TECHNIUM

Technium Vol. 6, pp.1-22 (2023)
ISSN: 2668-778X
www.techniumscience.com

Object-based Classification of Natural Scenes Using Machine Learning Methods

MOHAMMED SAADULDEEN JASIM [1]

MOHAMMED CHACHAN YOUNIS [1]

[1] Computer Sciences dept., College of Computer Sciences and Mathematics, University of Mosul, Mosul. Iraq.
Mohammed Chachan Younis, mohammed.c.y@uomosul.edu.iq

Abstract. The replication of human intellectual processes by machines, particularly computer systems, is known as artificial intelligence (AI). AI is an intelligent tool that is utilized across sectors to improve decision making, increase productivity, and eliminate repetitive tasks. Machine learning (ML) is a key component of AI since it includes understanding and developing ways that can learn or improve performance on tasks. For the last decade, ML has been applied in computer vision (CV) applications. In computer vision, systems and computers extract meaningful data from digital videos, photos, and other visual sources and use that information to conduct actions or make suggestions. In this work, we have solved the image segmentation problem for the natural images to segment out water, land, and sky. Instead of applying image segmentation directly to the images, images are pre-processed, and statistical and textural features are then passed through a neural network for the pixel-wise semantic segmentation of the images. We chose the 5X5 window over the pixel-by-pixel technique since it requires less resources and time for training and testing.

Keywords. Image Segmentation, Machine Learning, Features

1 Introduction

For the last decade, ML has been applied in computer vision (CV) applications [1]. In computer vision, systems and computers extract meaningful data from digital videos, photos, and other visual sources and use that information to conduct actions or make suggestions. Artificial intelligence allows computers to comprehend, but computer vision enables them to see, watch, and understand. Human vision and computer vision are comparable, yet humans have an edge. The human visual system has decades of context to learn how to distinguish things apart, how far away the things are, if they're moving, and if a picture is inaccurate [2].

In computer vision (CV), models are trained to do the activities using algorithms, cameras, and data rather than optic nerves, visual cortex, and retinas [3]. A machine trained to inspect products or track a production asset could examine hundreds of processes or products per minute, identifying flaws or irregularities that are otherwise invisible [4]. Computer vision is used in a broad variety of

sectors, from manufacturing and automobiles to energy and utilities [5]. It is estimated to reach $ 48.6 billion by the end of 2022 [6]. One of the key applications of computer vision is image recognition, which involves training models to classify and identify objects, scenes, and people in images. One example of a computer vision application is the development of self-driving cars. These vehicles rely on a combination of sensors, such as cameras and lidar, to gather visual data about their surroundings. Advanced machine learning algorithms are then used to analyze this data and make real-time decisions about how to navigate the environment and avoid obstacles. Other examples of computer vision applications include security and surveillance systems, which can be trained to detect and identify suspicious activity; medical image analysis, which can be used to diagnose diseases or identify abnormalities; and industrial inspection, which can be used to identify defects or problems in manufacturing processes.

Image segmentation is a commonly used technique in digital image processing and analysis. It is mainly used to partition an image into multiple parts or regions, often based on the characteristics of the pixels in the image. Image segmentation could involve separating foreground from background, or clustering regions of pixels based on similarities in color or shape. Figure 4 provides an example of image segmentation. Recently, with the advancement of modern computer vision-based techniques, image segmentation has been a focused research topic for the industry and research community. A plethora of research methods are available for fast image segmentation. However, most of the methods are overly complicated and resource exhaustive. Due to higher resource requirements and human dependency, they are not very helpful in real-time application.

In this work, we have solved the image segmentation problem for the natural images to segment out water, land, and sky. Instead of applying image segmentation directly to the images, images are preprocessed, and statistical and textural features are then passed through a neural network for the pixel-wise semantic segmentation of the images. We chose the 5X5 window over the pixel-by-pixel technique since it requires less resources and time for training and testing.

The rest of the paper is structured as follows. The section 2 shows the detailed literature review on this topic. Section 3 goes for the mathematical principles employed in this study and methodology. Section 4 discusses the suggested strategy, while section 5 discusses the findings. Finally, in section 6, the findings are summarized.

2 Literature Review

This section discusses the existing literature related to Feature extraction and image segmentation. The studies [7], [8] make use of red, blue, and green, band combinations to maximize overall distinction between foreground & background elements such as plants and soil. Excess green minus excess red (ExGR) or excess green (ExG) indices were employed incorporation with automated and positive criteria for automatic background or foreground segmentation. However, the indices of other color have been created to increase segmentation quality and to deal with ambient lighting [9]. The article referred in [10] refers to research that presents a cloud detection approach based on deep learning (DL) called multi-scale convolutional feature fusion using various sensors' remote sensing. The symmetric encoder-decoder module of network architecture offers both global & local environment by increasing density of feature maps using convolution which can be trained filter edges used to retrieve high-level multi-scale majority of the spatial data covered in the study. The spatial features can be used to determine contrast, energy, uniformity, and various other mentioned features of consideration in this study. The study in [11] provides a completely automated pipeline that detects every object in a class videos set. Based on motion segmentation, the technique identifies candidate spatio-temporal tubes and afterwards picks one tube for one video jointly across all videos. The studies for the method used in [12] show that it can segment several kinds of pictures such as gray scale, texture, & color, and outperforms the old technique.

The research alluded in [13] employed multi-level regressions using random effects of neighborhood-level to investigate associations among scores of GDS-15. These blue & green space exposed

measurements where depressed symptoms were measured using the abbreviated Geriatric Depression Scale (GDS-15). The paper's findings reveal that blue and green space metrics produced from photos of street view, are not associated with the measures based on satellite. While normalized difference vegetation index (NDVI), a dimensionless measure that distinguishes between plant cover's reflectance in the visible and near-infrared, was substantially connected to GlobeLand30 green space (a typical dataset), NDWI was only modestly connected with it. The study [14] refers to an article that explores the relationships between socioeconomic (SES) & street green space elements within Los Angeles (LA) County, California. The photos of Microsoft Bing Maps combined with DL algorithm were utilized to assess overall categories of the street view green space, that were contrasted to the generally employed satellite-based green space metric also known as NDVI. NDVI was substantially connected to green space consisted of total street views-& trees, but only weakly related to low-lying grass and vegetation. Generally, green space, as well as three categories of green space, had substantial negative connections to local SES.

A correlation criterion is employed in [15] to determine the relationship between psychological perceptions of residents and urban visual space, examine the visual space metrics impact on the subjective view of residents, and concentrate on if blue-green space has been affected by enclosure, openness, traffic flow, walkability, and, imageability within the process of its impact on the psychology of city residents. The research of [16] also analyzes correlation like an image's distinctive feature. The study in [17] investigates the relationship between socio-economic variables, and inhabitants' demographic such as gender, monthly income, education, and housing location, and their desire to live near to urban green space to take part in leisure activities (the recreation need). The RGB color data of item was extracted from background as well as other objects employing multilevel thresholding. A range of nature images were used to study color information. The color information can also be used to analyze the image using various other spatial and statistical features mentioned in the considered features for literature study. The research in [18] showed that convolutional neural networks (CNN) are appropriate for studying 3D scene layout using noisy labels and deliver a 7% relative improvement over the baseline by conducting qualitative and quantitative testing on publicly accessible datasets. Furthermore, merging color planes offers a road areas' statistical description with maximum homogeneity and an 8% improvement over the baseline. It is improved further by combining collected and current data from a single picture.

The work in [19] proposes a multi-feature learning approach for quantifying outdoor field vegetation development. The presented technique is compared to cutting edge as well as other digital picture learning techniques. All approaches are assessed and compared using a variety of environmental situations as well as the following criteria: (1) comparing ground-truth photos, (2) variance over the course of a day due to variations in changing light, (3) comparison with human measurements, and (4) performance assessment throughout the course of wheat canopy's whole life cycle. The suggested technique in [20] was tested on the KITTI dataset that was publicly available and had been supplemented with extra pixel & point-wise semantic annotations for the sidewalk, building, sky, vegetation, road, automobile, pedestrian, sign/pole, bicycle, and fence areas.

In [21], features are introduced that project 3-dimensional cues back toward the 2-dimensional picture plane while representing spatial context and layout. A randomly selected decision forest integrates several of these variables in order to accomplish coherent 2-dimensional segmentation and detect the item categories available. The research in [22] employs machine learning algorithms on a machine vision strategy to deal with color, spectral, texture, and shape analysis from object images. This study thoroughly examined the current use of statistical ML algorithms in systems of machine vision, analyzed every technique's capability for specialized application, and provided a review of instructive instances in several agricultural fields. In [23], a comparison examination is made between LAB & HSV color spaces in terms of color picture segmentation. When mean square error (MSE) & PSNR are used as performance criteria, it is discovered that HSV outperforms LAB. Color is the most often used attribute for visual plant segmentation in the articles cited in [24] because of its cheap processing cost,

resilience, particle occlusion, and resolution variations. Color qualities are readily retrieved and remain reasonably stable when seen from different angles. Color-based approaches, on the other hand, have difficulty preserving color consistency within and between photographs of the same item owing to variations in lighting conditions, shadows, inter-reflections with many objects, and so on. Color space conversions like HSL, YCbCr, CIELab, HSV, and CIELuv could help to reduce these impacts.

