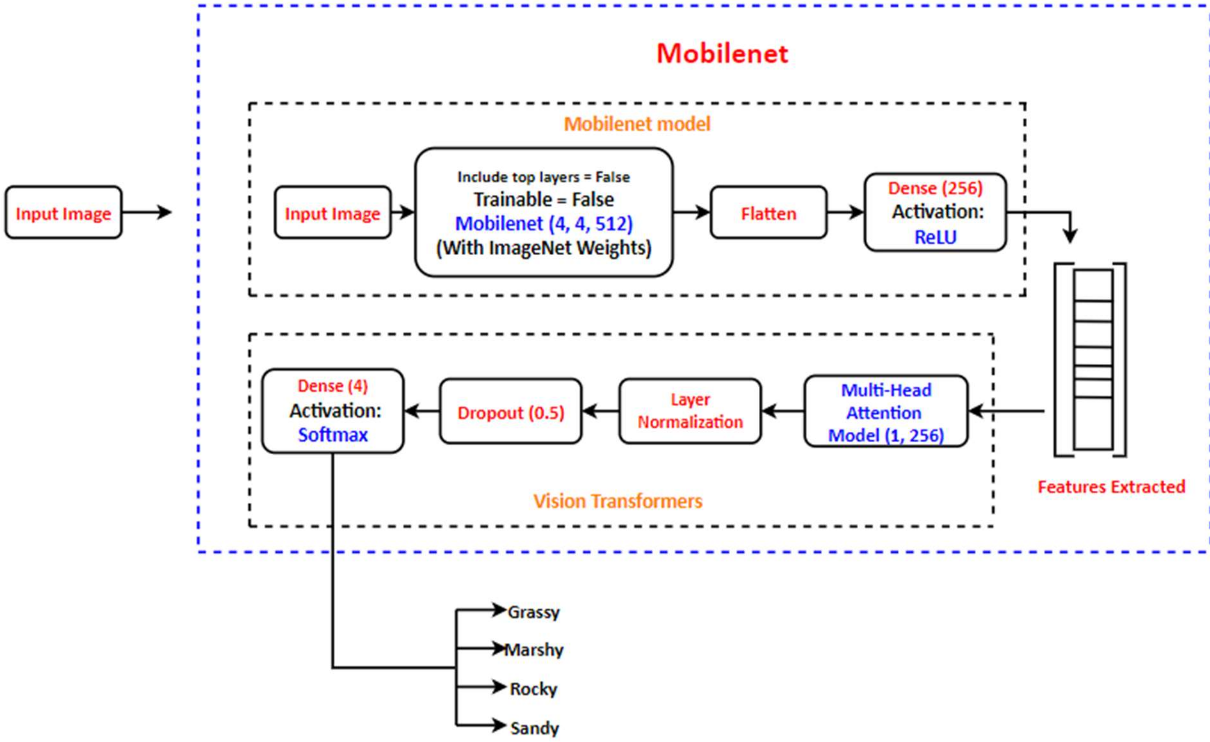

Fig 12: Architecture of MobileNet with Vision Transformers

Image Input: Augmented and preprocessed images are fed into the model to ensure dataset diversity.

Transfer Learning: MobileNetV1, with pre-trained weights from ImageNet, is utilized for feature extraction. Top layers are excluded to retain original features.

Feature Extraction: MobileNetV1 extracts features through depth-wise separable convolutions, reducing computation and parameters while maintaining performance.

Multi-Head Attention: Extracted features undergo multi-head attention to capture global dependencies and spatial relations.

Normalization and Dropout: Layer normalization and dropout are applied post-attention for stabilization and regularization.

Classification: Processed features are forwarded to a dense layer with softmax activation for classification.

The flow aligns with the MobileNetV1 model, with the addition of a multi-head attention mechanism post-feature extraction. Extracted features from MobileNetV1 are subjected to multi-head attention to capture global dependencies in the image. The processed features then undergo normalization, dropout, and classification.

Inception V3 Model:

Inception V3 is a convolutional neural network architecture characterized by its inception modules, which consist of convolutional layers with multiple filter sizes to capture diverse image features. It employs factorized convolutions and parallel branches to enhance feature representation while managing computational complexity.

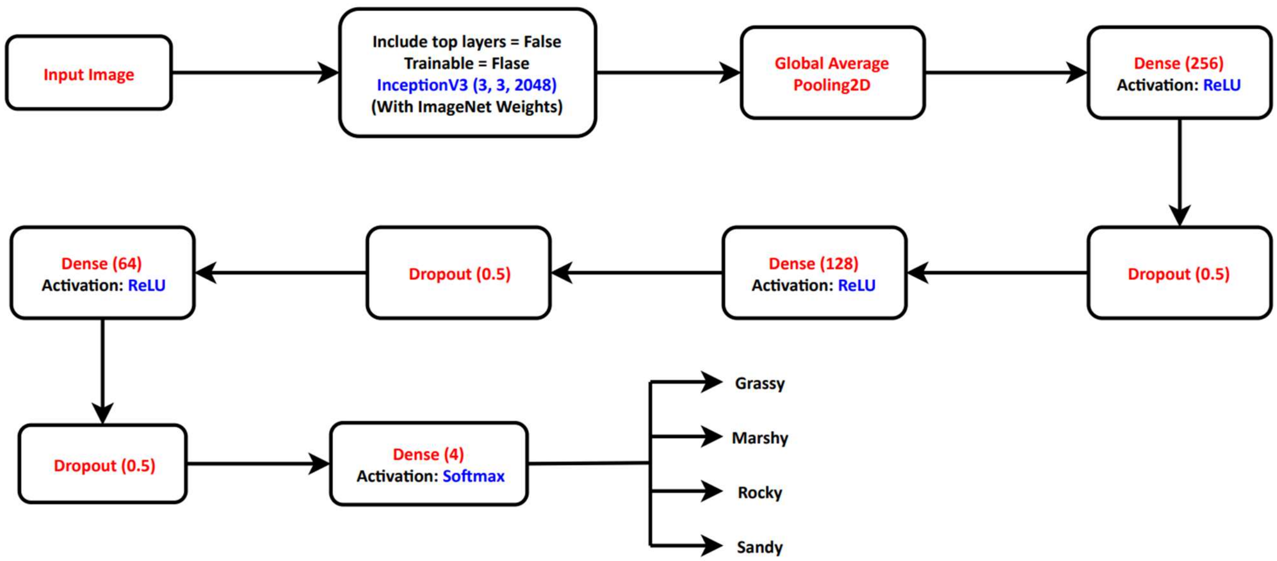

Fig 13: Architecture of InceptionV3

Vision Transformer Model for Inception V3:

The vision transformer model for Inception V3 follows a similar architecture to that described previously. The pre-trained Inception V3 base is augmented with additional layers for feature extraction and classification. After preprocessing the input image, it is passed through the Inception V3 base with pre-trained weights. The top layers of the Inception V3 model are excluded to prevent fine-tuning. Following the Inception V3 base, a global average pooling layer and dense layers are added for feature extraction. These features are then processed using a multi-head attention mechanism, followed by layer normalization, dropout regularization, and a final dense layer for classification.

