MIS 400 MANAGEMENT INFORMATION SYSTEMS

DESIGN PROJECT

GEOTHERMAL IMAGE CLASSIFICATION

Project Team:

Engin TEPE 20161708021 (MIS)

Berna DERYA 20161708040 (MIS)

Erkan ÖZDEMİR 20172305032 (MIS)

Project Advisor: Dr. Mahmut ÇAVUR

Faculty of Management

Kadir Has University

Spring 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

ACKNOWLEDGEMENTS

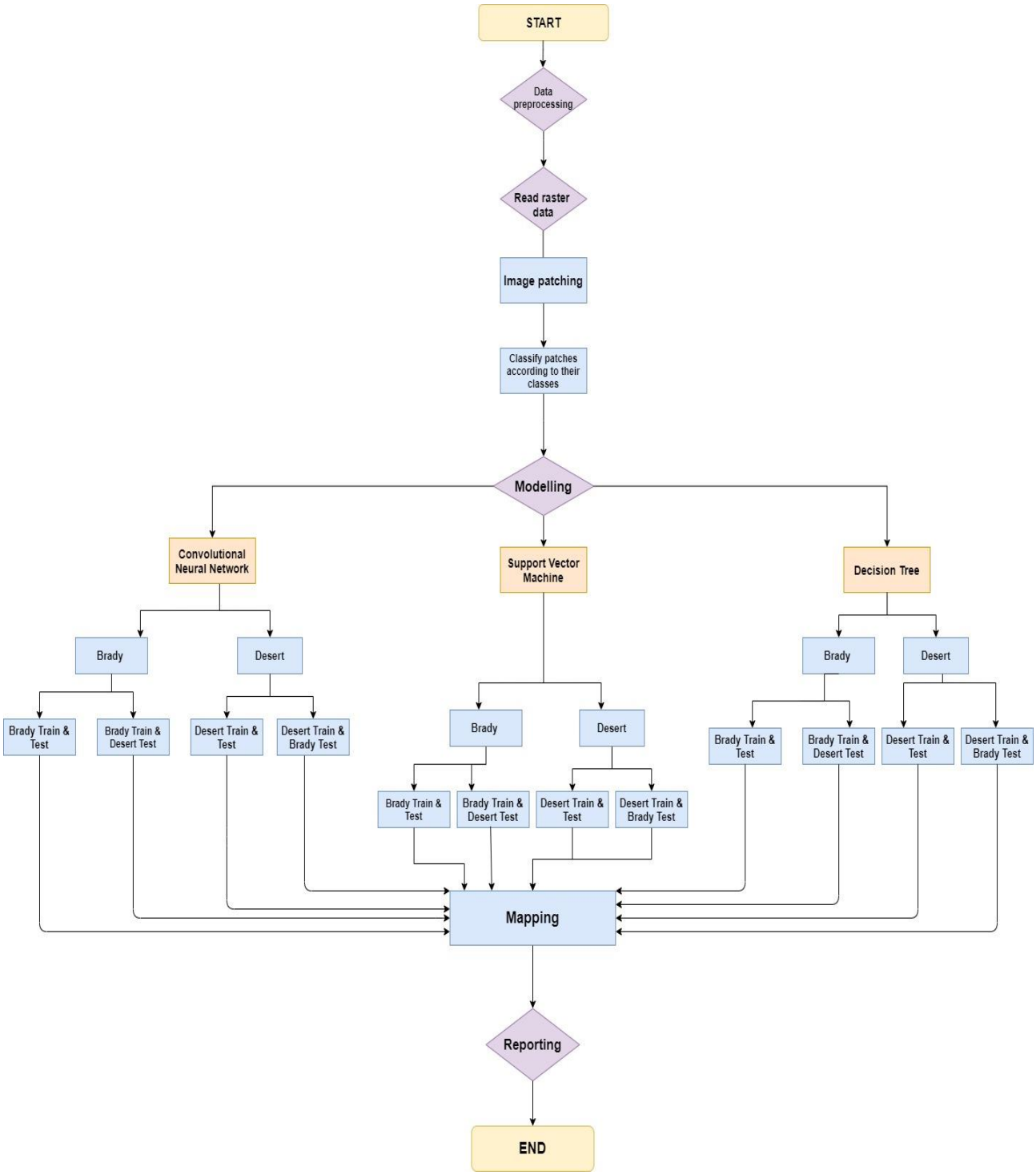
We would like to thank Dr. Mahmut Çavur who shared his knowledge and experiences with us in the planning, research, execution, and creation of this research. And we would like to thank Research Assistant Aykut Çayır who did his best to help us. Also, we would like to thanks the rector of Kadir Has University Prof. Dr. Sondan Durukanoğlu Feyiz, Head of Management Information Systems Department Doç. Dr. Mehmet Nafiz Aydın for letting us use all facilities within the university.

