National College of Ireland

Enhancing the Classification and Identification of Natural Rocks using Swin-Transformer Architecture

MSc Research Project
MSc. in Data Analytics

Subhashree Bera
Student ID: x20241062

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Subhashree Bera
Student ID:	x20241062
Programme:	MSc in Data Analytics **Year:** 2022
Module:	MSc Research Project
Supervisor:	Jorge Basilio
Submission Due Date:	17/09/2022
Project Title:	Enhancing the Classification and Identification of Natural Rocks using Swin-Transformer Architecture
Word Count:	7044 **Page Count** 20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Subhashree Bera

Date: 17/09/2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing the Classification and Identification of Natural Rocks using Swin-Transformer Architecture

Subhashree Bera

x20241062

Abstract

The identification and categorization of rock lithology are crucial topics in the geological survey. The reliability of the classification cannot be assured since the identification technique is based on rock-thin layers and seems to have a long identification duration and high cost. Additionally, the aforementioned approach is unable to offer a practical answer. The majority of geological survey employees carry devices, which are transceivers with several detectors. During the extraction of the rocks, natural calamities are always a risk therefore, in this study deep learning-based methods are discussed which can recognize the rocks by images and these images can be collected through drones. This method also ensures reliability and consumes less time in comparison to the traditional methods. Deep learning is succeeding in recognition and classification, which is simplifying the laborious process of categorization and recognizing images. This field of study is still fully unexplored. Four distinct deep learning models are used in this task which are VGG-19, Inception V3, Custom Model, and Swin Transformers. Important results have been discovered by employing the state-of-the-art design of deep learning models. Here, a bespoke model is created using convolutional blocks, and a swin transformer is a cutting-edge model. Vgg-19 and Inception V3 are based on a transfer learning technique. All models are evaluated on test data and assessed using various metrics after being trained on the rock's image data. The categorization of images of rocks into several groups in real-world tasks may be accomplished by using the swin transformers model, which has proven to be superior to other models after assessment of the model's using metrics like Accuracy, validation loss, Precision, and Recall.

Keywords - Swin-Transformer, Rock classification, VGG-19, Geographical, Inception-V3, Deep learning algorithms

1 Introduction

The main building block of Earth is rocks. They are therefore essential to almost all undertakings of an evolved society since for all manufacture and modern construction raw materials are provided by them. Mining, excavation, and drilling provide the material source for polymers, metals, and fuels for the use of rocks. Different rock kinds have

different origins and purposes. In accordance with different traits, three substantial categories of Rocks— igneous, metamorphic and sedimentary are also more segmented on different categories. Identification of the different types of rocks is an important step in geological exploration and the hunt for resources of minerals. Numerous methods exist for identifying rocks, including chemical analysis and visually analyzing under a microscope. Working conditions within the field for the most part limit identification to visual strategies, counting employing an amplifying spectacle for finely grained rocks. The visual analysis evaluates properties including composition, structure, tint, and grain's magnitude. Rocks' properties are a reflection of their origin, setting, and mineral and chemical composition. Rocks' chemical makeup is reflected in their color. The rapid advancement of image capturing and digital image pattern identification technologies have enabled the modifications in automatic systems to identify rock type from field photographs. By increasing identification efficiency and accuracy, these systems will be of considerable use to geologists. They will also allow students and newly certified geologists to practice identifying different types of rocks. The geological mapping systems and automatic remote sensing used by Unmanned Aerial Vehicles (UAVs) can contain identification systems.

The largest data science community in the world, Kaggle provides us with powerful resources and tools to help us accomplish our data science goals. This dataset was created by the Brac University Mars Rover Team with the intention of identifying comparable rocks on a surface resembling Mars. The task of identifying minerals in the field is one that is fraught with many difficulties. Traditional methods are error-prone and demand skill. Deep learning techniques can assist in overcoming some of these obstacles to offer straightforward and efficient methods of mineral identification. However, current methods tend to favor a manual feature extraction pipeline and mostly use characteristics of the minerals as seen under a microscope. Datasets for igneous metamorphic sedimentary rocks and minerals are also available.

Deep learning, commonly referred to as deep neural networks, has drawn interest recently in a number of scientific domains. Numerous deep learning techniques have been put forth. Deep Convolutional Neural Networks (CNNs) can improve classification efficiency and accuracy without the use of artificial feature selection by automatically learning the characteristics needed for identifying images from the data of the image. Deep learning algorithms have been presented in recent studies to produce significant improvements in a variety of like image speech recognition, classification, clinical diagnosis, object detection, identification of diseases of plants, and recognition of traffic signals and humans.

In this research, the four different models of deep learning will be deployed for accurate classification. Among these 4 models, two pre-trained models will be utilized which are VGG-19 and Inception-V3. Other than that custom model, with very limited number of Convolutional layers has been developed from scratch and also another transformer-based model, Swin Transformer will be utilized for this research. The efficiency of each model could be measured with respect to Validation Accuracy, Validation loss, Precision and Recall score over the test data.

1.1 Research Question

RQ1: To what extent the Advance Deep neural architecture such as Swin-Transformer can accurately classify the rock type based on the images?

RQ2: Which Deep neural Network architecture is the most optimal model for rock classification among the Inception-V3, VGG-19, Custom CNN Model and Swin-Transformer?

1.2 Research Objectives

The objectives of this project are as follow:

- A critical review of Literature on Rock classification
- Execution of Data Pre-processing and Image augmentation along with SWIN-Transformer model.
- Implement and evaluate the outcomes of Inception-V3, VGG-19, Custom-CNN and Swin-Transformer.
- Correctly Identification of different classes of Rock (Igneous, Metamorphic, Minerals and Sedimentary).
- Compare and identify the most optimal model for Rock classification.

2 Related Work

An overview of the prior research has been presented here on rock classification. The structure of the related work will be as follow. (i). Rock classification using advance deep learning algorithms, (ii). Critical analysis of Transfer learning techniques for Rock classification, (iii). Analysis of Machine learning based algorithm for rock classification.

2.1 Rock Classification Using Advanced Deep Learning Algorithms

Tong et al. (2017) pointed out that while deep neural networks produce results with large range of Accuracy, the computational expense and duration required to execute that make them challenging to use in quick on-site examinations. The study is a proposal that aims to enhance the application of rock spectral imaging in order to address these drawbacks. Neighborhood Component Analysis (NCA) will be used to accomplish Dimensionality Reduction (DR), transforming hyperspectral imagery into multispectral imaging. Goldberger et al. (2004) proposed the spectral characteristics of rocks, environment, and minerals could be done afterward without the requirement for needless data processing by linking DR of hyperspectral data through NCA to multispectral photography, in addition to Machine Learning (ML).