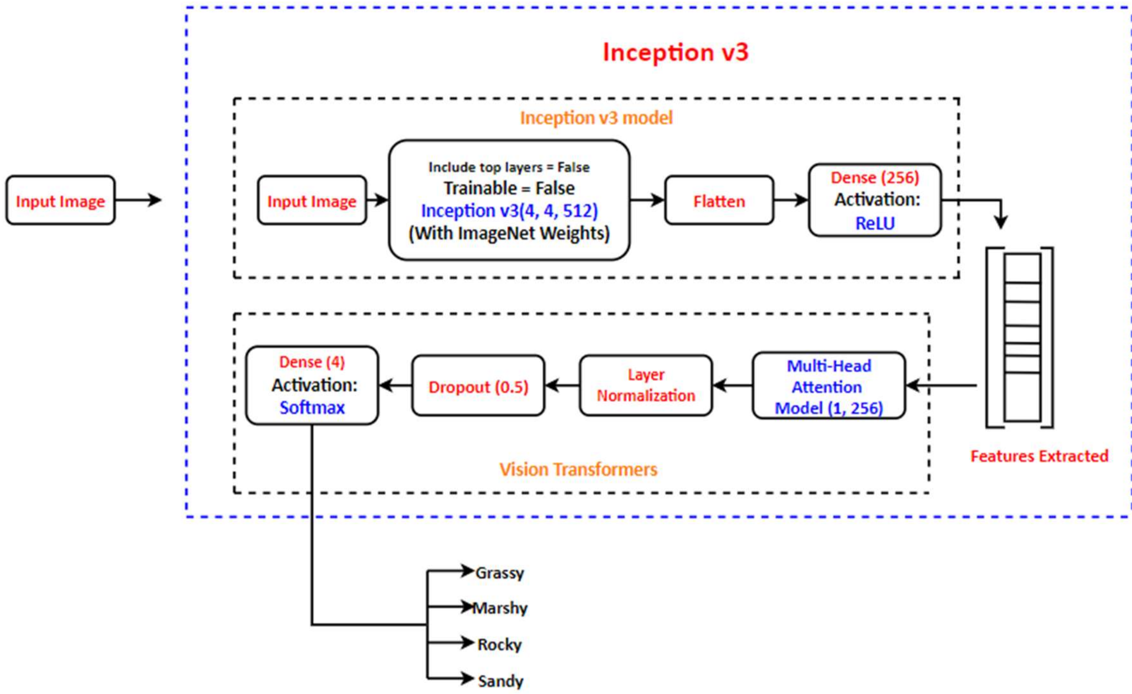

Fig 14: Architecture of InceptionV3 with Vision Transformers

Image Input: Augmented and preprocessed images serve as input, ensuring diversity in the dataset.

Transfer Learning: InceptionV3, pre-trained on ImageNet, is employed for feature extraction. Top layers are excluded to preserve original features.

Feature Extraction: InceptionV3 utilizes inception modules with various filter sizes to capture diverse image features efficiently.

Multi-Head Attention: Extracted features are subjected to multi-head attention to capture global dependencies and spatial relations.

Normalization and Dropout: Following attention, layer normalization and dropout are applied to stabilize and regularize the features.

Classification: Processed features are then passed through a dense layer with SoftMax activation for classification.

Like the InceptionV3 model, the vision transformer incorporates a multi-head attention mechanism post-feature extraction. Extracted features undergo multi-head attention to capture global dependencies in the image. The processed features are subsequently normalized, subjected to dropout, and classified.

VGG16 Model:

VGG16 is a classic convolutional neural network architecture known for its simplicity and effectiveness. It consists of multiple convolutional layers with small 3x3 filters followed by max-pooling layers for spatial down sampling. The network architecture is characterized by its deep stack of convolutional layers, which enables it to learn complex hierarchical features from input images. VGG16 has been widely used in various computer vision tasks, including image classification, object detection, and segmentation.

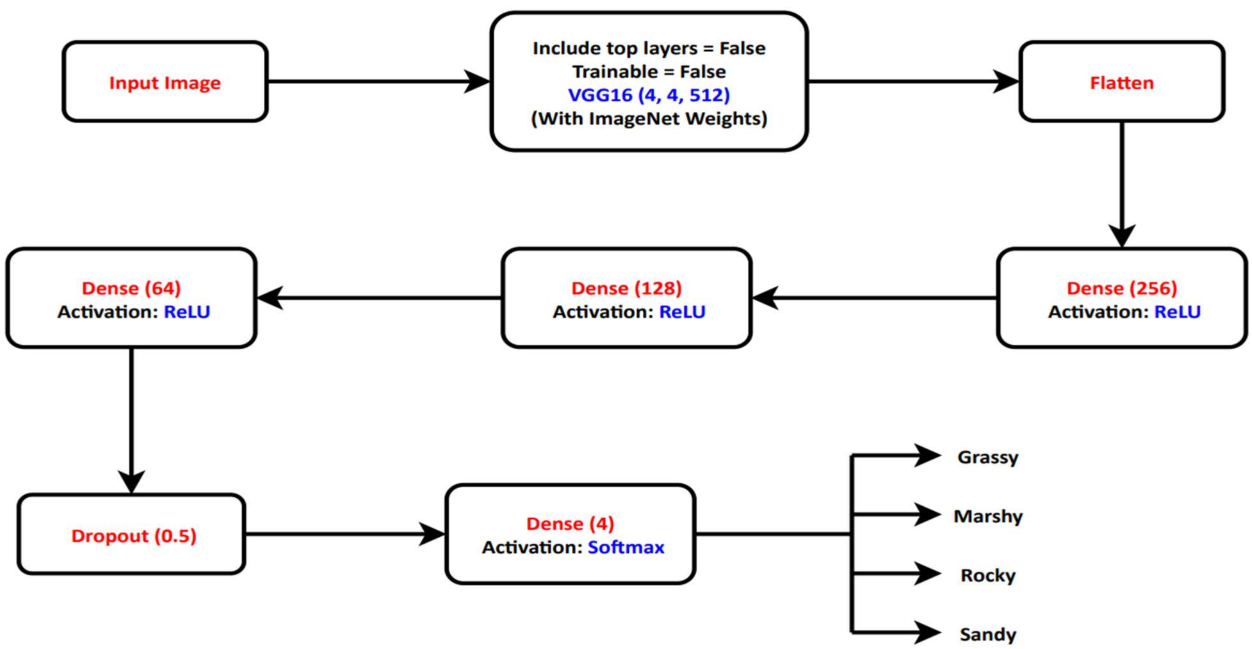

Fig 15: Architecture of Vgg16

Vision Transformer Model for VGG16:

The vision transformer model for VGG16 follows a similar architecture to that described previously. The pre-trained VGG16 base is augmented with additional layers for feature extraction and classification. After preprocessing the input image, it is passed through the VGG16 base with pre-trained weights. The top layers of the VGG16 model are excluded to prevent fine-tuning. Following the VGG16 base, a global average pooling layer and dense layers are added for feature extraction. These features are then processed using a multi-head attention mechanism, followed by layer normalization, dropout regularization, and a final dense layer for classification.