The study in [25] investigates the resilience of several color spaces (for example, HSV and RGB) for the water segmentation, as well as how to employ multiple color channels for testing and training to optimize robustness. We assess segmentation accuracy on a difficult, previously unreported test dataset collected under radically varied settings and with various equipment. The research [26] examined the clustering-based image segmentation approaches used on magnetic resonance pictures of the human brain to discover white matter hyperintensities. They carried out assessment approach in MATLAB to demonstrate the proper valuation procedure. By using a computer-assisted tool, they determined the best approach for watching and evaluating medical pictures. This study [27] provides a quick overview of the most prevalent segmentation approaches, including as model-based segmentation, thresholding, clustering, edge detection etc. highlighting their benefits and downsides. They expressed that some of the strategies work well with noisy photos. They concluded that Markov Random Field was the most effective approach of picture noise cancellation, while thresholding was the most basic technique for segmentation. The picture segmentation techniques covered in [28] could be graded according to their application, appropriateness, computational cost and performance. They considered Gray-level segmentation methods like as region-based algorithms and thresholding as the most basic segmentation method that had limited uses. They observed that the efficiency of these method, however, would be enhanced by combining them with AI approaches. They discussed that textural feature techniques based on atlases or look-up tables provided great results in medical picture segmentation. They also described the disadvantage of the atlas-based approach that it had become difficult to choose and label data under specific conditions; it had issues segmenting complicated structures with changeable form, features, and size. They concluded that in such cases, unsupervised approaches like the fuzzy-c-means methods were preferable. The study cited in [29] the typical picture segmentation approaches like single linkage, clustering, hybrid linkage, spatial clustering, split & merge, and region expanding. They observed that these methods could be strengthened if they relied on some form of statistical testing for equal distribution of means.

3 Materials and Methods

Image segmentation is described as a function that processes image inputs and outputs a segmented image. The result is a matrix or a mask with different components indicating which instance or object class each pixel belongs to. Image segmentation could benefit from a number of high-level image properties or relevant heuristics. These characteristics serve as the foundation for traditional image segmentation techniques that employ clustering methods such as histograms and edges. The most typical heuristic example is Color. Graphic designers mostly utilize the green screen to assure that the picture background is consistent in color, allowing for automated background replacement and detection in post-production.

Contrast is another important image segmentation tool which can readily discern between a bright background such as sky and a dark figure. Based on sharply contrasting data, the algorithm determines pixel boundaries. Conventional image segmentation algorithms relying on these heuristics could be quick and easy, but they sometimes need extensive fine-tuning to meet particular use cases via manually generated heuristics. These heuristics are not always precise enough to be used for complicated pictures. ML and DL are used in newer segmentation approaches to improve flexibility and accuracy.

Model training is used in image segmentation algorithms based on machine learning to increase the program's capacity to recognize relevant characteristics. The technology of deep neural network is very useful for the image segmentation. There are several neural network architectures and

implementations that are appropriate for the image segmentation. They all have the same essential components that are given below:

An encoder is a set of layers to extract picture data by applying deeper and narrower filters as the layers proceed. The encoder could have been pre-trained on a repetitive task for example, image detection, enabling it to use its prior expertise to execute segmentation tasks. A decoder is a sequence of layers that progressively turn the output of encoder into a segmentation matrix according to the pixel resolution of the input picture. Skip connections are multiple long-distance neural network connections that enable the model to recognize characteristics at various sizes in order to improve model accuracy. We employed following method for the image segmentation of natural photos in our research.

3.1 Feature Extraction

The idea of this project is to use natural images and segment out Water, Land and Sky area pixels by extracting statistical and textural features [30] and utilizing machine learning algorithms [31]. The features of interest under consideration used for the accurate segmentation of the images are as follows:

3.1.1 Contrast

The contrast of a picture is the variation in brightness between its dark and light portions [32]. It also controls how many shades are in the picture. A low-contrast picture contains detail but lacks depth and seems soft [33]. Normal contrast images are sharp and maintain detail and depth.

3.1.2 Correlation

Correlation is a measure of similarity between two images [34]. The correlation of two signals is determined mathematically by Equation (1) below.

$$\int_{-\infty}^{\infty} a(t)b(t-\tau)dt \tag{1}$$

The cross-correlation of two images (i.e., matrices) is calculated by taking out the sum of element-by-element multiplication for every possible overlap of the two images; using the full-size image, this sum would need to be calculated approximately 2 million times [35].

3.1.3 Energy of a grayscale image

Energy of a grayscale image can be mapped to ‘intensity’ or ‘brightness’[36]. It is named energy due to some historical perspective, where, if the images are considered to be transmitted via some kind of analog modulation [37], the bright pixels having high intensity mean the ones with high voltage and dark pixels having low intensity mean the ones with low voltage [38]. The white parts of image in cathode ray tube (CRT) [39] television will be significantly at higher energy than the dark part of images. When attempting to improve a codec or algorithm by analyzing the difference between the encoded-decoded and original signal [40], the term "energy minimization" is also used [41]. If the difference was entirely zeros, the codec would be flawless, hence the goal is to reduce the difference signal energy.

3.1.4 Uniformity

Uniformity is also known as Angular Second Moment (ASM) [42]. Then energy is simply the square root of ASM. This is a relative term in image processing [43]. Equation (2) provides the mathematical interpretation of ASM. The energy also means the level of steadiness (uniformity) [44] in pixel intensities as explained in Equation (3).

$$\text{ASM} = \sum_{i,j=0}^{N-1} (P_{ij})^2 \quad (2)$$

$$\text{Energy} = \sqrt{(ASM)} \quad (3)$$

where, P_{ij} refers to the probability (i.e. the normalized histogram [45]) of the color intensity at position (i,j) and N is the gray level (e.g., 256 in typical grayscale images).

3.1.5 HUE

The simplest definition of hue is the wavelength in the visible light spectrum where a source produces the highest energy [46]. The intensity vs. wavelength graph shows this as the apex of curves [47]. Hue describes the primary wavelength of light which the human eye perceives as color, however it is only the fundamental color on color wheel, while saturation indicates the color's strength [48].

3.1.6 The HSV Color Scale

HSV is an abbreviation for 'Hue Saturation Value' [49]. It is sometimes referred to as HIS (for intensity)[50] and HSB (for brightness) [51]. An HSV color wheel [52] can take the shape of a cylinder or a cone, but it always has these 3 parts. It is a scale that offers a numerical interpretation of the picture and correlates to the colors included in it. Designers use the HSV color models when picking colors for ink or paint because it better captures how humans connect to colors as contrasted to system of RGB color [53]. This HSV color wheel additionally assists to create high-quality instances. Despite HSV is not so popular than CMYK and RGB [54], it is a technique that is supported by many high-end photo editing software packages. Choosing an HSV color starts with selecting one of the accessible colors, followed by adjusting the shade and brightness values. Hue is expressed in degrees ranging from 0 – 360 [55].

3.1.7 Saturation

Saturation defines the percentage of gray within a color that ranges from 0 to 100% [56]. When this factor is set to 0, more gray is injected, results in a fading appearance. Saturation is often depicted on a 0–1 scale, with 0 representing gray and 1 being the primary color [57].

3.1.8 Value or Brightness

Value or Brightness [58] works in the tandem with saturation to represent the color's intensity or brightness on a scale of 0 to 100%, with 0 being absolutely black and 100 being the brightest.

3.1.9 The average/mean

The hue value is computed by taking the circular quantities' mean. Because color is a periodic variable [59], the conventional method of determining the average cannot be applied (modulo 360).

Given a collection of N hues H = {h_1, h_2, ...} in range of $h_i \in [0,359]$, transform every hue to the radians employing $h_i^r = \pi h_i/180$, and afterwards calculate the angle generated by average vector of overall hue angles as depicted in equation (4) below.

$$\bar{h}^r = \arctan\left[avg_i \sin(h_i^r), avg_i \cos(h_i^r)\right] \quad (4)$$

Then $\bar{h}^r$ is converted to degrees using the basic formula $(180/\pi)*\bar{h}^r \pmod{360}$.