Li et.al (2021) presented the practicality of the hyperspectral data DR using the NCA algorithm, ranging from 204 - 100, 50, 25, and 10, bands. The standard five bands are considered multispectral by the industry. On the basis of NCA, the researchers concluded the most practical multispectral bands for discrimination having the following wavelengths- 741 nm, 897nm, 441 nm, 791 nm, and 535nm. These 5 multispectral bands are most feasible in the igneous rock's classification such as granodiorite, diorite, rhyolite, dacite, granite, basalt, gabbro, and andesite. The data plots 2D with DR help such as frequency density histograms and band-against-band scatter plots offer better data interpretation, visualization, and some data prediction capabilities. As a result, it can be claimed that DR can be a valuable strategy for many datasets that suffer from the dimensionality curse because it eliminates redundant features. The suggested approach seamlessly combines with various ML models. Machine learning models such as the Support Vector Machine (SVM) model in classifying igneous rock database has rock performed better than all other models, is one example of how the authors have given quantifiable outputs relevant to the classification capabilities of each ML model.

In Sharma et.al (2021), the authors proposed the fact that NCA gives the option to "discard redundant data-heavy bands", but it does not contribute much evidence on the, "retained classification accuracy", therefore an ML approach is required (s). The study employed ML for a wide range of "data-related jobs" and issues. It has developed into a branch of "Artificial Intelligence (AI)" which consists of prototypes that can extract relevant evidence from data and utilize knowledge to support individual learning that facilitates accurate categorization or prediction. Due to ML's precision and dependability, it has rapidly grown in popularity. Machine learning algorithms that analyze data more quickly and provide trustworthy conclusions in a short amount of time have been made possible by improved components of machine systems.

By claiming that an NCA model demonstrates a direct change that raises the adequacy of k-NN, Koren Carmel (2004) also promote the use of the "NCA model". These aforementioned benefits informed the decision to use NCA to separate the rocks from the database of hyperspectral rock images. The recent researchers point out that another important distinction between these 2 models is that the one, non-convex optimization issue is NCA and the standard Principal Component Analysis (PCA) model is convex. Besides, PCA model has an analytical solution. This means that each time NCA is run, a different result may be obtained. As with other nonconvex algorithms and K-means, therefore advised to perform NCA multiple times and select the best result.

2.2 Critical Analysis of Transfer Learning Techniques for Rock Classification

Rock categorization knowledge is used to perform an image recognition task known as "naked eye rock type identification." Ran et.al (2019) the rapid advancement of image capturing and digital image pattern identification technologies have enabled the invention

of automated processes in order to detect rock from field photographs. By increasing identification efficiency and accuracy, these systems will be of considerable use to geologists. They will also allow students and newly certified geologists to practice identifying different types of rocks. The geological mapping systems and automatic remote sensing used by UAVs can contain identification systems. Lepist¨o et al. (2003) proposed that it is now possible to collect and analyze a range of rock features digitally as digital cameras became widely available and computerized image processing has been advanced. The color, size, and texture of rocks can be clearly displayed in photographs. Despite the fact that photographs of rocks do not always show the same textures, color, or shape, the "digital image analysis" can be employed to categorize different kinds of rock shots. "Gray-Level Co-occurrence Matrices" were utilized by Partio et al. (2002) to extract rock texture from photos. Furthermore, rock photos were categorized in both studies using "spectral and textural characteristics".

The researchers have looked into computerized rock categorization on the basis of conventional algorithms of machine learning and have enhanced the speed and accuracy of rock identification using ML approaches employed into computer image assessments. Lepist¨o et al. (2005) used image analysis to study the characteristics of bedrock, while Chatterjee (2013) tried a genetic based algorithm on sample images from a limestone mine for developing a model for visual classification of rocks built upon imaging and SVM technique. Perez et al. (2015) used the "SVM method" to categorize the types of rocks by extracting features from photographs of boulders that were being transported on a conveyor belt. Deep learning, commonly referred to as "deep neural networks", has drawn interest recently in a number of scientific domains. Numerous deep learning techniques have been put forth. "Deep Convolutional Neural Networks (CNNs)" can improve classification efficiency and accuracy without the use of artificial feature selection by automatically learning the characteristics needed for identifying images from the data of the image. Deep learning algorithms have been presented in recent studies to produce significant improvements in a variety of like image speech recognition, classification, clinical diagnosis, object detection, identification of diseases of plants, and recognition of traffic signals and humans.

Geologists have been researching the use of CNNs in classifying different types of rocks as a result of their success in image recognition. Several researchers have used deep learning to categorize distinct rock kinds pictures. Zhang et al. (2018) employed transfer learning to detect breccia, phyllite and granite, obtaining the Accuracy of 85 percent using the GoogLeNet, Inception v3 models. Cheng et al. (2017) developed a deep learning technique relying on the CNNs in order to precisely detect image pieces of 3 different varieties of sandstones. These studies demonstrate that when used for rock-type identification and geological surveying. CNNs have produced positive results. Without the need for manually selecting image attributes, deep CNNs can recognize different sorts of rocks from photos.

Xu et al. (2021) proposed that deep CNNs haven't yet been used in the real world, hence the accuracy of the results above wasn't good enough to identify rocks. Because of

advances in deep learning technology, CNN's models' accuracy has been steadily increasing. Since models have gotten deeper throughout time, they have become more computationally intensive and time-consuming, leading to these advancements. In this study, an RTCNNs model for categorizing different types of rocks in the wild is suggested. In terms of calculation window, RTCNNs model is relatively low than any model with 10 or more layers. The computation is run on GPUs and CPUs from commonly used devices due to the minimal hardware requirements (GPUs). For the accurate categorization of rock types from field-captured pictures, a deep CNNs model having optimum parameters is recommended. To make original collected picture images more suitable for training the model, the authors creatively cut and patched them. The sliced samples clearly preserve the important rock features and increase the training dataset. Zhu et al. (2020) proposed deep CNNs model ultimately achieved an overall accuracy of 97.96 percent utilizing 24,315 patches of sample images for testing and training. The model provides development in the autonomous classification of rocks in the field due to its better accuracy level than GoogleNet Inception v3, SVM, VGNet-16, and AlexNet models.