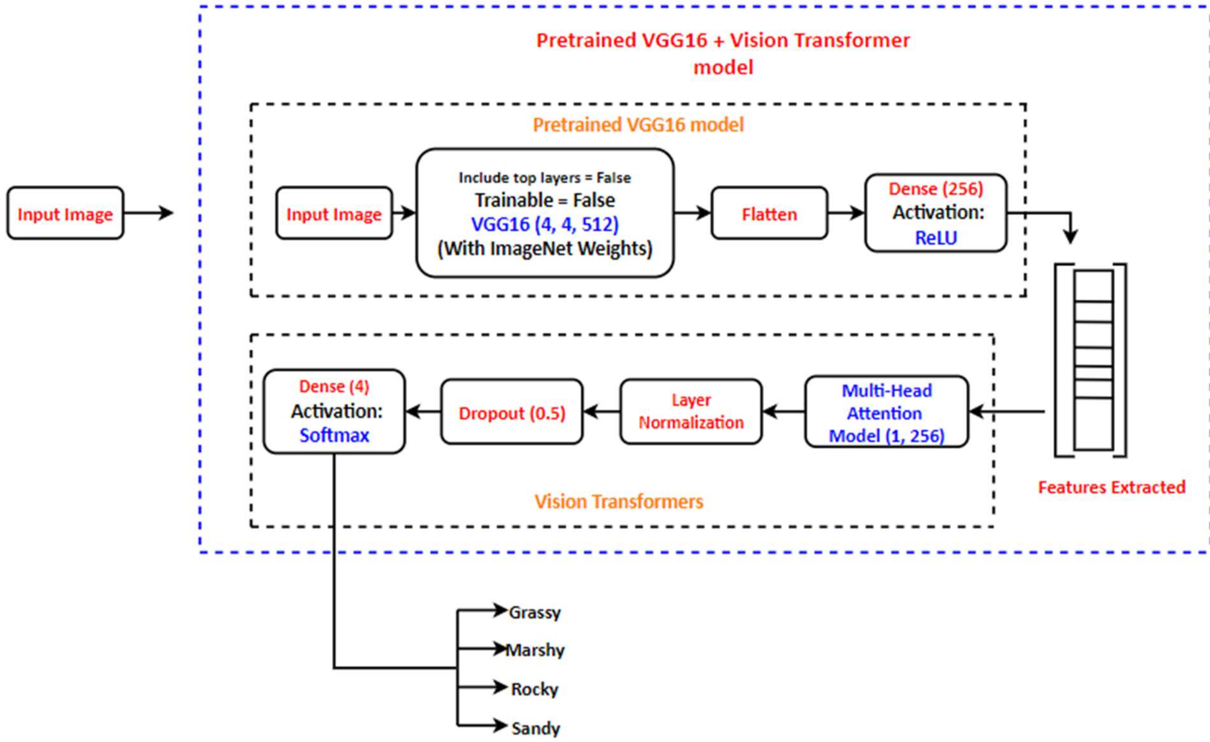

Fig 16: Architecture of Vgg16 with Vision transformers

Image Input: Augmented and pre-processed images serve as input, ensuring diversity in the dataset.

Transfer Learning: VGG16, pre-trained on ImageNet, is utilized for feature extraction. Top layers are excluded to preserve original features.

Feature Extraction: VGG16 consists of multiple convolutional and fully connected layers, capturing hierarchical image features.

Flatten Layer: The output from the convolutional layers is flattened to prepare for classification.

Dense Layers with ReLU Activation: Flattened features are passed through dense layers with ReLU activation, facilitating feature transformation.

Dropout: Dropout regularization is applied to prevent overfitting by randomly dropping units during training.

Classification: The processed features are classified into different classes using a dense layer with SoftMax activation.

Like the VGG16 model, the vision transformer incorporates a multi-head attention mechanism post-feature extraction. Extracted features undergo multi-head attention to capture global dependencies in the image. The processed features are subsequently normalized, subjected to dropout, and classified.

Chapter – 6

RESULTS & DISCUSSION

6. **RESULTS & DISCUSSION**

The image reveals predictions made by the top-performing model, DenseNet201, enhanced with vision transformers, achieving an exceptional accuracy rate of 99.65% on the test dataset. Each row in the image corresponds to a specific test sample, presenting both the true label and the predicted label, classified into four distinct classes: grassy, sandy, marshy, and rocky. This amalgamation of cutting-edge technologies underscores the model's ability to discern intricate features within images, facilitating precise classifications essential for various real-world applications. The transparent presentation of true and predicted labels enhances interpretability, enabling validation of the model's efficacy and identification of potential areas for improvement. With such remarkable accuracy achieved, the model's potential for practical deployment across domains like environmental monitoring, geological surveying, and land classification becomes increasingly promising, signifying a significant advancement in machine learning capabilities for image recognition tasks.