The average vector's length can be computed using equation (5) below.

$$\left\|\bar{h}\right\| = \sqrt{avg_i \sin\left(h_i^r\right)^2 + avg_i \cos(h_i^r)^2} \quad (5)$$

Both $\bar{h}^{r}$ and $\bar{h}$ are reported. $\|\bar{h}\|$ is close to zero if there is a balance of complementary hues in the image and is close to 1 if all the hues in the photo are same.

3.1.10 Skewness

Saturation/Value mean for an HSV image is calculated by simply taking the maximum and minimum value of saturation/value of an image [60]. The basic value of skewness can be estimated by equation (6) below.

$$\text{Skewness} = \sum_{i=1}^{N} \frac{(X_i - X)^2}{(N-1)\sigma^3} \tag{6}$$

Where the 'σ' is the standard deviation and is determined by using equation (7) below:

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{(X_i - \bar{X})^2}{N}} \tag{7}$$

The formula can be used to find the hue/saturation/value skewness [61], which are useful statistical parameters to analyze the image in terms of its statistical characteristic features [62].

3.2 Image Segmentation

As per digital image processing and CV [63], segmentation refers to the technique of splitting a digital picture into multiple segments, also described as objects or regions of the photo [64]. They are the pixels sets [65]. The goal of segmentation process is to convert a picture into a more understandable and simpler representation [66], [67]. Image segmentation has also been used to determine the borders of items like lines, curves, and so on. Photo segmentation, in a nutshell, allocates tags to every single pixel within an image. Certain features are shared by pixels having the same label. Following image segmentation, a collection of contours taken from the picture or segments set encompassing the full image is returned [68]. Based on calculated qualities and attributes, each picture is assigned a class. These calculated attributes could include intensity, texture, color, and so on.

Image segmentation [69] produces a collection of segments which encompass the full image, or a group of contours taken from the picture as described in edge detection. Every pixel in an area has a comparable characteristic or calculated attribute [70], like intensity, color, or texture. Every region is given a unique color. If employed to a heap of photos, as is common in medical imaging, the contours generated after image segmentation could be utilized to produce 3-dimensional reconstructions using interpolation methods such as marching cubes. Image segmentation could be further divided into three different categories that are instance segmentation[71]–[74], semantic segmentation, and panoptic image segmentation.

3.2.1 Instance Segmentation

This type of image segmentation detects instances of objects and demarcates their boundaries [75]. Among its many real-world applications are self-driving cars, medical imaging, aerial crop monitoring, and more [76]. When multiple objects of a similar type need to be monitored separately, instance segmentation is particularly useful [77].

3.2.2 Semantic Segmentation

Semantic image segmentation entails classifying every pixel of a picture with a class that corresponds to what it depicts [78]. This task is commonly known as dense prediction because we are predicting for every pixel within the photo. It is very important that we are not separating instances of the same class, only the category matters for each pixel [79]. The segmentation map does not distinguish between two objects of the same category if they are in the same input image. It differs from instance segmentation models such that they distinguish between separate objects belonging to the same class [80].

3.2.3 Panoptic segmentation

In panoptic segmentation, the prediction from both instance and semantic segmentation is combined into a unified output [81]. Studying things and stuff is part of panoptic segmentation.

Many products requiring vast amounts of information can benefit from panoptic segmentation. A self-driving vehicle, for instance, must be able to capture and understand its surroundings quickly and accurately [82]. A panoptic segmentation algorithm can be used to segment a live stream of images [83].

Neural Networks such as deep neural networks or convolution neural networks are used for image segmentation [84]. Convolution neural networks generally take the raw images as input and convolve its weight kernel over the image to learn features and finally a flattened layer concatenates all features and then each pixel is assigned a label [85]. However, a deep neural network computes the features before passing it through the number of densely connected layers to compute the class of each pixel. In this work, we have used an artificial neural network (ANN) that is a sub-class of deep neural networks (DNN) [86], [87].

4 Proposed scheme

In this chapter we will be discussing the proposed scheme for semantic segmentation of natural images. The proposed scheme classifies pixels of a given image into three different classes that are sky, land, and water. A natural scene usually contains all of these different classes, and it is really important to segment them properly for a better understanding of the natural scenes. The overall work can be seen in the following block diagram in Figure 1.

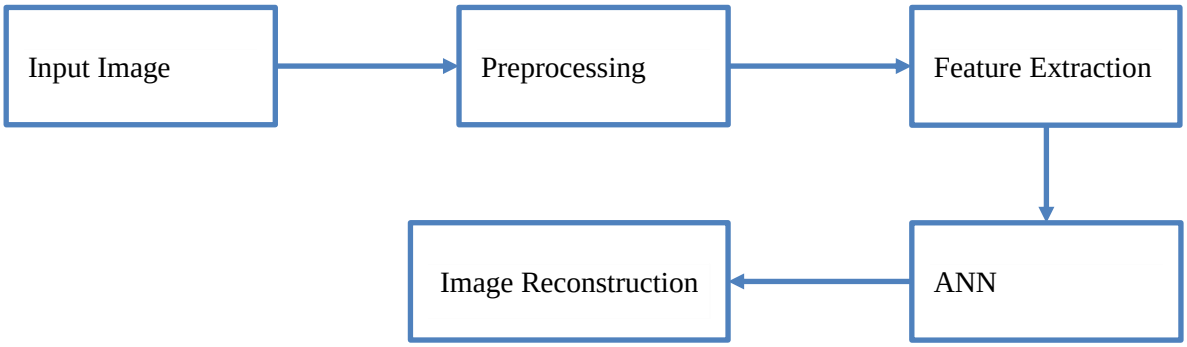

Figure 1. Proposed Technique's Block Diagram

The image is first passed through a preprocessing stage. This stage not only splits the picture into a non-overlapping window size of 5 X 5 but also the label for each window is saved. In this work, we have divided each image into three different classes that are each pixel or more precisely each window of pixels is assigned one of the labels. We have used label 1 for the water label 2 is used for land, and for sky label 3 is used. Once each image chunk is preprocessed these image windows are passed through a feature extraction stage. The feature extraction stage calculates different features that include the textural and statistical features. The overall feature that are extracted in this stage include Contrast (from grayscale image), Correlation (from grayscale image), Energy (from grayscale image), Homogeneity (from grayscale image), Hue variation (from HSV image), Saturation variation (from HSV image), Value (intensity) variation (from HSV image), Hue standard deviation (from HSV image), Saturation standard deviation (from HSV image), Value (intensity) standard deviation (from HSV image), Hue Mean (from HSV image), Saturation Mean (from HSV image), Value (intensity) Mean (from HSV image), Hue skewness (from HSV image), Saturation, skewness (from HSV image) and Value (intensity) skewness (from HSV image).

Extracted features and the assigned label collectively generate the data, ground truth labels pair. This data is divided into testing and training data. In the training phase, the training data (i.e. the extracted features from the 5X5 window) is given to a feed-forward Artificial Neural Network then trains the model for these 5X5 window data. The error is calculated between random prediction and ground truth labels. The model weights are tuned using this calculated error. Once the learning of the artificial neural network is completed. The model is used to predict the unseen testing data.

The Neural architecture used in this work takes a batch size of 16 whereas, the cross-entropy loss is utilized for model. A scaled conjugate gradient algorithm is used to train the model on cross-entropy loss. Below is the block diagram of the neural network model shown in Figure 2.

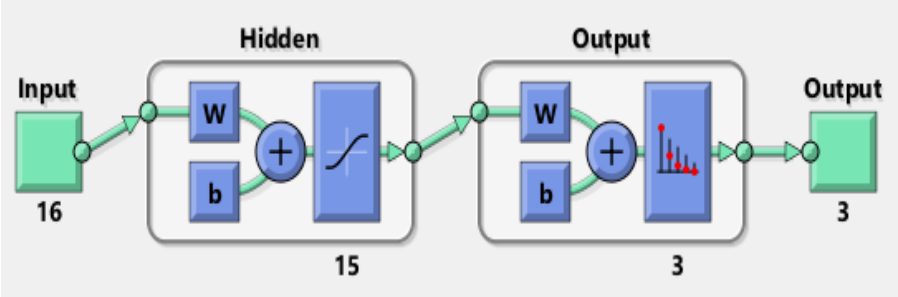

Figure 2. Proposed ANN model's Block diagram.

The overall processing is summarized as follows:

Step 1: Image preprocessing (i.e., dividing the image into 5X5 windows and assigning labels to each window and data balancing).

Step 2: Calculate textural and statistical features from each window.

Step 3: Training an artificial neural network for classifying image windows features.

Step 4: After the training model, the trained model is used to make predictions on each window and finally the windows are joined together to reconstruct a segmented image.

The Graphical Use interface is designed in MATLAB that make this application user friendly. In this interface we can load data, extract the features that are saved as a mat file. We can also perform the data balancing so that each class contains approximately same number of samples for training the model to avoid any biasness. We may also train the model given the features extracted or we may also test on a given test data. The GUI is shown in fig. 3.