Geological surveys can benefit because it is quick and simple to apply in the field to identify the type of rock. In the meanwhile, after retraining the necessary parameters, the method for classifying rock types suggested in the research can be used to detect various textures, such as photographs of rock thin sections, and fossils. Yang et al. (2018) proposed as CNNs have helped in classifying and identifying different kinds of rocks, significant difficulties still exist. First, there is still room for improvement in recognition accuracy. The suggested model was able to correctly identify 97.96 percent of the test dataset's images, which left ninety-nine photos with incorrect identifications. Given that sandstone and limestone have comparable colorations and tiny particle sizes, the model only managed to identify them with a rather low degree of accuracy. Additionally, this study only took into account a small number of sample types (a total of six different rock kinds). According to the mineral makeup, the three primary forms of rock— sedimentary, igneous, and metamorphic are also further subdivided into different categories. Therefore, in the future, the knowledge that has data regarding the connections between various types of rocks and the rocks themselves for the classification of numerous kinds of rocks is combined with the deep learning models. This can increase the scope and accuracy of the identification of several types of rocks in the field.

Currently, most of the manual effort involved in interpreting the geology of these "UAV high-resolution images" is done by hand. Therefore, the large-scale geological mapping will be much more effective in locations with good outcrops where rock types can be automatically identified. UAVs can gather a lot of high-resolution outcrop photographs in these places (like western China). The recommended method might be used to evaluate these photos, helping in both geological interpretation and mapping while minimizing expenses and maximizing efficiency. In order to detect various rock kinds, in addition, the study intends to use other cutting-edge deep learning models like the "Mask RCNN".

2.3 Analysis of Machine Learning approaches for Rock Classification

Yanto et al. conducted research on "fuzzy soft set for igneous rock classification" in 2018. In his study, he used a ML technique, known as 'Fuzzy Soft Set Classifier' (FSSC) to discriminate rocks type and identify a specific category of rocks. FSSC identified igneous according to the chemical formula. The ML algorithm used in this research was a hybrid of two types of machine learning models, 'soft set theory', and 'fuzzy theory on data-classification'. According to this study, the hybridized model, FFSSC were profoundly equipped for precise discrimination and identification of igneous rocks among other rock categories. The authors achieved significant results based on the model's precision, accuracy and recall.

The general functional independency of a machine model is the basic goal of machine learning. In the past years, researchers have focused on advancing kernel-based models to design more flexible and reliable output and input representation sets. The researchers in (Seng et al., 2009) published a paper that described the implementation of two kernel-based machine algorithms. In his paper, he proposed that SVM and Rough Set (RS) algorithms can be practically employed for the classification of ore-rocks. The study also discussed the construction of SVM algorithm using kernel mapping technique. Prior to the classification, the data for original ore-rocks' sample was pre-run with 'Knowledge Reduction Algorithm' of Rough Set theory. The conflicting samples and the tautological attributes were eradicated from the experimental sample sets. This was done to lessen area and dimension of the training data. The study concluded that both Rough Set and SVM can enhance the precision and training speed of ore-rocks' categorization.

Rocks can also be classified on the bases of their penetration power. Similar research was conducted by Hegde and Gray in 2018 on the penetration optimization using machine learning techniques. In this study, the researchers used Random Forest (RF) algorithm. This algorithm was employed to construct a model that coupled weight and penetration rate of the rocks based on rotation speed, drilling rate and, free-range rock strength. The results for this machine learning optimization showed that RF algorithm optimized the penetration rate up to 12 % (Hegde et al., 2018).

Hidalgo et al. (2021) proposed that the essential components of the spectral data extraction system were employed for recording the "pixel-by-pixel fingerprints" of the rock using a "VNIR Specim IQ hyperspectral camera" which has a wavelength of 400-1000 nm, and have multispectral 204 bands. Salles et al. (2017) proposed that this data collection requires standardizing the spectral signature recording technique. Van de Meer (2006) In order to filter out the noise and ensure that the following data is collected under the same standard settings, the manufacturer, "Specim", provided a white reference board for initializing the camera. Because "tungsten-halogen lights" have large output capacities over the Visible Near Infrared Range (VNIR)., which is also the camera capturing range, they are used in the experimental setup to illuminate the stage. According to experts,

Sinaice et al. (2020) “hyperspectral imaging” is the process of gathering information about a topic from the “electromagnetic spectrum” at a scale of hundreds of pixels. The “400–1000 nm electromagnetic range” is where such data were collected, that fell inside the VNIR. When compared to multispectral imaging, hyperspectral imaging is a step up because it has a better resolution over the same spectral range, making it easier to extract fine-grained spectral signatures (Sinaice et al., 2019).

Li et al. (2021) captured the interaction of the subject with light when a hyperspectral image, such as one captured by 204 band Specim IQ capturing camera. As a result, every 204 VNIR spectral bands located within every single roughly 3 nm broad spectral bands detect a unique signal. The researchers noted that the different camera manufacturers might provide different spectral ranges and spectral band counts in their specifications. This essentially influences each spectral band’s breadth, but it does not affect the underlying fingerprints displayed by certain minerals and rocks.
Zhang et al. (2014) obvious that advanced analysis software is needed for the examination of hyperspectral data. This is because the depth of information bands in this type of data—commonly known as dimensionalities, hence the phrase Dimensionality-curse makes it computationally expensive to analyze. A technique known as DR must be used to decrease or delete redundant information in order to combat this issue. To do this, it is necessary to choose the most representative spectral bands that can identify between rocks in their database without impacting or changing the intrinsic distinctions in their spectral signatures.

2.4 Identified Gaps and Conclusion

After studying several sets of research papers, analyzing various dataset, techniques and methods executed by different researchers, it has been identified that very limited amount of Sedimentary rock data is available for research, which makes this research more challenging to get the optimal results from the dataset. Also, the pre-trained models consisting of complex architecture of deep neural network are even not able to correctly classify among the different classes of rocks. Therefore, in this research a Transformer based approach has been utilized, which has not been explored much by previous researchers. Swin Transformer can learn efficiently from very limited amount of image data and generate the optimal results.