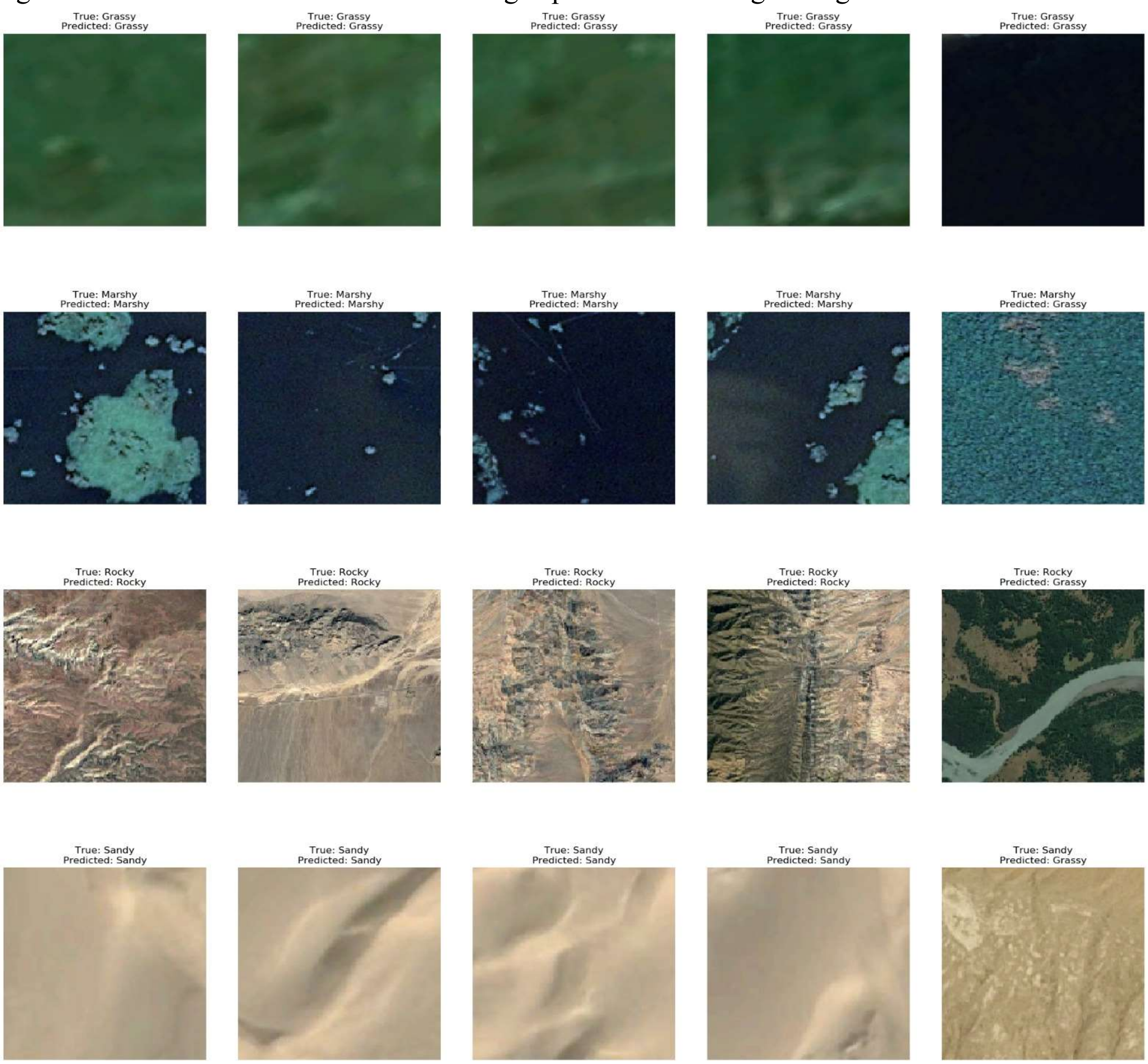

Fig 17: Sample predictions

The table provides a comprehensive analysis of the top seven models, encompassing both transfer learning models and vision transformer models, showcasing their performance in terms of class-wise accuracy, precision, recall, and support. Notably, the DenseNet201 Vision Transformer (ViT) emerged as the top-performing model across all metrics. For marshy terrain, the DenseNet201 ViT achieved an impressive accuracy, precision, and recall of 99.52%, 99%, and 99%, respectively, with a support of 1657 instances. Similarly, for rocky terrain, the model demonstrated a remarkable accuracy of 99.33%, with a precision and recall of 99% each, supported by 1644 instances. In the case of grassy terrain, the DenseNet201 ViT excelled further, achieving a near-perfect precision of 100%, with a corresponding recall and accuracy of 100%, supported by 1824 instances. Moreover, for sandy terrain, the model attained a precision and recall of 99% each, alongside a perfect accuracy of 99.94%, supported by 1644 instances. These exceptional performance metrics underscore the efficacy and reliability of the DenseNet201 ViT in accurately categorizing diverse landscape features. With such high precision, recall, and support across multiple terrain types, the model holds immense potential for applications in environmental monitoring, land cover classification, and military reconnaissance tasks, reaffirming its superiority among the evaluated models.

Model Name	**Class Name**	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**Support**
DENSENET 201 VIT	MARSHY	99.52	0.99	0.99	0.99	1657
	ROCKY	99.33	0.99	0.99	0.99	1644
	GRASSY	99.84	1	1	1	1824
	SANDY	99.94	0.99	1	1	1644
DENSENET 201	MARSHY	99.33	1	0.99	0.99	1657
	ROCKY	99.75	0.93	1	0.96	1644
	GRASSY	99.78	1	1	1	1824
	SANDY	93.24	1	0.93	0.96	1644
DENSENET 121 VIT	MARSHY	95.9	0.96	0.95	0.96	1657
	ROCKY	95.86	0.97	0.97	0.97	1644
	GRASSY	98.38	0.98	0.99	0.98	1824
	SANDY	99.51	0.99	0.99	0.99	1644
DENSENET 121	MARSHY	94.87	0.96	0.95	0.95	1657
	ROCKY	96.35	0.97	0.96	0.97	1644
	GRASSY	98.46	0.98	0.98	0.98	1824
	SANDY	99.51	0.99	1	0.99	1644
MOBILENET VIT	MARSHY	94.87	0.94	0.95	0.95	1657
	ROCKY	94.53	0.97	0.96	0.97	1644
	GRASSY	98.03	0.98	0.97	0.98	1824
	SANDY	99.03	0.99	0.99	0.99	1644
MOBILENET	MARSHY	92.45	0.95	0.92	0.94	1657
	ROCKY	95.74	0.97	0.96	0.96	1644
	GRASSY	98.30	0.96	0.98	0.97	1824
	SANDY	99.20	0.98	0.99	0.99	1644
VGG 16 VIT	MARSHY	94.45	0.94	0.95	0.95	1657
	ROCKY	91.06	0.97	0.96	0.97	1644
	GRASSY	98.19	0.98	0.97	0.98	1824
	SANDY	98.18	0.99	0.99	0.99	1644

Table 1: Class wise performance metrics achieved for the top 7 models

The below table provides a comprehensive overview of the performance metrics for the top 7 models, focusing on the DenseNet201 architecture. Among these models, DenseNet201 exhibited the highest training accuracy of 99.89% with a corresponding training loss of 0.0034. During testing, DenseNet201 maintained exceptional accuracy, achieving 99.656% accuracy with a test loss of 0.009622. Furthermore, its validation accuracy closely mirrored its training and test accuracies, standing at 99.66%, while the validation loss remained low at 0.01. This detailed breakdown underscores the robustness and reliability of the DenseNet201 model across various evaluation stages, making it a standout performer in the study.

MODEL NAME	**EPOCHS**	**TRAINING ACC**	**TRAINING LOSS**	**VALIDATION ACC**	**VALIDATION LOSS**	**TEST ACC**	**TEST LOSS**
DENSENET 201 VIT	20	0.9989	0.0034	0.9966	0.01	0.99656	0.009622
DENSENET 201	20	0.9939	0.022	0.9713	0.1489	0.9807	0.1709
DENSENET 121 VIT	16	0.9889	0.036	0.9733	0.0947	0.97408	0.08092
DENSENET 121	20	0.9812	0.058	0.9713	0.1028	0.9734	0.0927
MOBILENET VIT	13	0.9914	0.027	0.9717	0.1065	0.96638	0.110856
MOBILENET	14	0.9867	0.0429	0.9696	0.1201	0.9647	0.11201
VGG 16 VIT	15	0.9969	0.1719	0.968	0.1719	0.95527	0.18248

Table 2: Performance summary of top 7 models

The below graph illustrates a comparison of accuracy among different models utilized in transfer learning, including both traditional transfer learning models and those augmented with vision transformers. Notably, the highest accuracy of 99.65% was attained by the combination of DenseNet201 with Vision Transformers (ViT). This standout performance underscores the effectiveness of leveraging both DenseNet201 and Vision Transformers in tandem, resulting in superior accuracy compared to other models evaluated in the study. This highlight emphasizes the potential synergy between established transfer learning approaches and emerging techniques like vision transformers, offering promising avenues for advancing performance in various image-related tasks.