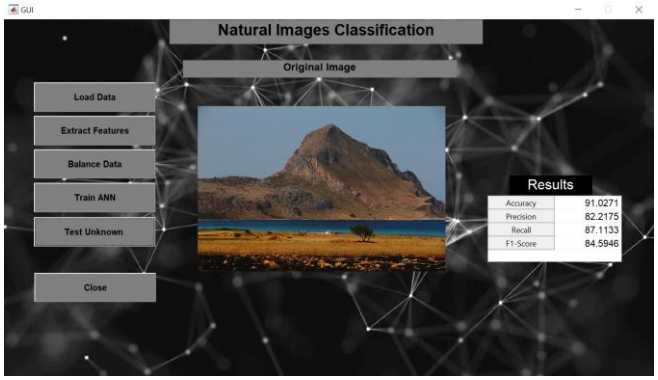

Figure 3. Graphical User Interface.

4.1 Dataset

The data used in this work is gathered from various internet sources. A few Sample training images for each class are shown in the following Figure 3. Where images in the first column of Figure 3 depict the training pictures for the water class, second column images depicts the training pictures for land class and third column images in Figure 3 represents the training images for sky class. Where all of these images are divided into 5x5 block images and features are extracted from each block to compile a set of features and class labels. Although the training images contain only one class however, the test image contains all three classes in an image. The sample test images, and their respective ground truth are shown in the Figure 4.

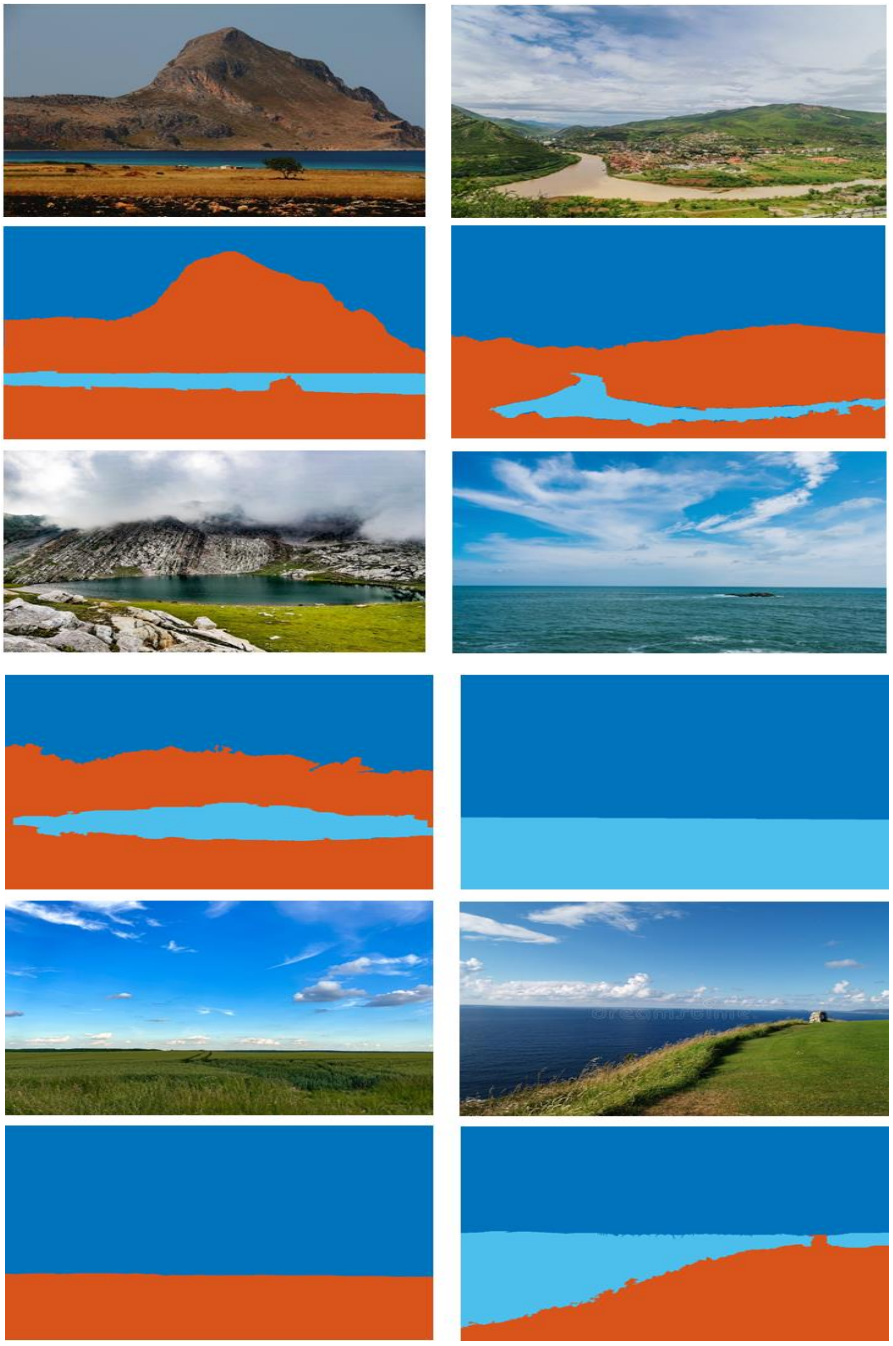

Figure 4. Test images and their respective ground truths.

4.2 Comparison parameters

In this section we will be discussing the comparison parameters that we have used to evaluate the proposed scheme. These parameters include AUC-ROC curve, Cross Entropy loss and Confusion matrix.

4.2.1 AUC-ROC Curve

While working with the classification problems in machine learning, the AUC-ROC curve [88] is widely used as a performance parameter where, AUC is a separability measure. It describes the ability of a model to differentiate between classes. If the value of AUC increases the accuracy increases so the probability that the class 0 will be classified as 0 and 1 as 1 increase. The ROC curve is plotted such that False Positive Rate (FPR) [89] is the domain with True Positive Rate (TPR) [90] is the range.

The True positive rate or TPR is the percentage of properly categorized positive instances, such as feature vectors of malicious apps [91], while the False positive rate or FPR is the percentage of wrongly diagnosed negative cases such as feature vectors of benign apps. Mathematically TPR and FPR can be calculated as equation (8) and equation (9), respectively.

$$TPR = \frac{TP}{TP+FN} \tag{8}$$

$$FPR = \frac{FP}{FP+TN} \tag{9}$$

Where TP shows the true positive, FN represents the false negative, FP depicts the false positive and TN denotes the true negative. The Graphical representation is shown below in Figure 5.

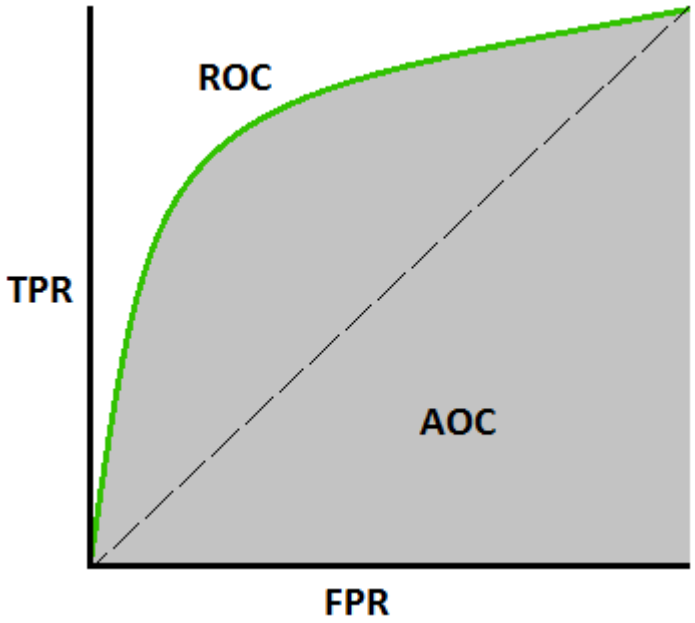

Figure 5. AUC-ROC Curve

4.2.2 Cross Entropy Loss

Cross-entropy loss also called log loss [92], logarithmic loss or logistic loss. Based on how far a predicted class probability differs from the actual expected value [93], a score/loss is calculated to penalize it. For differences close to 1, the penalty is large, and for small differences, the penalty is small [94]. During training, model weights are adjusted according to cross-entropy loss. Ideally, the

loss should be minimized, as the loss is less, the model will be better. A perfect model has zero cross-entropy loss [95]. Mathematically it can be depicted as equation (10) below.

$$L_{ce} = -\sum_{i=1}^{n} t_i \log(p_i) \;\; for\; n\; number\; of\; classes \tag{10}$$

Where, t_i denotes the truth label and p_i shows the $ith\ class$' SoftMax probability.