3 Research Methodology

The identification of minerals and rocks is not only a significant component in the geological survey but also fundamental area of exploration. The conventional technique of identification necessitates that the spectator possesses extensive geological expertise and knowledge. Strong subjectivity, a protracted identification phase, and a mediocre capacity for field recognition are just a few of the issues the approach faces. Leveraging deep learning algorithms is suggested for recognizing and classifying rocks using images because traditional methods are a highly laborious and time-consuming process. The primary

objective of this research is to identify the best deep learning algorithm that, when trained on a collection of rock images, can more reliably detect, and categorize rocks into their respective classifications. A uniform structure is described that includes several methods, such as data collection, data pre-processing, algorithm setup, model training, and assessment, to accomplish this objective. Each stage is addressed in great depth in this section. The framework for rock classification is shown in Figure 1.

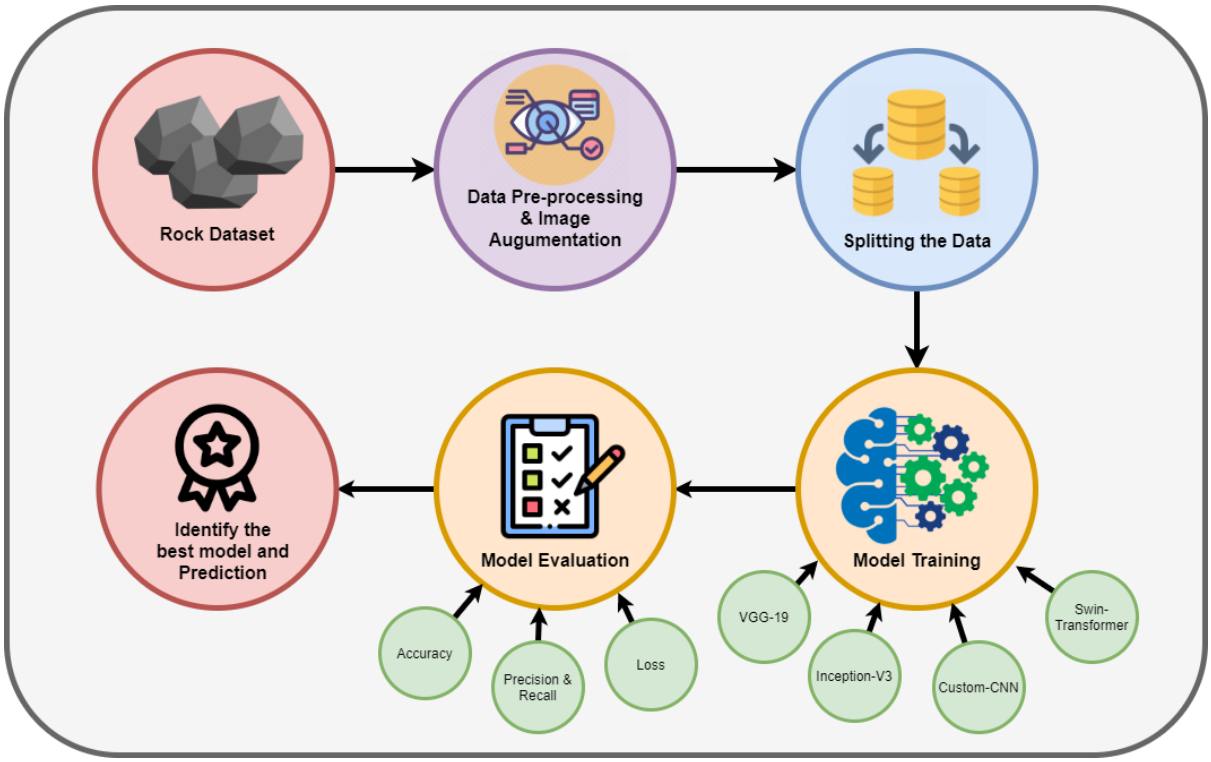

Figure 1 Framework for Rock Classification

3.1 Dataset Description

In geological research, the classification of rocks plays a vital role because it assists the researchers to understand the constituent particles, their nature, and their contribution to the rock's formation. Therefore, for the classification of the rocks the dataset is collected from the igneous metamorphic sedimentary rocks and minerals dataset via authenticated Kaggle website (igneous metamorphic sedimentary rocks and minerals, 2022). This data contains the four categories of different rocks which are Igneous rocks, Metamorphic rocks, Minerals, and Sedimentary rocks. These classes are further divided into 27, 13, 38, and 17 sub-classes. The total number of images by combining all classes is 5887 and format of images is jpg. The size of images is varying, and the channel of the images is RGB. A snap of sample images from dataset is shown in Figure 2.

Figure 2 Sample images from Dataset[1]

3.2 Data Pre-processing

Pre-processing the data is the next action taken after gathering it from a credible source. So, because raw data may include distortion and have varying picture dimensions, pre-processing is necessary when using deep learning-based algorithms for categorization. First, the pre-processing is executed with the help of Keras Image Data generator then the scaling of the images is performed. All images are set to the standard size of 128 X 128. The images are then enhanced in subsequent processes. Together with scaling the data, this augmentation also incorporates techniques like rotation, tilting, zooming, and shear range. If the dataset is tiny, this augmentation and pre-processing aid in reducing model skewing and producing better outcomes.

3.3 Model Training

Once pre-processing is done, the processed data has been split into training and test data with the ratio of 80:20. First, all the images are combined, and then training and test are splitted. After this step number of images in training, data is 4716 which are comprised of four different classes and the test data contains 1171 images for four classes. Here, 4 models based on deep-learning have been used and they are categorized as Custom Model, Transfer Learning and Swin Transformer models. Under transfer-learning methods, Vgg-19 and Inception V3 models are employed whereas, Swin transformer and custom model have been created from scratches utilizing different Convolutional layers. First, all of these models have been trained on train data, then tested using test data.

3.4 Model Evaluation

Once the algorithms have successfully been trained over the training data, so every model has been evaluated using test data. As the activity involves with classification, the classification-based metrics are used to evaluate the model such as Accuracy, Validation loss, Precision and Recall. The finest algorithm is the one that obtains the maximum accuracy, precision and recall score along with minimum validation loss.

[1] <https://www.kaggle.com/datasets/mahmoudalforawi/igneous-metamorphic-sedimentary-rocks-and-minerals>.

4 Design Specification

Deep learning techniques are essential for making inferences and evaluating attributes. Four deep learning methods are thus utilized in this study to categorise the set of images from rock. In order to carry out the transfer learning model and achieve the desired outcomes, two pre-trained models have been employed. Among the final two methods employed in this study, Convolution neural Bespoke designs have been developed on a deep convolutional neural system and another on a swin transformer.