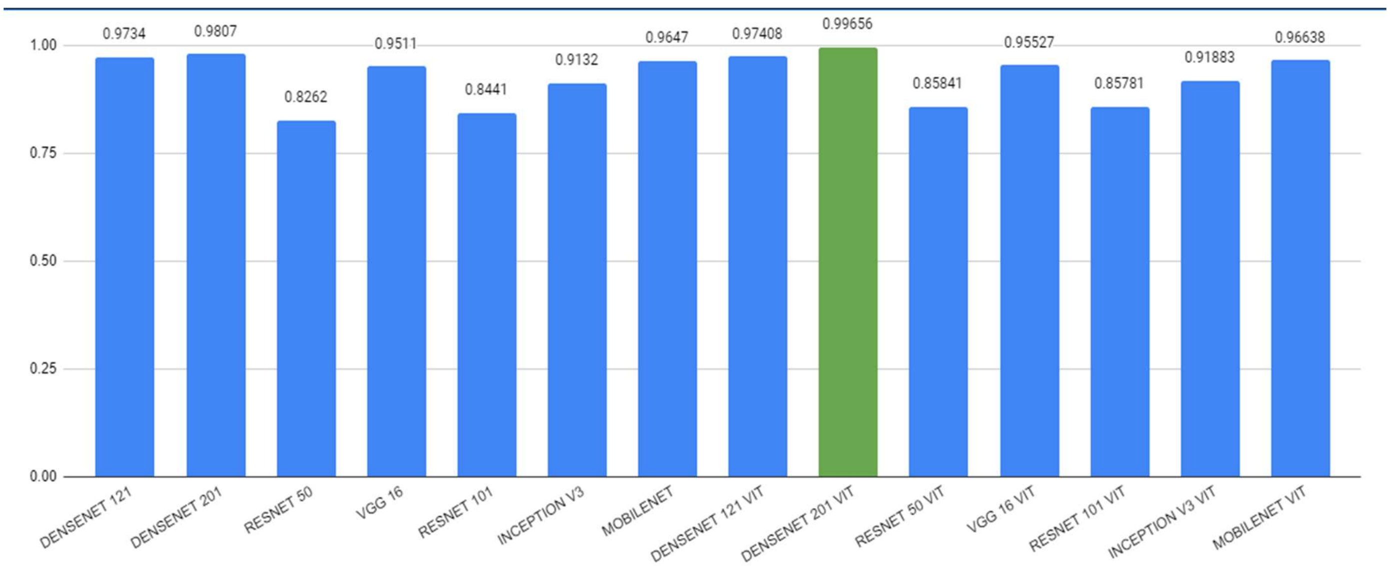

Figure 18: Accuracy comparison of models performed

The below UI images depict a user interface designed for image prediction tasks, featuring a home page displaying three primary buttons: "Choose Image," "Upload," and "Clear Predictions." Upon selecting "Choose Image," users navigate to their local directory to upload an image of their choice. Subsequently, upon clicking "Upload," the system executes the prediction process, generating results for the uploaded image. The subsequent images showcase the prediction outputs for sample images from each label, providing users with insights into the model's classification performance across various categories. This intuitive interface streamlines the prediction workflow, allowing users to seamlessly upload images and obtain accurate predictions with just a few clicks. Overall, the UI design fosters user engagement and facilitates efficient interaction with the image prediction system, enhancing usability and accessibility for users of varying technical backgrounds.