4.2.3 Confusion matrix

Confusion matrices are used to define a classification algorithm's performance [96]. An algorithm's confusion matrix visualizes and summarizes its performance [97]. For instance, the following table in Figure 18 depicts an example confusion matrix for the classification of normal vs. cancer images in which benign tissue is referred to as healthy and marked as class 1 while malignant tissue is referred to as cancerous and marked as class 2.

4.2.4 Simulation Results

After the collection of features dataset which includes different features attributes and class labels form 5x5 blocks of all training images, the features data is balanced out by keeping equal number of 5x5 samples for each class. Further, the data is divided into training, validation and testing. The model is the trained for total 663 epochs or iterations where in each iteration the loss is reduced. In other words, the accuracy is increasing while training the model. The overall loss for the trained model is shown in the following graph in Figure 6.

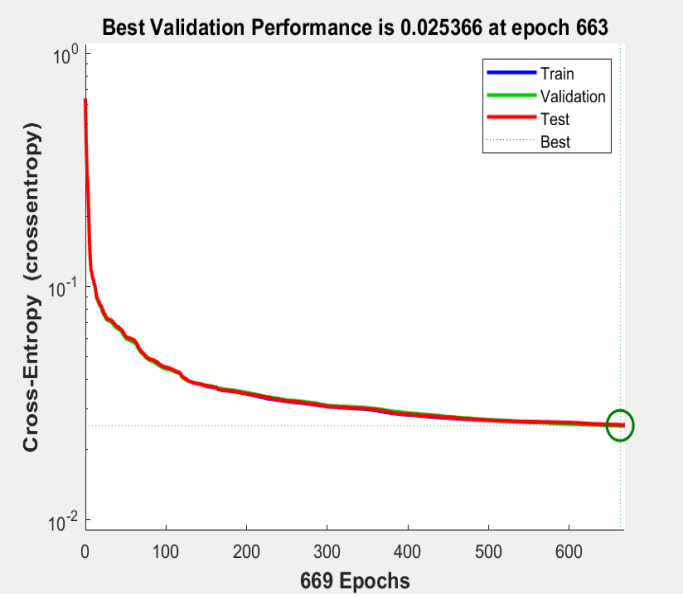

Figure 6. Cross-Entropy Loss Vs the number of Epochs.

Figure 20 shows the confusion matrix of trained work. Where Top Left confusion matrix shows the confusion matrix for training data whereas, the validation confusion matrix is presented in Top Right. Bottom Left confusion matrix in Figure 7 describes the testing confusion matrix and the overall confusion matrix is presented in Bottom Right of Figure 6. It can clearly be seen that in all cases, and Accuracy of more than 97.5 % is achieved.

Figure 7. The confusion matrix

Figure 8 depicts the graph between false negative rate and the true positive rate for test, train, validation and collectively for all data. This graph is known as receiver operating characteristic curve (ROC). All the curves are inclined more towards top-left corner of the graph which shows that the performance for each class is good.

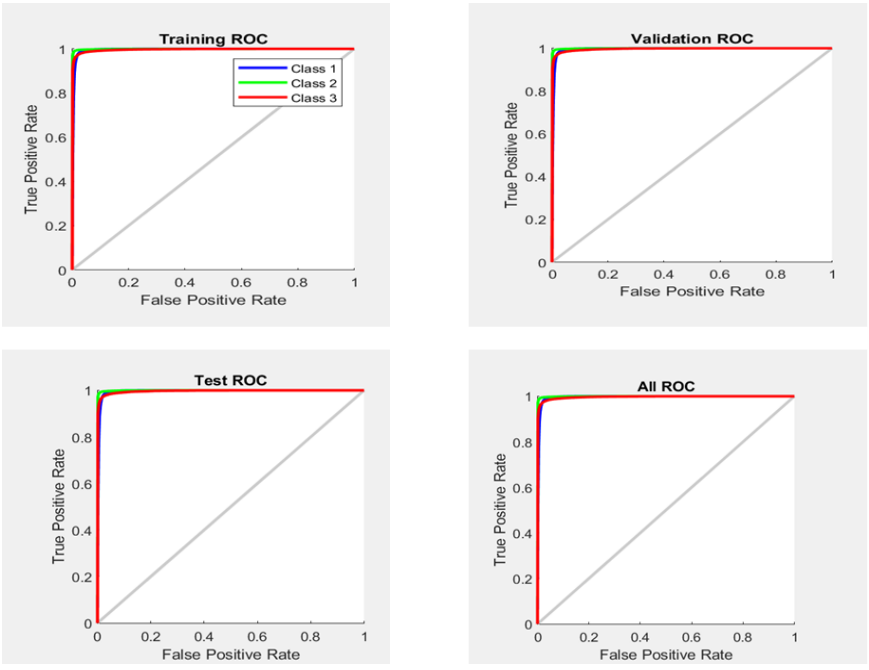

Figure 8. The Roc curve for Training, validation, Test and Overall data

After training and validation of model, the trained ANN model is saved and is used for testing of unknown test images shown in Figure 16. The test image is also going to be divided into 5x5 blocks and the feature values are to be extracted from each block and then each block will be tested through the trained model which will predict a class label for each block. After testing all the blocks and collecting the predicted labels, we have assigned the color values to all blocks pixels and created a predicted image which can be visually compared with the already manually prepared ground truth images to assess the visual similarity between the predicted and actual ground truth images. The testing of unknown images also provides a confusion matrix and ROC curve for the numerical performance evaluation. Figure 9-11 shows the testing done on three unknown images where the first image in each figure shows the actual test image, second image is the actual ground truth image for that test image, third image is the predicted image after testing which can be compared with ground truth image visually to see the performance of trained model in case of an unknown image testing. All three images show good, predicted images which shows that the trained model is performing well. The same results are confirmed from fourth and fifth images in each Figure in terms of numerical results from confusion matrices and ROC plots where the Accuracy for the unknown tested images are 91.0%, 94.2% and 96.8% which shows that the trained model is performing well on unknown images.

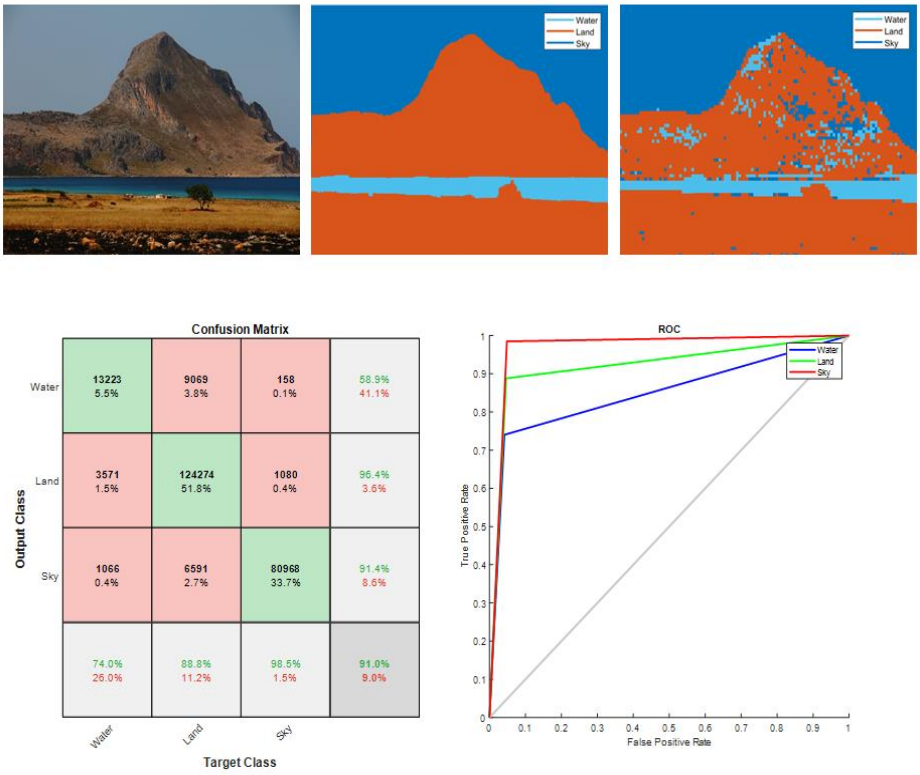

Figure 9. Test Image 1 Results.