4.1 VGG-19 Model

There are 3 fully linked layers on top of the 16 convolutional layers, bringing the total number of layers to 19. This is explained as having convolutional neural network meaning. The employed database is Image Net, a model that has been used to train just under 1 million photos. In Image Net, which offers photos in over 1000 distinct categories, several categories are defined, including those for automobiles, bikes, trucks, tools, different kinds of animals, numerals, and more. The required image size to process is 224x224, and it is thought to have a good photo category. The proposed model, which uses pre-trained weights, is used to develop an introduction to the idea of transfer learning. As implied by the name, transfer learning includes moving knowledge from one model to another. This model's operation is described in a way that avoids having to start the modelling process from scratch each time and instead uses one model as the basis for another, saving a significant amount of time and processing resources. In many instances, the model comes with a pre-trained model that can be used as a base and the results can be transferred to a new model for increased accuracy and less processing. Let's look at an example in which the model was taught to identify animals based on their eyes, but this led to a new route that might be utilized to train a new model that identifies animals based on their ears or noses. VGG-19's construction is depicted in Figure 3. Two CNNs, VGG19 and ResNet15, were employed as pre-trained models in this investigation along with the transfer learning technique. The VGG-19 model architecture (Wang, et al., 2020) has been demonstrated in Figure 3.

Figure 3 VGG-19 Model Architecture (Wang, et al., 2020)

4.2 Inception_V3 Model

CNNs serve as the model's foundations. As this work investigates the Inception V3 deep learning model that executed for distinguish between images, it could be state that the

upgraded version of Inception V1 is V3, considering basic model's initial release just like Google-Net in 2014. The model was created by a team of Google employees, as the name implies. Due to the data used and the inclusion of numerous deep convolutional layers, the model was deemed to be overfitted. The Conception V1 model uses the same surface as the surface considered above, but with several different-sized filters to avoid the problem. The model's conclusion was that results may be obtained by employing parallel layers in place of deep layers, which makes the model larger rather than deeper. Following the research, this work concluded that the Inception V3 model is a more enhanced and altered version of the Inception V1 model. After researching the Inception V3, it has been discovered that numerous techniques were employed to enhance model compatibility and optimize the network. The model has a wider network than the Inception V1 and V2 models, but it is slower because it is more effective. To reduce the computational expense of normalization, an auxiliary classifier has been used. The 42-layers Inception V3 model, which got significantly less errors than it's predecessor, was introduced in 2015. An architecture of Inception-V3 model (Ali, et al., 2021) has illustrated in Figure 4.

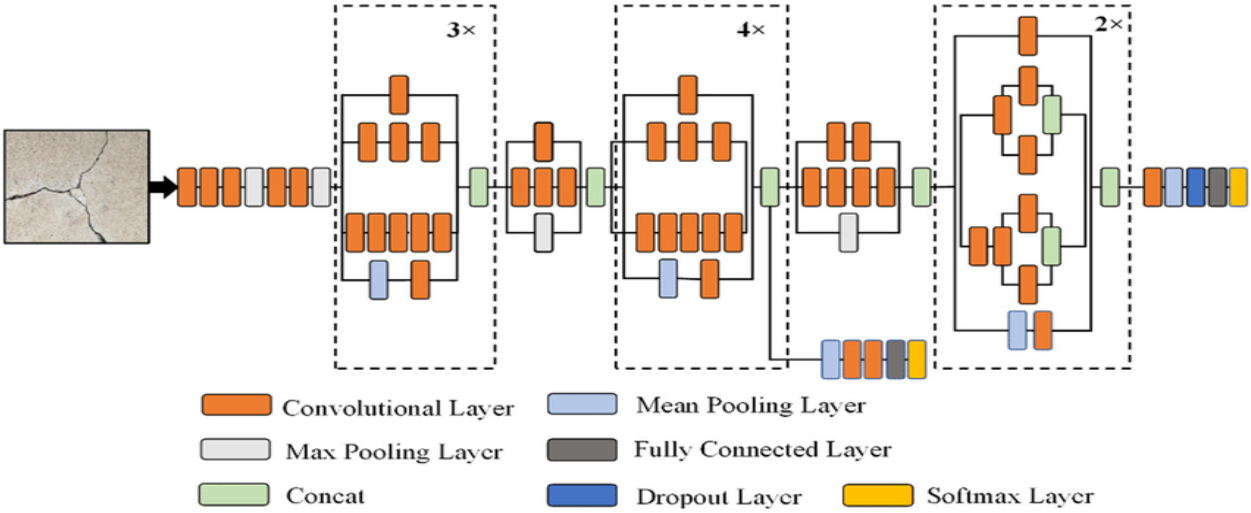

Figure 4 Inception V3 Model Architecture (Ali, et al., 2021)

4.3 Custom Model

This sequence model was developed using TensorFlow and Keras for this study objectives. In this network, 64 neurons were distributed across 4 convolutional 2-dimensional layers with a (3,3) filter, with the pad remaining constant. The keyword " Rectified Linear Activation Unit" (ReLu) is used to activate the ReLu function. Then after, the Batch normalization layer is applied after the max pool layer with the max - pooling layers set to (2,2) has been applied. Now, everything in the 2nd layer, which comes after the first layer, is the same as it was in the first layer aside from neurons. 128 neurons have been employed in the second layer, after which a max pool layer having a pool-size of (2,2) and a batch normalization layer have been added. Following that, the identical layers with 256 neurons were inserted in the third and fourth layers, and then batch normalization and maximum pooling were used. Following the addition of the Flatten layer, the Dense layers having the ReLu activation-function and input capacity 256 were then incorporated. In addition, a Dropout layer of 0.5 followed by batch

normalization has been included to prevent overfitting. Since a dense layer containing 128 neurons that may are activated by Relu is once more used. The output unit of CNN architecture that can forecast a multinomial probability distribution finally adds a dense layer with four neurons that have a softmax activation function. Categorical cross entropy and the Adam optimizer are employed in this model's loss function and optimization, respectively. Basic Architecture of CNN (Kang, et al., 2019) is shown in Figure 5.