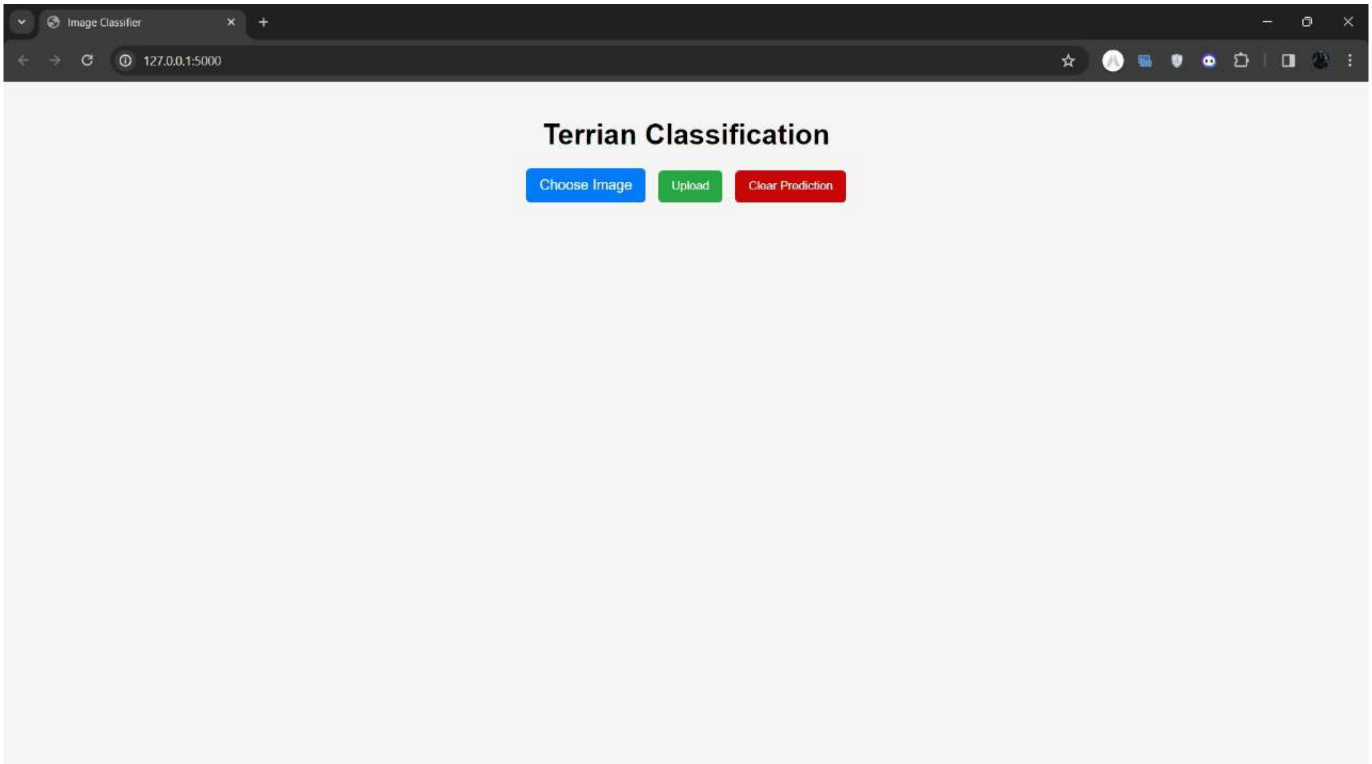

Fig 19:UI HomePage

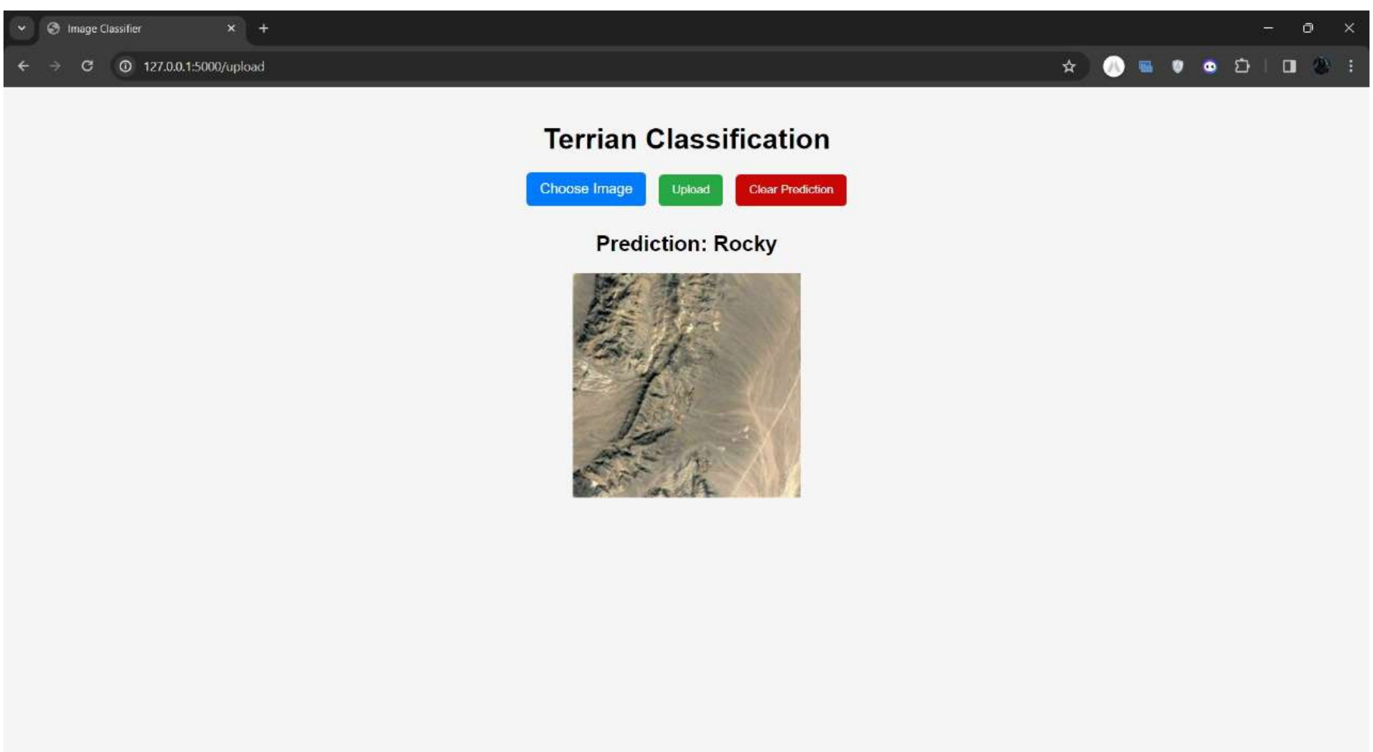

Fig 20:Sample prediction 1

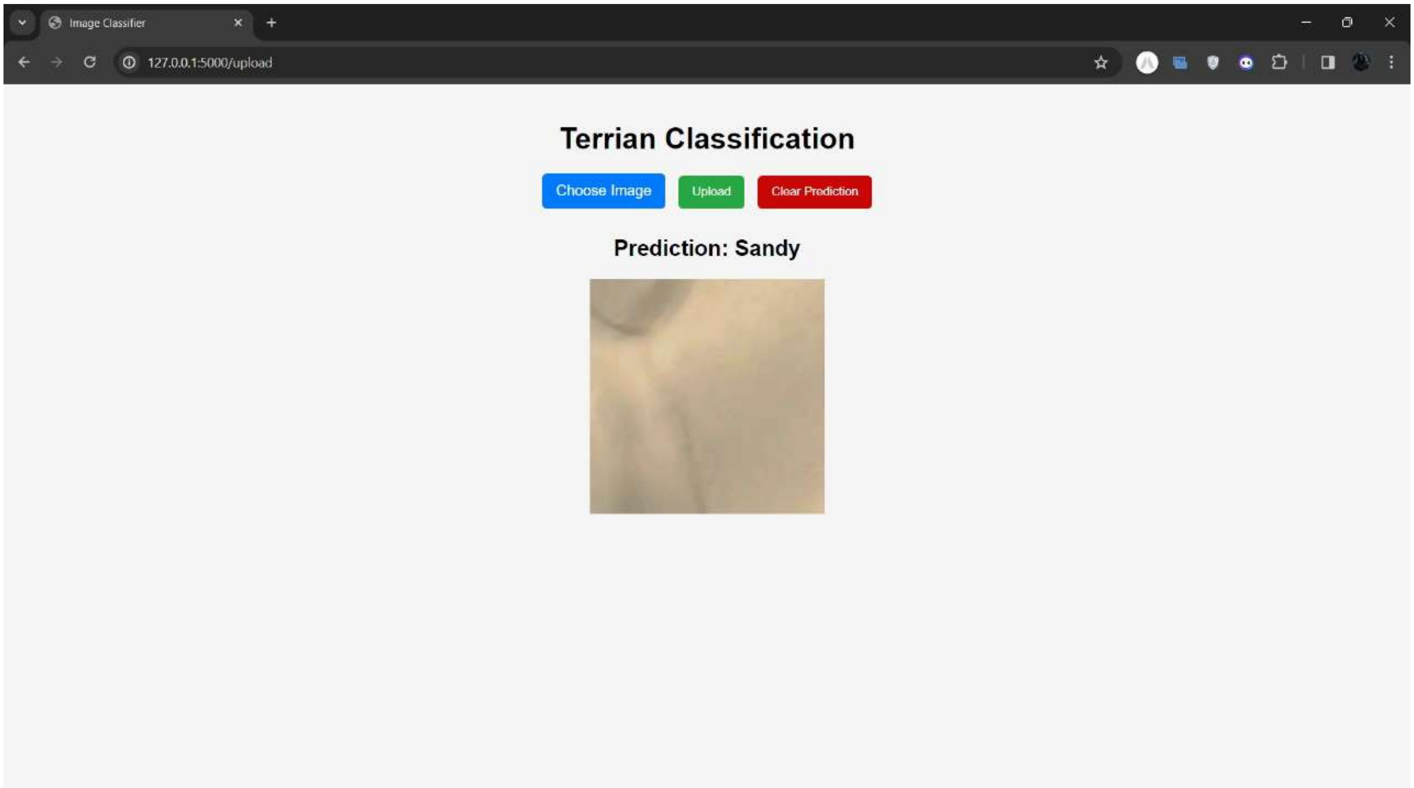

Fig 21:Sample prediction 2

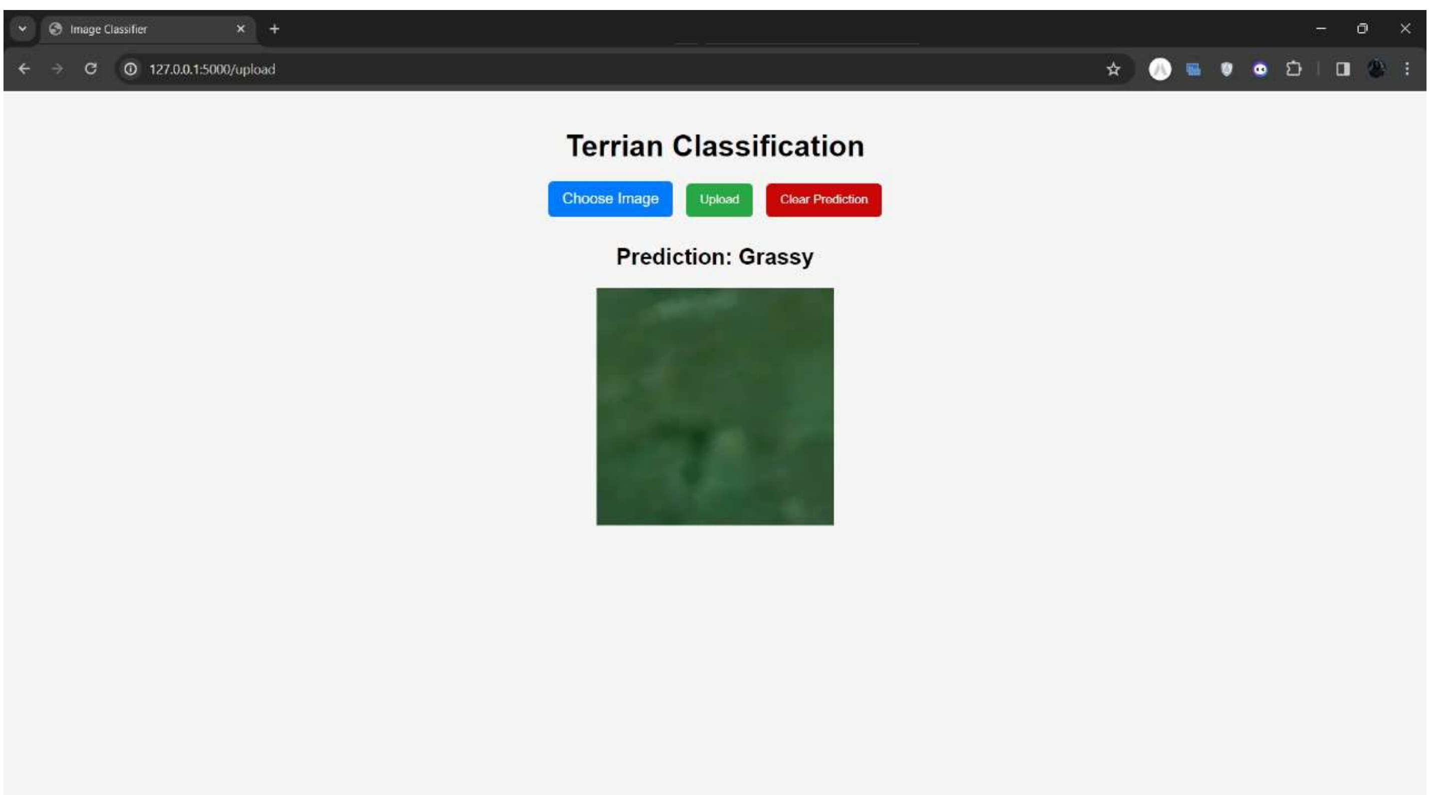

Fig 22:Sample prediction 3

Chapter – 7

CONCLUSION

7. CONCLUSION

The terrain classification project utilizing transfer learning techniques and vision transformers has yielded promising results, achieving a remarkable accuracy of 99% on unseen test data. The project addressed a problem statement provided by SIH 2023, presented by DRDO, and utilized a dataset available on Kaggle, comprising four terrain classes: grassy, marshy, sandy, and rocky.

Through the implementation of transfer learning, the project leveraged pre-trained models to efficiently learn relevant features from the dataset, despite its limited size. This approach enabled the development of robust terrain classification models capable of accurately distinguishing between different terrain types, essential for applications in robotics, autonomous vehicles, and military operations.

Vision transformers augmented the project's methodology by introducing self-attention mechanisms, enhancing the model's ability to capture long-range dependencies in terrain imagery. By combining transfer learning with vision transformers, the project demonstrated a synergistic approach that leverages the strengths of both techniques to achieve superior performance in terrain classification tasks.

Looking ahead, the project lays the groundwork for future research and development in terrain classification. The success achieved with the limited data set suggests potential for further improvement and expansion. Future work could involve training the models on larger datasets encompassing a broader range of terrain classes and landforms, thereby enhancing the models' ability to generalize to diverse environments.

Additionally, the insights gained from this project can inform the design of more sophisticated terrain classification systems, capable of addressing real-world challenges in navigation, reconnaissance, and decision-making. By continuing to refine and optimize the models, researchers can contribute to advancements in autonomous systems, military operations, and various other fields reliant on accurate terrain analysis.

In summary, the terrain classification project represents a significant step forward in leveraging deep learning techniques for understanding and interpreting complex terrain landscapes. With its high accuracy and potential for expansion, the project sets a solid foundation for future endeavors aimed at enhancing our capabilities in terrain classification and related applications.

.

REFERENCES:

[1] Yang, J., Xu, J., Lv, Y., Zhou, C., Zhu, Y., & Cheng, W. (2023). Deeplearning-based automated terrain classification using high-resolution DEM data. International Journal of Applied Earth Observation and Geoinformation, 118, 103249.

[2] Bai, C., Guo, J., & Zheng, H. (2019). Three-dimensional vibration-based terrain classification for mobile robots. IEEE Access, 7, 63485-63492.

[3] Shi, W., Li, Z., Lv, W., Wu, Y., Chang, J., & Li, X. (2020). Laplacian support vector machine for vibration-based robotic terrain classification. Electronics, 9(3), 513.

[4] Christie, J., & Kottege, N. (2016, May). Acoustics based terrain classification for legged robots. In 2016 IEEE international conference on robotics and automation (ICRA) (pp. 3596-3603). IEEE.

[5] Pingel, T. J., Clarke, K. C., & McBride, W. A. (2013). An improved simple morphological filter for the terrain classification of airborne LIDAR data. ISPRS journal of photogrammetry and remote sensing, 77, 21-30.

[6] Li, S., Xiong, L., Tang, G., & Strobl, J. (2020). Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. Geomorphology, 354, 107045.

[7] Zhang, L., Chen, Z., Zou, B., & Gao, Y. (2018, July). Polarimetric SAR terrain classification using 3D convolutional neural network. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 4551-4554). IEEE.

[8] Yu, Z. (2020, September). Research on Remote Sensing Image Terrain Classification Algorithm Based on Improved KNN. In 2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 569-573). IEEE.

[9] Zürn, J., Burgard, W., & Valada, A. (2020). Self-supervised visual terrain classification from unsupervised acoustic feature learning. IEEE Transactions on Robotics, 37(2), 466-481.

[10]Abete, J. M., Abooee, A., Acharya, P., Afzal, M. R., Agnus, J., Ahlers, D.,... & Cai, C. J. (2021). 2021 Index IEEE/ASME Transactions on Mechatronics Vol. 26. IEEE/ASME TRANSACTIONS ON MECHATRONICS, 26(6).

[11] Abete, J. M., Abooee, A., Acharya, P., Afzal, M. R., Agnus, J., Ahlers, D., ... & Cai, C. J. (2021). 2021 Index IEEE/ASME Transactions on Mechatronics Vol. 26. IEEE/ASME TRANSACTIONS ON MECHATRONICS, 26(6).

[12] Manduchi, R., Castano, A., Talukder, A., & Matthies, L. (2015). Obstacle detection and terrain classification for autonomous off-road navigation. Autonomous robots, 18, 81-102.

[13] Khaleghian, S., & Taheri, S. (2017). Terrain classification using intelligent tire. Journal of Terramechanics, 71, 15-24.

[14] Vulpi, F., Milella, A., Marani, R., & Reina, G. (2021). Recurrent and convolutional neural networks for deep terrain classification by autonomous robots. Journal of Terramechanics, 96, 119-131.

[15] Shaban, A., Meng, X., Lee, J., Boots, B., & Fox, D. (2022, January). Semantic terrain classification for off-road autonomous driving. In Conference on Robot Learning (pp. 619-629). PMLR.