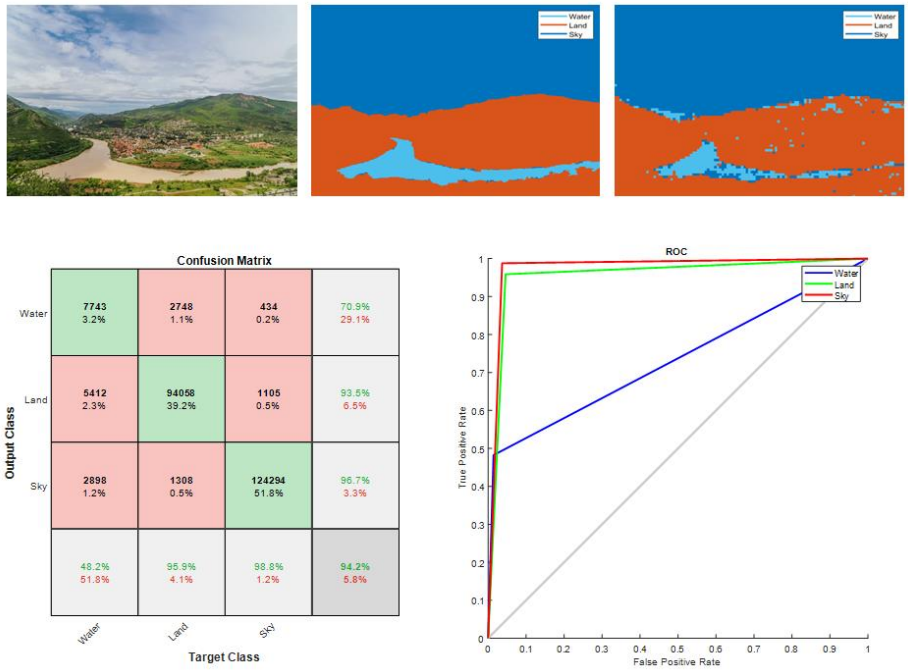

Figure 10. Test Image 2 Results.

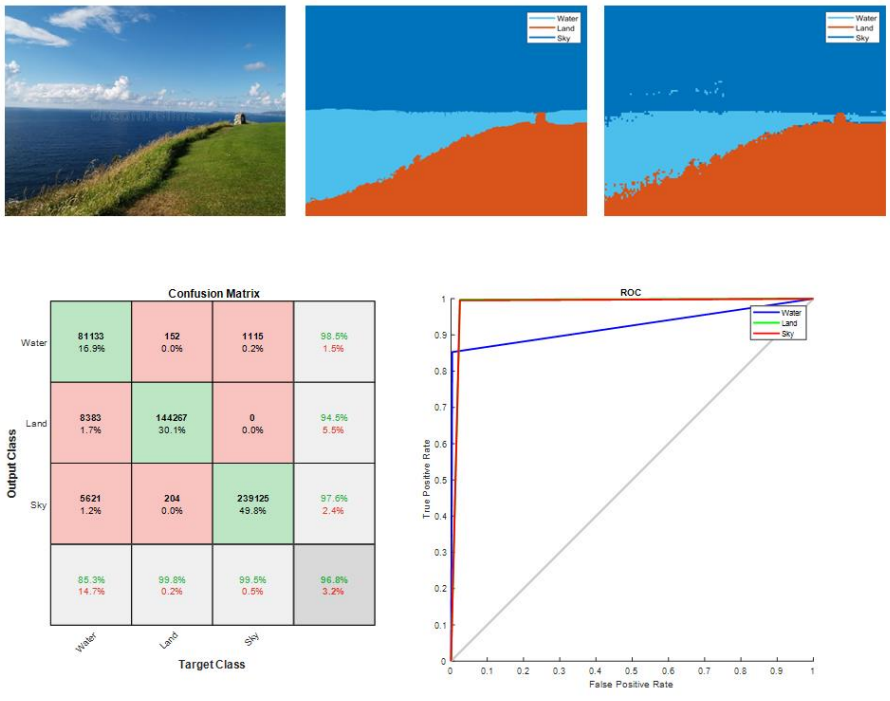

Figure 11. Test Image 3 Results.

5 Conclusions

In this project, we developed a machine learning model for semantic segmentation of natural scene images, specifically targeting the segmentation of sky, land, and water. The model used 5x5 windows to extract features from the input images, which were then passed through a neural network for training. Our trained model achieved an accuracy of 97.5% and required fewer floating-point operations compared to convolutional networks. To train the model, we used a dataset of images with only one class present, which simplifies the labelling process. However, to further improve the complexity and performance of the model, we suggest adding additional classes such as birds, mountains, and grass. Additionally, modifying the neural network architecture may also be beneficial in maintaining accuracy with the increased number of classes. Exploring other feature extraction methods or feature selection techniques could also potentially improve the model's performance. It is also important to test the model on a diverse range of natural scene images to ensure its robustness and generalizability. A graphical user interface was also developed to facilitate inputting images, extracting features, training the model, and evaluating performance on unseen data.

References

[1] A. I. Khan and S. Al-Habsi, "Machine Learning in Computer Vision," in *Procedia Computer Science*, 2020, vol. 167, pp. 1444–1451. doi: 10.1016/j.procs.2020.03.355.

[2] P. A. Kragel, M. C. Reddan, K. S. LaBar, and T. D. Wager, "Emotion schemas are embedded in the human visual system," *Sci Adv*, vol. 5, no. 7, 2019, doi: 10.1126/sciadv.aaw4358.

[3] L. N. Ayton *et al.*, "An update on retinal prostheses," *Clinical Neurophysiology*, vol. 131, no. 6. pp. 1383–1398, 2020. doi: 10.1016/j.clinph.2019.11.029.

[4] S. Taghizadegan, *Essentials of Lean Six Sigma*. 2006. doi: 10.1016/B978-0-12-370502-0.X5000-0.

[5] N. V. Gavrilovskaya, V. P. Kuvaldin, I. S. Zlobina, D. E. Lomakin, and E. E. Suchkova, "Developing a robot with computer vision for automating business processes of the industrial complex," in *Journal of Physics: Conference Series*, 2021, vol. 1889, no. 2. doi: 10.1088/1742-6596/1889/2/022024.

[6] "What is Computer Vision? | IBM."

[7] J. Casadesús, D. V.-J. of integrative plant biology, and undefined 2014, "Conventional digital cameras as a tool for assessing leaf area index and biomass for cereal breeding," *Wiley Online Library*.

[8] G. Meyer, J. N.-C. and electronics in agriculture, and undefined 2008, "Verification of color vegetation indices for automated crop imaging applications," *Elsevier*.

[9] E. Hamuda, M. Glavin, E. J.-C. and electronics in agriculture, and undefined 2016, "A survey of image processing techniques for plant extraction and segmentation in the field," *Elsevier*.

[10] Z. Li *et al.*, "Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors," *Elsevier*.

[11] A. Prest, C. Leistner, J. Civera, … C. S.-2012 I. C., and undefined 2012, "Learning object class detectors from weakly annotated video," *ieeexplore.ieee.org*.

[12] H. Deng, D. C.-P. recognition, and undefined 2004, "Unsupervised image segmentation using a simple MRF model with a new implementation scheme," *Elsevier*.

[13] M. Helbich, Y. Yao, Y. Liu, J. Zhang, P. L.-E. international, and undefined 2019, "Using deep learning to examine street view green and blue spaces and their associations with geriatric depression in Beijing, China," *Elsevier*.

[14] Y. Sun, X. Wang, J. Zhu, L. Chen, … Y. J.-S. of T. T., and undefined 2021, "Using machine learning to examine street green space types at a high spatial resolution: Application in Los Angeles County on socioeconomic disparities in," *Elsevier*.

[15] L. Dai, C. Zheng, Z. Dong, Y. Yao, … R. W.-C. and E., and undefined 2021, "Analyzing the correlation between visual space and residents' psychology in Wuhan, China using street-view images and deep-learning technique," *Elsevier*.

[16] H. Zhang, B. Chen, Z. Sun, Z. B.-U. F. & U. Greening, and undefined 2013, "Landscape perception and recreation needs in urban green space in Fuyang, Hangzhou, China," *Elsevier*.

[17] N. Kulkarni, "Color Thresholding Method for Image Segmentation of Natural Images," *International Journal of Image, Graphics and Signal Processing*, vol. 4, no. 1, pp. 28–34, 2012, doi: 10.5815/ijigsp.2012.01.04.

[18] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez, "Road scene segmentation from a single image," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7578 LNCS, no. PART 7, pp. 376–389, 2012, doi: 10.1007/978-3-642-33786-4_28.

[19] P. Sadeghi-Tehran, N. Virlet, K. Sabermanesh, and M. J. Hawkesford, "Multi-feature machine learning model for automatic segmentation of green fractional vegetation cover for high-throughput field phenotyping," *Plant Methods*, vol. 13, no. 1, Nov. 2017, doi: 10.1186/S13007-017-0253-8.

[20] R. Zhang, S. Candra, K. V.-2015 I. international, and undefined 2015, "Sensor fusion for semantic segmentation of urban scenes," *ieeexplore.ieee.org*.