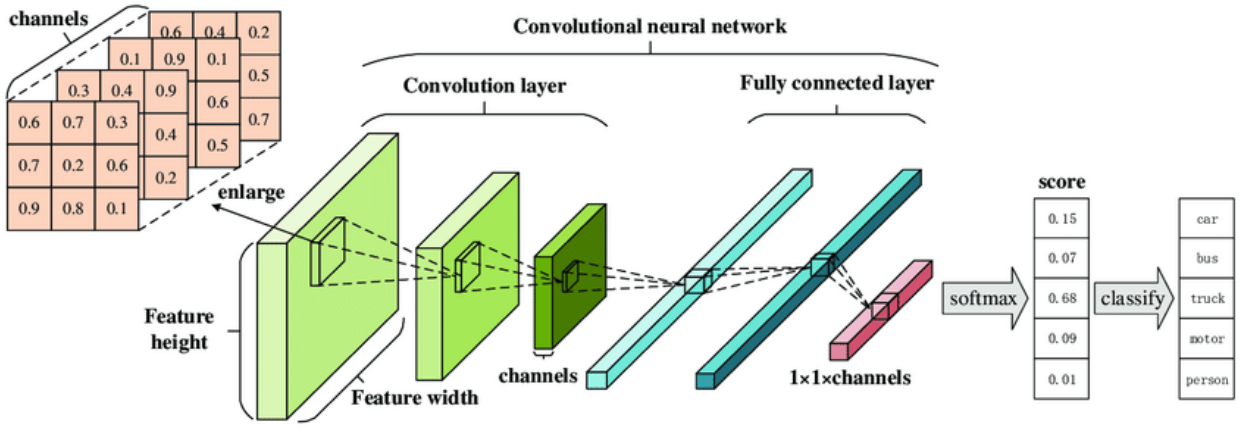

Figure 5 Architecture of Basic CNN (Kang, et al., 2019)

4.4 Swin Transformer

Recently in the year 2021, transformers are introduced for the image vision tasks. These are the state-of-the-art models in computer vision and its related field. First, the transformers are constructed for the NLP tasks but further by incorporating changes and improvements these models are applicable for the image-related tasks too. Swin Transformers stands for the shifted windows transformers which consist of encoder and decoder blocks. These encoders and decoders are different from the conventional encoder-decoder model due to the presence of self-attention and positional encoding layers between the encoder and decoder layers. Swin transformers consist of patch partition, linear embeddings, swin transformer, and patch merging blocks only in stage 1 of the model and this similar architecture is followed for the next 3 more stages except for the linear embedding layer. When an image is input to the model it is partitioned into patches which are further inputted to the vanilla neural network (linear embedding layer) and through the transformer block all patches are merged once again. The transformer block contains input embedded patches, a normalization layer followed by a multi-head attention block again a normalization layer, and finally a multi-layer perceptron. Inside the attention block shifting windows are integrated. When an image patch after embedding enters in attention block it is further divided into small patches and a part of the patches is aligned for the window and this window shifts the patches array thus these are known as shifted window transformers. This division of the image into patches and shift of windows extract features very finely at the pixel level because of which higher accuracies are achieved and these models are the new generation model for the computer

vision task. These models are solutions for the small dataset challenges. The architecture of the swin transformers (Liu, et al., 2021) is shown in Figure 6.

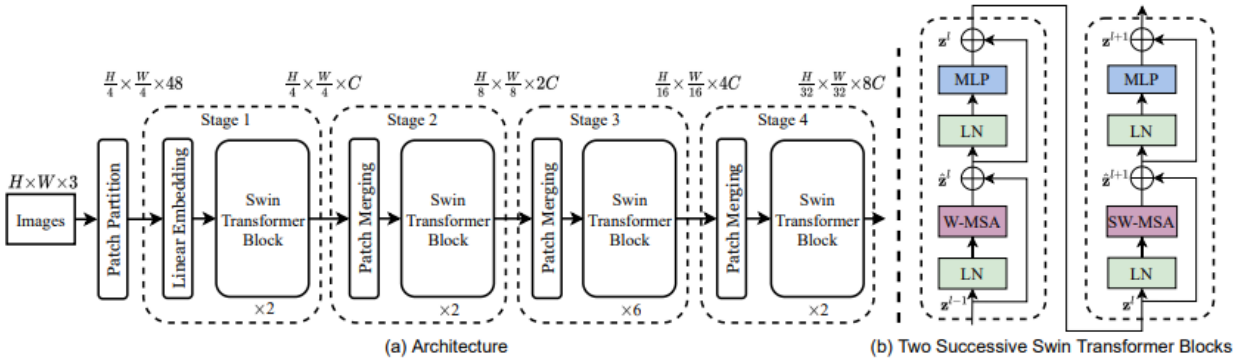

Figure 6 Architecture of swin transformer (Liu, et al., 2021)

5 Implementation

In this project work, the VGG-19, Inception V3, custom model, and swin transformer are four deep learning-based algorithms that are executed. The spectacular model is then taken away, whose score of, precision, accuracy and recall over testing data would be maximum. Each of the algorithms employs categorical cross entropy as their loss function, and Adam has been utilized for optimizing the model's performance. Every evaluation metric is determined after each model has been trained on the exact same data and validated on the exact same data. Numerous libraries, including OpenCV, SciPy, Sklearn, NumPy, plotly, tqdm, Keras, and TensorFlow are used during the execution. Kaggle Framework is chosen for the learning of the algorithms because it offers free GPU resources, and the suggested technique incorporates CNN. Python is being utilized in this case for programming. The following provisions are required in order to execute the model.

Operating System	Windows 10
RAM	16 GB
Disk Space	200 GB
Programming Language	Python 3.8.8
Framework	Jupyter-Notebook
Libraries	Sklearn, OpenCV, plotly, Matplotlib, Keras, TensorFlow

6 Evaluation

The objective of this project is to determine the most appropriate model for classifying rocks relying on pictures for geology research and in particular their utilization. Rocks image

dataset is divided into four classes: igneous, metamorphic, minerals, and sedimentary, thus it is crucial to evaluate each model using the classification metrics. This categorization is actually a challenge for many classes. Metrics including accuracy, precision, and recall are generated for each of the four deep learning-based algorithms and tested using test data. Each model is evaluated based on its biases for accuracy, precision-recall, and validation loss after a series of training. For making future predictions, the model with the greatest value in these metrics is chosen. For a better understanding of the scores observed line plots are also plotted to make a visualization of comparison.

6.1 Experiment 1 (Evaluation Based on Accuracy)

First model implemented in this study is vgg-19 which has been trained upon 50 epochs and while training, the observation depicts that it was learning as its accuracy was continuously increasing throughout all the epochs with a very slow rate on this data because this model is deep in nature but due to simple convolutional layers, the bottleneck features could not extract out from the image data. After the post-training period of the model, the accuracy achieved on the validation data is 32.54%. On the analysis of the training curve, it is noticed that the slope of the curve is very low but increases at each epoch and on the validation, similar observations are assessed which are indicating that the model is not overfitting but in some sort of underfitting which can be justified because the dataset is small for such deep CNNs.