[16] Ahmadi, A., Nygaard, T., Kottege, N., Howard, D., & Hudson, N. (2021). Semi-supervised gated recurrent neural networks for robotic terrain classification. IEEE Robotics and Automation Letters, 6(2), 1848-1855.

[17] Li, W., & Hsu, C. Y. (2020). Automated terrain feature identification from remote sensing imagery: a deep learning approach. International Journal of Geographical Information Science, 34(4), 637-660.

[18] Anantrasirichai, N., Burn, J., & Bull, D. (2014). Terrain classification from body-mounted cameras during human locomotion. IEEE transactions on cybernetics, 45(10), 2249-2260.

[19] Grezmak, J., Graf, N., Behr, A., & Daltorio, K. (2021). Terrain classification based on sensed leg compliance for amphibious crab robot. IEEE Sensors Journal, 21(20), 23308-23316.

[20] Li, Z., Li, Y., Rong, X., & Zhang, H. (2020). Grid map construction and terrain prediction for quadruped robot based on c-terrain path. IEEE Access, 8, 56572-56580.

[21] Li, L., Wang, X., Meng, Q., & Yu, H. (2023). A Wearable Computer Vision System With Gimbal Enables Position-, Speed-, and Phase-Independent Terrain Classification for Lower Limb Prostheses. IEEE transactions on neural systems and rehabilitation engineering, 31, 4539-4548

[22] Wu, X. A., Huh, T. M., Mukherjee, R., & Cutkosky, M. (2016). Integrated ground reaction force sensing and terrain classification for small legged robots. IEEE Robotics and Automation Letters, 1(2), 1125-1132.

[23] Zhao, Y., Gao, F., Sun, Q., & Yin, Y. (2021). Terrain classification and adaptive locomotion for a hexapod robot Qingzhui. Frontiers of Mechanical Engineering, 16(2), 271-284..

[24] Na, J., Ding, H., Zhao, W., Liu, K., Tang, G., & Pfeifer, N. (2021). Object-based large-scale terrain classification combined with segmentation optimization and terrain features: A case study in China. Transactions in GIS, 25(6), 2939-2962.

[25] Iwahashi, J., Kamiya, I., Matsuoka, M., & Yamazaki, D. (2018). Global terrain classification using 280 m DEMs: segmentation, clustering, and reclassification. Progress in Earth and Planetary Science, 5, 1-31.
[26] Shi, J., Jin, H., & Li, X. (2020). A novel multi-feature joint learning method for fast polarimetric SAR terrain classification. IEEE Access, 8, 30491-30503.
[27] Lv, W., Kang, Y., Zheng, W. X., Wu, Y., & Li, Z. (2020). Feature-temporal semi-supervised extreme learning machine for robotic terrain classification. IEEE Transactions on Circuits and Systems II: Express Briefs, 67(12), 3567-3571..
[28] Otsu, K., Ono, M., Fuchs, T. J., Baldwin, I., & Kubota, T. (2016). Autonomous terrain classification with co-and self-training approach. IEEE Robotics and Automation Letters, 1(2), 814-819.
[29] Dai, W., Na, J., Huang, N., Hu, G., Yang, X., Tang, G., ... & Li, F. (2020). Integrated edge detection and terrain analysis for agricultural terrace delineation from remote sensing images. International Journal of Geographical Information Science, 34(3), 484-503.
[30] Barrett, A. M., Wright, J., Favaro, E., Fawdon, P., Balme, M. R., Woods, M. J., ... & Joudrier, L. (2023). Oxia Planum, Mars, classified using the NOAH-H deep-learning terrain classification system. Journal of Maps, 19(1), 2112777.
[31] Zou, X., Hwu, T., Krichmar, J., & Neftci, E. (2020, October). Terrain classification with a reservoir-based network of spiking neurons. In 2020 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 1-5). IEEE.
[32] Karimzadeh, S., Feizizadeh, B., & Matsuoka, M. (2019). DEM-based Vs30 map and terrain surface classification in nationwide scale—a case study in Iran. ISPRS International Journal of Geo-Information, 8(12), 537.
[33] Lunsaeter, S. F., Iwashita, Y., Stoica, A., & Torresen, J. (2020, October). Terrain classification from an aerial perspective. In 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 173-177). IEEE.
[34] Ai, J., Wang, F., Mao, Y., Luo, Q., Yao, B., Yan, H., ... & Wu, Y. (2021). A fine PolSAR terrain classification algorithm using the texture feature fusion-based improved convolutional autoencoder. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-14.

[35] Liu, H., Yang, S., Gou, S., Liu, S., & Jiao, L. (2018). Terrain classification based on spatial multi-attribute graph using polarimetric SAR data. Applied Soft Computing, 68, 24-38.
[36] Wellhausen, L., Dosovitskiy, A., Ranftl, R., Walas, K., Cadena, C., & Hutter, M. (2019). Where should i walk? predicting terrain properties from images via self-supervised learning. IEEE Robotics and Automation Letters, 4(2), 1509-1516.
[37] Liu, H., Shang, F., Yang, S., Gong, M., Zhu, T., & Jiao, L. (2019). Sparse manifold-regularized neural networks for polarimetric SAR terrain classification. IEEE Transactions on Neural Networks and Learning Systems, 31(8), 3007-3016.
[38] Zhang, L., Chen, Z., Zou, B., & Gao, Y. (2018, July). Polarimetric SAR terrain classification using 3D convolutional neural network. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 4551-4554). IEEE.
[39] Cui, X., Xing, Z., Yang, F., Fan, M., Ma, Y., & Sun, Y. (2020). A method for multibeam seafloor terrain classification based on self-adaptive geographic classification unit. Applied Acoustics, 157, 107029.
[40] Wang, W., Zhang, B., Wu, K., Chepinskiy, S. A., Zhilenkov, A. A., Chernyi, S., & Krasnov, A. Y. (2022). A visual terrain classification method for mobile robots' navigation based on convolutional neural network and support vector machine. Transactions of the Institute of Measurement and Control, 44(4), 744-753.
[41] Nampoothiri, M. H., Vinayakumar, B., Sunny, Y., & Antony, R. (2021). Recent developments in terrain identification, classification, parameter estimation for the navigation of autonomous robots. SN Applied Sciences, 3, 1-14.
[42] Yang, C., Hou, B., Ren, B., Hu, Y., & Jiao, L. (2019). CNN-based polarimetric decomposition feature selection for PolSAR image classification. IEEE Transactions on Geoscience and Remote Sensing, 57(11), 8796-8812.
[43] Yin, Q., Hong, W., Zhang, F., & Pottier, E. (2019). Optimal combination of polarimetric features for vegetation classification in PolSAR image. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(10), 3919-3931.
[44] Pokonieczny, K. (2018). Methods of using self-organising maps for terrain classification, using an example of developing a military passability map. In Dynamics in GIscience 4 (pp. 359-371). Springer International Publishing.

[45] Zeng, R., Zhong, B., Liu, W., Liu, S., Guo, Q., & Kang, Y. (2023, June). A Heterogeneous Ensemble Approach for Terrain Classification with Tracked Mobile Robots. In Journal of Physics: Conference Series (Vol. 2478, No. 10, p. 102001). IOP Publishing.