[21] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," *Lecture Notes in Computer Science*, vol. 5302 LNCS, no. PART 1, pp. 44–57, 2008, doi: 10.1007/978-3-540-88682-2_5.

[22] T. Rehman, M. Mahmud, Y. Chang, … J. J.-… and electronics in, and undefined 2019, "Current and future applications of statistical machine learning algorithms for agricultural machine vision systems," *Elsevier*.

[23] D. J. Bora, A. K. Gupta, and F. A. Khan, "Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation," Jun. 2015.

[24] S. Khan *et al.*, "Human Action Recognition: A Paradigm of Best Deep Learning Features Selection and Serial Based Extended Fusion," *mdpi.com*, 2021, doi: 10.3390/s21237941.

[25] J. Taipalmaa, … N. P.-2020 I. international, and undefined 2020, "Different color spaces in deep learning-based water segmentation for autonomous marine operations," *ieeexplore.ieee.org*.

[26] Dr. S. Manoharan, "Performance Analysis of Clustering Based Image Segmentation Techniques," *Journal of Innovative Image Processing*, vol. 2, no. 1, pp. 14–24, 2020, doi: 10.36548/jiip.2020.1.002.

[27] R. Yogamangalam and B. Karthikeyan, "Segmentation techniques comparison in image processing," *International Journal of Eng. and Technology*, vol. 5, no. 1, pp. 307–313, 2013.

[28] N. Sharma *et al.*, "Automated medical image segmentation techniques," *Journal of Medical Physics*, vol. 35, no. 1. pp. 3–14, 2010. doi: 10.4103/0971-6203.58777.

[29] H. J. He, C. Zheng, and D. W. Sun, "Image Segmentation Techniques," in *Computer Vision Technology for Food Quality Evaluation: Second Edition*, 2016, pp. 45–63. doi: 10.1016/B978-0-12-802232-0.00002-5.

[30] E. Akbal, "An automated environmental sound classification methods based on statistical and textural feature," *Applied Acoustics*, vol. 167, 2020, doi: 10.1016/j.apacoust.2020.107413.

[31] D. R. Patel, V. Vakharia, and M. B. Kiran, "Texture classification of machined surfaces using image processing and machine learning techniques," *FME Transactions*, vol. 47, no. 4, pp. 865–872, 2019, doi: 10.5937/fmet1904865P.

[32] M. R. Rudnick, S. Goldfarb, and J. Tumlin, "Contrast-induced nephropathy: Is the picture any clearer?," *Clinical Journal of the American Society of Nephrology*, vol. 3, no. 1, pp. 261–262, 2008, doi: 10.2215/CJN.04951107.

[33] K. Hasikin and N. A. Mat Isa, "Adaptive fuzzy contrast factor enhancement technique for low contrast and nonuniform illumination images," *Signal Image Video Process*, vol. 8, no. 8, pp. 1591–1603, 2014, doi: 10.1007/s11760-012-0398-x.

[34] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1998, vol. 1496, pp. 1115–1124. doi: 10.1007/bfb0056301.

[35] M. Van Heel, "Similarity measures between images," *Ultramicroscopy*, vol. 21, no. 1, pp. 95–100, 1987, doi: 10.1016/0304-3991(87)90010-6.

[36] T. Wei, J. Wei, K. Zhang, H. Zhao, and L. Zhang, "Grayscale image recording on Ge2Sb2Te5 thin films through laser-induced structural evolution," *Sci Rep*, vol. 7, 2017, doi: 10.1038/srep42712.

[37] R. K. Mohammed and H. A. Abdullah, "Implementation of digital and analog modulation systems using FPGA," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 485–493, 2020, doi: 10.11591/ijeecs.v18.i1.pp485-493.

[38] Y. Asaoka *et al.*, "29.1: Polarizer-Free Reflective LCD Combined with Ultra Low-Power Driving Technology," *SID Symposium Digest of Technical Papers*, vol. 40, no. 1, p. 395, 2009, doi: 10.1889/1.3256797.

[39] E. Y. Lin, A. Rahmawati, J. H. Ko, and J. C. Liu, "Extraction of yttrium and europium from waste cathode-ray tube (CRT) phosphor by subcritical water," *Sep Purif Technol*, vol. 192, pp. 166–175, 2018, doi: 10.1016/j.seppur.2017.10.004.

[40] S. Hwang, Y. Cheon, S. Han, I. Jang, and J. W. Shin, "Enhancement of Coded Speech Using Neural Network-Based Side Information," *IEEE Access*, vol. 9, pp. 121532–121540, 2021, doi: 10.1109/ACCESS.2021.3108784.

[41] E. A. Boonstra and H. A. Slagter, "The Dialectics of Free Energy Minimization," *Front Syst Neurosci*, vol. 13, 2019, doi: 10.3389/fnsys.2019.00042.

[42] A. Frost, "Overvaluing uniformity," *Virginia Law Review*, vol. 94, no. 7. pp. 1567–1639, 2008.

[43] S. Van Der Walt *et al.*, "Scikit-image: Image processing in python," *PeerJ*, vol. 2014, no. 1, 2014, doi: 10.7717/peerj.453.

[44] M. Tixier, C. Cian, P. A. Barraud, R. Laboissiere, and S. Rousset, "The interaction between long-term memory and postural control: Different effects of episodic and semantic tasks," *Motor Control*, vol. 25, no. 2, pp. 182–197, 2021, doi: 10.1123/MC.2020-0061.

[45] G. Singh and M. A. Ansari, "Efficient detection of brain tumor from MRIs using K-means segmentation and normalized histogram," in *India International Conference on Information Processing, IICIP 2016 - Proceedings*, 2017. doi: 10.1109/IICIP.2016.7975365.

[46] Y. Kinoshita and H. Kiya, "Hue-correction scheme based on constant-hue plane for deep-learning-based color-image enhancement," *IEEE Access*, vol. 8, pp. 9540–9550, 2020, doi: 10.1109/ACCESS.2020.2964823.

[47] M. J. Pearson *et al.*, "Evidence of Intrinsic Impairment of Osteoblast Phenotype at the Curve Apex in Girls With Adolescent Idiopathic Scoliosis," *Spine Deform*, vol. 7, no. 4, pp. 533–542, 2019, doi: 10.1016/j.jspd.2018.11.016.

[48] S. Kunz, S. Haasova, and A. Florack, "Fifty shades of food: The influence of package color saturation on health and taste in consumer judgments," *Psychol Mark*, vol. 37, no. 7, pp. 900–912, 2020, doi: 10.1002/mar.21317.

[49] J. K. Patil and R. Kumar, "Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features," *Engineering in Agriculture, Environment and Food*, vol. 10, no. 2, 2017, doi: 10.1016/j.eaef.2016.11.004.

[50] L. Y. Hanger, A. Lotz, and S. Lepeniotis, "Descriptive profiles of selected high intensity sweeteners (HIS), HIS blends, and sucrose," *J Food Sci*, vol. 61, no. 2, pp. 456–459, 1996, doi: 10.1111/j.1365-2621.1996.tb14216.x.

[51] M. Domino *et al.*, "Selection of Image Texture Analysis and Color Model in the Advanced Image Processing of Thermal Images of Horses following Exercise," *Animals*, vol. 12, no. 4, 2022, doi: 10.3390/ani12040444.

[52] S. Pearl and A. Dorothy, "Converting RGB to HSV," *Mechatronics*, 2022.

[53] B. Robot, "Hsv_Writeup," *Mechatronics*, 2005.

[54] P. M. Nishad and R. Manicka Chezian, "Various Colour Spaces and Colour Space Conversion Algorithms," *Journal of Global Research in Computer Science*, vol. 4, no. 1, pp. 44–48, 2013.

[55] L. B. Wexner, "The degree to which colors (hues) are associated with mood-tones," *Journal of Applied Psychology*, vol. 38, no. 6, pp. 432–435, 1954, doi: 10.1037/h0062181.

[56] M. M. Hennink, B. N. Kaiser, and M. B. Weber, "What Influences Saturation? Estimating Sample Sizes in Focus Group Research," *Qual Health Res*, vol. 29, no. 10, pp. 1483–1496, 2019, doi: 10.1177/1049732318821692.

[57] R. R. Burton and J. S. Brown, "An investigation of computer coaching for informal learning activities," *Int J Man Mach Stud*, vol. 11, no. 1, pp. 5–24, 1979, doi: 10.1016/S0020-7373(79)80003-6.

[58] S. N. Utami, B. Kiswanjaya, S. I. Syahraini, and P. Ustriyana, "Tolerance limit value of brightness and contrast adjustment on digitized radiographs," in *Journal of Physics: Conference Series*, 2017, vol. 884, no. 1. doi: 10.1088/1742-6596/884/1/012052.