The second model executed is the Inception V3 model which is another set of transfer learning model. Inception-V3 model is also trained over 50 epochs. During the training process, the model was observed to be learning the features as the validation loss was continuously decreasing. The highest level of accuracy of Inception V3 model was 46.29% which was achieved during the model training.

The third model executed in series is Custom CNN (Convolutional Neural Network) model which has been trained over 50 epochs. This model is constructed using simple convolutional layers and due to the sparse training data size, this has been noted that the model is unable to extract the relevant features but still it was learning from the data. The maximum accuracy of validation data observed for this model is 29.04%.

The last model applied in this task is the swin transformer and this model is also built from the scratch by relying on the concept of the transformers which includes the encoder and decoder model. This model is also trained upon the 50 epochs on the same training data used for the previous models and after complete training, it is assessed on the validation data and the observed accuracy on the validation data is 73.10%.

After the comparison among executed models in terms of validation accuracy depicted in Figure 7 and after the inspection it has been highlighted that swin transformer got a better accuracy score of 73.10% followed by the Inception_V3 model at 46.29% and then the Vgg-19 model having an accuracy of 32.54%.

Figure 7 Comparison of Models Based on Accuracy

6.2 Experiment 2 (Evaluation Based on Precision and Recall Score)

In the second experiment, the models that have been used are assessed using precision and recall. Precision is related to the values of the false positives and the recall concerns about the value of the false negatives. In this experiment for each model value of precision and recall is calculated. For the vgg-19 model, the value of precision and recall is 0.3254 and 0.3254 respectively. These values remained for the first three epochs and then vanished while for the other pre-trained model inception_v3 model the observed value of precision and recall are 0.4684 and 0.4500 respectively. The custom model did not perform well on this data; therefore, the value of precision and recall are 0.3360 and 0.1067 respectively but the swin transformer which is also from scratch achieved the values 0.8036 and 0.6430 for precision and recall respectively. Based on PR scores, all of these models that have been implemented are compared and illustrated in Fig 8 and 9. The swin transformer model achieved the highest PR score and outperformed the inception_v3.

Figure 8 Comparison of Models Based on Precision

Figure 9 Comparison of Models Based on Recall

6.3 Experiment 3 (Evaluation Based on Validation Loss)

In the third experiment, validation loss is used for the evaluation of the executed models. Validation loss depicts the training of the model and is used for the identification of the overfitting or the underfitting of the algorithms. For better results, the validation loss must be continuously decreasing while training because its opposite is the indication of the overfitting of the algorithms. The vgg-19 model achieves the value of 1.3849 while the value of the validation loss for inception_v3 value is 5.9049. The custom model accomplishes the validation loss value of 2.2761 and the swin transformer model achieved 0.9855 as the validation loss value. The comparison of validation loss has been displayed in Figure 10.

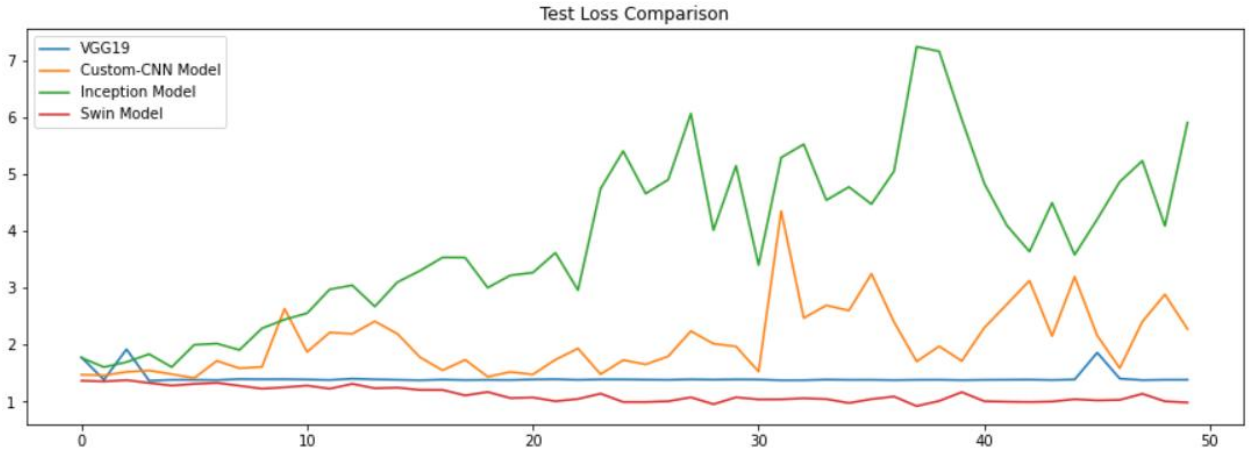

Figure 10 Comparison of Models Based on Validation Loss

6.4 Discussion

In this research task after carrying out three trials on the dataset using the deep learning algorithms for this research objective, it was discovered that the Swin-transformer outperformed the other models. Transfer learning techniques have also been used in this work, where a model like Inception V3 or VGG-19 that was trained over an Imagenet dataset

was performed, yet a cutting-edge algorithm swin transformer outperformed rest of the models. In this work, test Precision and test Recall have been applied for the assessment of the test accuracy to compare the models. Swin transformer accomplished 73.10% Accuracy over test data while the model's acquired value of recall and precision were 0.8036 and 0.6430 respectively. This model may be further used to provide excellent predictions because it produced the best outcomes. This model may be used in real-time applications in place of conventional techniques of classification to classify rocks based on images. Additionally, a customized model is evaluated in this task which exhibits some poor performance than other applied models due to the minimal dataset used for model training. Large amount of data is necessary to execute these tasks for sequential custom models to function properly. Besides, the structure of custom architecture is simple compared to the discussed models. The performance of the pre-trained models such as vgg-19 and inception_v3 is also poor because in the transfer learning techniques, the models required a large amount of data to be trained. Small datasets are not good for model training because the feature extraction from images using complex architecture or deep models requires a large amount of data. As a result, neither model can be used to make predictions in real world executions.