[59] B. Voisiat, W. Wang, M. Holzhey, and A. F. Lasagni, "Improving the homogeneity of diffraction based colours by fabricating periodic patterns with gradient spatial period using Direct Laser Interference Patterning," *Sci Rep*, vol. 9, no. 1, 2019, doi: 10.1038/s41598-019-44212-4.

[60] I. Kurniastuti, E. N. I. Yuliati, F. Yudianto, and T. D. Wulan, "Determination of Hue Saturation Value (HSV) color feature in kidney histology image," in *Journal of Physics: Conference Series*, 2022, vol. 2157, no. 1. doi: 10.1088/1742-6596/2157/1/012020.

[61] L. L. P. Nguyen, L. Baranyai, D. Nagy, P. V. Mahajan, V. Zsom-Muha, and T. Zsom, "Color analysis of horticultural produces using hue spectra fingerprinting," *MethodsX*, vol. 8, 2021, doi: 10.1016/j.mex.2021.101594.

[62] Y. P. Huang, T. H. Wang, and H. Basanta, "Using Fuzzy Mask R-CNN Model to Automatically Identify Tomato Ripeness," *IEEE Access*, vol. 8, pp. 207672–207682, 2020, doi: 10.1109/ACCESS.2020.3038184.

[63] C. E. Widodo, K. Adi, and R. Gernowo, "Medical image processing using python and open cv," in *Journal of Physics: Conference Series*, 2020, vol. 1524, no. 1. doi: 10.1088/1742-6596/1524/1/012003.

[64] A. Bhatnagar, R. Patel, M. Gupta, M. Pal, and L. Kumar, "Customized sorting and packaging machine," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 4, pp. 1326–1333, 2021, doi: 10.12928/TELKOMNIKA.v19i4.16786.

[65] X. Wang, C. Song, C. Yang, and Y. Xie, "Process working condition recognition based on the fusion of morphological and pixel set features of froth for froth flotation," *Miner Eng*, vol. 128, pp. 17–26, 2018, doi: 10.1016/j.mineng.2018.08.017.

[66] "Linda G. Shapiro and George C. Stockman (2001):... - Google Scholar."

[67] L. Barghout, 543 L Lee - US Patent App. 10/618, and undefined 2004, "Perceptual information processing system," *Google Patents*, no. 10.

[68] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 7, pp. 3523–3542, 2022, doi: 10.1109/TPAMI.2021.3059968.

[69] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[70] S. Khan *et al.*, "Human Action Recognition: A Paradigm of Best Deep Learning Features Selection and Serial Based Extended Fusion," *mdpi.com*, 2021, doi: 10.3390/s21237941.

[71] A. Hafiz, G. B.-I. journal of multimedia information, and undefined 2020, "A survey on instance segmentation: state of the art," *Springer*.

[72] Y. Yi, M. L.-P. Recognition, and undefined 2016, "Human action recognition with graph-based multiple-instance learning," *Elsevier*.

[73] S. Liu, L. Qi, H. Qin, J. Shi, J. J.-P. of the IEEE, and undefined 2018, "Path aggregation network for instance segmentation," *openaccess.thecvf.com*.

[74] B. Romera-Paredes and P. H. S. Torr, "Recurrent instance segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9910, pp. 312–329, 2016, doi: 10.1007/978-3-319-46466-4_19.

[75] N. Gao, Y. Shan, Y. Wang, X. Zhao, and K. Huang, "SSAP: Single-Shot Instance Segmentation with Affinity Pyramid," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 661–673, 2021, doi: 10.1109/TCSVT.2020.2985420.

[76] R. Fan, M. M. Cheng, Q. Hou, T. J. Mu, J. Wang, and S. M. Hu, "S4Net: Single stage salient-instance segmentation," *Comput Vis Media (Beijing)*, vol. 6, no. 2, pp. 191–204, 2020, doi: 10.1007/s41095-020-0173-9.

[77] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "YOLACT++ Better Real-Time Instance Segmentation," *IEEE Trans Pattern Anal Mach Intell*, vol. 44, no. 2, pp. 1108–1121, 2022, doi: 10.1109/TPAMI.2020.3014297.

[78] I. Papadeas, L. Tsochatzidis, A. Amanatiadis, and I. Pratikakis, "Real-time semantic image segmentation with deep learning for autonomous driving: A survey," *Applied Sciences (Switzerland)*, vol. 11, no. 19, 2021, doi: 10.3390/app11198802.

[79] S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: a review," *Artif Intell Rev*, vol. 54, no. 1, pp. 137–178, 2021, doi: 10.1007/s10462-020-09854-1.

[80] X. Liu, Z. Deng, and Y. Yang, "Recent progress in semantic image segmentation," *Artif Intell Rev*, vol. 52, no. 2, pp. 1089–1106, 2019, doi: 10.1007/s10462-018-9641-3.

[81] O. L. F. de Carvalho *et al.*, "Panoptic Segmentation Meets Remote Sensing," *Remote Sens (Basel)*, vol. 14, no. 4, 2022, doi: 10.3390/rs14040965.

[82] Q. Chen, A. Cheng, X. He, P. Wang, and J. Cheng, "SpatialFlow: Bridging All Tasks for Panoptic Segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 6, pp. 2288–2300, 2021, doi: 10.1109/TCSVT.2020.3020257.

[83] R. Mohan and A. Valada, "EfficientPS: Efficient Panoptic Segmentation," *Int J Comput Vis*, vol. 129, no. 5, pp. 1551–1579, 2021, doi: 10.1007/s11263-021-01445-z.

[84] H. Gouk, E. Frank, B. Pfahringer, and M. J. Cree, "Regularisation of neural networks by enforcing Lipschitz continuity," *Mach Learn*, vol. 110, no. 2, pp. 393–416, 2021, doi: 10.1007/s10994-020-05929-w.

[85] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A Comprehensive Survey on Graph Neural Networks," *IEEE Trans Neural Netw Learn Syst*, vol. 32, no. 1, pp. 4–24, 2021, doi: 10.1109/TNNLS.2020.2978386.

[86] Y. V. R. Nagapawan, K. B. Prakash, and G. R. Kanagachidambaresan, "Convolutional Neural Network," in *EAI/Springer Innovations in Communication and Computing*, 2021, pp. 45–51. doi: 10.1007/978-3-030-57077-4_6.

[87] S. Lyu and J. Liu, "Convolutional recurrent neural networks for text classification," *Journal of Database Management*, vol. 32, no. 4, pp. 65–82, 2021, doi: 10.4018/JDM.2021100105.

[88] Sarang Narkhede, "Understanding AUC - ROC Curve," *Towards Data Science*, pp. 6–11, 2019.

[89] M. Latah and L. Toker, "Minimizing false positive rate for DoS attack detection: A hybrid SDN-based approach," *ICT Express*, vol. 6, no. 2, pp. 125–127, 2020, doi: 10.1016/j.icte.2019.11.002.

[90] C. S. Hong and T. G. Oh, "TPR-TNR plot for confusion matrix," *Commun Stat Appl Methods*, vol. 28, no. 2, pp. 161–169, 2021, doi: 10.29220/CSAM.2021.28.2.161.

[91] J. N. Jannath and M. S. B. S, "Detection of malicious Android applications using Ontology-based intelligent model in mobile cloud environment," *Journal of Information Security and Applications*, vol. 58, 2021, doi: 10.1016/j.jisa.2021.102751.

[92] X. Zeng, Y. Zhang, X. Wang, K. Chen, D. Li, and W. Yang, "Fine-Grained Image Retrieval via Piecewise Cross Entropy loss," *Image Vis Comput*, vol. 93, 2020, doi: 10.1016/j.imavis.2019.10.006.

[93] J. Lu and S. Steinerberger, "Neural collapse under cross-entropy loss," *Appl Comput Harmon Anal*, vol. 59, pp. 224–241, 2022, doi: 10.1016/j.acha.2021.12.011.

[94] X. Zeng, Y. Zhang, X. Wang, K. Chen, D. Li, and W. Yang, "Fine-Grained Image Retrieval via Piecewise Cross Entropy loss," *Image Vis Comput*, vol. 93, 2020, doi: 10.1016/j.imavis.2019.10.006.

[95] M. Martinez and R. Stiefelhagen, "Taming the Cross Entropy Loss," in *Lecture Notes in Computer Science*, 2019, vol. 11269 LNCS, pp. 628–637. doi: 10.1007/978-3-030-12939-2_43.

[96] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, "Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking," *IEEE Access*, vol. 8, pp. 90847–90861, 2020, doi: 10.1109/ACCESS.2020.2994222.

[97] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Inf Sci (N Y)*, vol. 507, pp. 772–794, 2020, doi: 10.1016/j.ins.2019.06.064.