7 Conclusion and Future Work

Due to the fact that these rocks are obtained from the earth's crust and have varied imperfections on their exterior, classifying and categorizing them into corresponding grades and phases is a highly laborious process and a very broad topic of study. To detect and classify images of rocks into the appropriate categories, various deep learning techniques are applied in this task. Due to the fact that deep learning models are effective in segmenting, categorizing, and classifying images into the appropriate categories regardless of the subject, this research task is extensive and constantly growing. This study is associated with four deep learning models i.e., Inception V3, Vgg-19, Custom Model and Swin Transformer that are trained and examined. The outcome depicts that the classification of rock images is adequate for the swin transformer and that is why the additional three models were excluded as the vgg-19, inception_v3 and custom model could not obtain the utmost accuracy owing to limited data and their simple architecture is limiting the learning of the algorithms by improper extracting the features from the images. The swin transformer model acquired an Accuracy of 73.10% with the Precision and Recall value of 0.8036 and 0.6430 respectively. The images are divided into four categories using this model. It has been noted that the data used for this assignment is quite little, which has an impact on the model's accuracy and overall efficiency. Future research can use a huge dataset to boost the models' accuracy and produce better predictions. Further efforts that might represent an advancement in the automated recognition of rocks could also integrate segmentation and masking of the various portions of the rock image.

Acknowledgement

I would like to showcase my sincere gratitude to Professor Jorge Basilio who has supported and guided me throughout the thesis module to accomplish this project. Without his enormous assistance, this work would not have been possible.

References

Chatterjee, S. (2013). Vision-based rock-type classification of limestone using multi-class support vector machine. Applied intelligence, 39(1), 14-27.

Cheng, G., Guo, W., & Fan, P. (2017). Study on rock image classification based on convolution neural network. Journal of Xi'an Shiyou University (Natural Science Edition), 32(4), 116-122.

Goldberger, J., Hinton, G. E., Roweis, S., & Salakhutdinov, R. R. (2004). Neighbourhood components analysis. Advances in neural information processing systems, 17.

Hegde, C., Daigle, H., & Gray, K. E. (2018). Performance comparison of algorithms for real-time rate-of-penetration optimization in drilling using data-driven models. Spe Journal, 23(05), 1706-1722.

Hidalgo, D. R., Cortés, B. B., & Bravo, E. C. (2021). Dimensionality reduction of hyperspectral images of vegetation and crops based on self-organized maps. Information Processing in Agriculture, 8(2), 310-327.

Kaggle.com. 2022. igneous metamorphic sedimentary rocks and minerals. [online] Available at: <https://www.kaggle.com/datasets/mahmoudalforawi/igneous-metamorphic-sedimentary-rocks-and-minerals>.

Koren, Y., & Carmel, L. (2004). Robust linear dimensionality reduction. IEEE transactions on visualization and computer graphics, 10(4), 459-470.

Lepistö, L., Kunttu, I., Autio, J., & Visa, A. (2003). Rock image classification using non-homogenous textures and spectral imaging.

Lepistö, L., Kunttu, I., & Visa, A. J. (2005). Rock image classification using color features in Gabor space. Journal of Electronic Imaging, 14(4), 040503.

Li, Y., Chai, Y., Zhou, H., & Yin, H. (2021). A novel dimension reduction and dictionary learning framework for high-dimensional data classification. Pattern Recognition, 112, 107793.

Partio, M., Cramariuc, B., Gabbouj, M., & Visa, A. (2002). Rock texture retrieval using gray level co-occurrence matrix. Proc. of 5th Nordic Signal Processing Symposium,

Perez, C. A., Saravia, J. A., Navarro, C. F., Schulz, D. A., Aravena, C. M., & Galdames, F. J. (2015). Rock lithological classification using multi-scale Gabor features from sub-images, and voting with rock contour information. International Journal of Mineral Processing, 144, 56-64.

Ran, X., Xue, L., Zhang, Y., Liu, Z., Sang, X., & He, J. (2019). Rock classification from field image patches analyzed using a deep convolutional neural network. Mathematics, 7(8), 755.

Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning applications-a vision. Global Transitions Proceedings, 2(1), 24-28.

Sinaice, B. B., Kawamura, Y., Kim, J., Okada, N., Kitahara, I., & Jang, H. (2019). Application of deep learning approaches in igneous rock hyperspectral imaging. International Symposium on Mine Planning & Equipment Selection,

Tong, Z., Gao, J., & Zhang, H. (2017). Recognition, location, measurement, and 3D reconstruction of concealed cracks using convolutional neural networks. Construction and Building Materials, 146, 775-787.

Van der Meer, F. (2006). The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. International journal of applied earth observation and geoinformation, 8(1), 3-17.

Xu, X., Feng, Z., Cao, C., Li, M., Wu, J., Wu, Z., . . . Ye, S. (2021). An improved swin transformer-based model for remote sensing object detection and instance segmentation. Remote Sensing, 13(23), 4779.

Yang, C., Li, W., & Lin, Z. (2018). Vehicle object detection in remote sensing imagery based on multi-perspective convolutional neural network. ISPRS International Journal of Geo-Information, 7(7), 249.

Yanto, I. T. R., Sutoyo, E., Apriani, A., & Verdiansyah, O. (2018). Fuzzy soft set for rock igneous clasification. 2018 International Symposium on Advanced Intelligent Informatics (SAIN),

Zhang, X., & Li, P. (2014). Lithological mapping from hyperspectral data by improved use of spectral angle mapper. International journal of applied earth observation and geoinformation, 31, 95-109.

Zhang, Y., Li, M., & Han, S. (2018). Automatic identification and classification in lithology based on deep learning in rock images. Yanshi Xuebao/Acta Petrologica Sinica, 34(2), 333-342.

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., & Dai, J. (2020). Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159.

Wang, W., Tian, J., Zhang, C., Luo, Y., Wang, X., & Li, J. (2020). An improved deep learning approach and its applications on colonic polyp images detection. BMC Medical Imaging, 20. https://doi.org/10.1186/s12880-020-00482-3

Ali, L., Alnajjar, F., Jassmi, H., Gochoo, M., Khan, W., & Serhani, M. (2021). Performance evaluation of deep cnn-based crack detection and localization techniques for concrete structures. Sensors, 21, 1688. https://doi.org/10.3390/s21051688

Kang, X., Song, B., & Sun, F. (2019). A deep similarity metric method based on incomplete data for traffic anomaly detection in iot. Applied Sciences, 9, 135. https://doi.org/10.3390/app9010135

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows, 9992–10002. https://doi.org/10.1109/ICCV48922. 2021.00986