ABSTRACT

Increasing energy demands in the world and encouraging the use of local resources have increased the importance of alternative energy sources. Geothermal energy has grown in popularity in recent years, so the need to explore and study geothermal energy sources has also increased. Investigation of geothermal resources for studies involves the management of many uncertainties that make decisions such as investment and operation difficult. Artificial Intelligence (AI), Machine Learning (ML), and Remote Sensing (RS) all have the potential to manage the challenges of geothermal exploration. In this article, we present a methodology that integrates AI, RS, and ML to create an initial assessment of geothermal potential, referring to known indicators of geothermal fields i.e. surface temperature, defects, mineral alteration, and deformation. We have demonstrated the application of the method in two sites that are close to each other but do not have the same characteristics (Brady and Desert geothermal fields).

This study will provide Machine Learning that will reduce the cost and time that its users (private mining firms or government agencies) would spend on-site detection when they need geothermal fields for their work. In this project, the effectiveness of various estimation methods in geothermal field detection on satellite images was reviewed. Our satellite imagery provided by Dr. Mahmut Çavur. Two satellite images named Desert and Brady were used. Different satellite images and different areas can be used. Three different algorithms were used: Convolutional Neural Network (CNN), Support Vector Machine (SVM), and Decision Tree. These algorithms' performance will be evaluated in the next chapters. The aim of this research is to help companies and government agencies save costs and time, and increase their performance efficiency. This model will allow us to identify and examine geothermal areas much more easily. This project will guide companies or governments wishing to establish a geothermal power plant in determining suitable geothermal areas as a basic study for future research and development. Python was used as a tool to perform prediction and ArcGIS Software for mapping. The performance results of each algorithm are presented in this paper. Best results were obtained with the Decision Tree method.

OUR APPROACHES IN THIS PROJECT



ABBREVIATIONS

ANN	Artificial Neural Networks
SVM	Support Vector Machine
CNN	Convolutional Neural Network
DT	Decision Tree
AI	Artificial Intelligence
RS	Remote Sensing
ML	Machine Learning

1 INTRODUCTION

Geothermal energy is heat within the earth. The word geothermal comes from the Greek words geo (earth) and therme (heat). Geothermal energy is a renewable energy source because heat is continuously produced inside the earth. Geothermal resources are hot water reservoirs located below the Earth's surface at varying temperatures and depths. On the other hand, the geothermal field is an area of the Earth characterized by a relatively high flow of heat. Geothermal energy use can be divided into three categories: direct-use applications, geothermal heat pumps, and electric power generation. People use geothermal heat for bathing, to heat buildings, and to generate electricity. [1]

The biggest advantage of geothermal energy is that it is environmental, besides it is known to be 80% cheaper than fossil fuels. Unlike other renewable energy sources such as solar and wind, they are constantly available. Although it is inexpensive, sustainable, and environmentally friendly, geothermal has its disadvantages as well. First, production is limited to areas close to tectonic plate boundaries. Additionally, some places may become cold after decades of use. Though cheaper than fossil fuels once a facility is installed, drilling and exploration of these areas are expensive. This is partly due to the amount of wear that drills and other tools have been subjected to in such aggressive environments. [2] Detection of such sources may be easy in some cases (hot springs) or difficult due to limited signs on the earth. The Discovery of economic resources is expensive and requires expert analysis, observation, and drilling as well as a great deal of time. Areas to explore can be remote. Also, unfortunately, expertise in this area is limited.

Depending on the development and growth rate of the countries, their energy needs are constantly increasing. Today, energy consumption is equated with the level of development. Approximately 90% of the energy consumption in the world is met from coal, oil, and natural gas, which are called fossil fuels. The fact that fossil energy resources will be depleted in the near future and that they create pollution due to the high rate of carbon dioxide they give to the air when they are burned makes it necessary for alternative energy sources to come into play. For this reason, studies on the search and utilization of alternative energy sources that can replace fossil fuels have accelerated in recent years. Geothermal energy is an important alternative energy source because it does not create air pollution with its low carbon dioxide emission rate and is also renewable.[3] Compared to renewable energy sources such as solar and wind, geothermal energy has an advantageous position due to its uninterrupted nature.

Geothermal energy, depending on its temperature, is used in various fields in industry, mainly for electricity generation, heating, and treatment purposes. It is possible to benefit from a high-temperature geothermal fluid in many areas as integrated. Geothermal resources have been used in the world for therapeutic and primitive heating purposes since ancient times. Industrial use was first started in 1904 in Italy by obtaining electricity from geothermal. As of 2000, the geothermal power generation capacity in the world has reached 7974 MWe. According to the data of 2000 in the world, 1.6% of the total electricity production is provided by geothermal energy. The share of other renewable energy sources in world electricity production is 0.6% in wind and 0.05% in sun. Among the countries that generate electricity from geothermal, USA with 2228 MWe, Flipins with 1909 MWe, Italy with 785 MWe, Mexico with 755 MWe, Indonesia with 590 MWe, and Japan with 547 MWe are at the top. Turkey, on the other hand, ranks 15th in the world with its installed power capacity of 20.4 MWe. [4] In recent years, the development of environmental awareness throughout the world and in our country has increased the use of geothermal energy, which is a clean and renewable energy source, especially in urban heating, apart from electricity generation. The direct use capacity of geothermal energy such as heating, cooling, and thermal energy gains importance in the world. [5] Considering all these, geothermal energy has been used in the world for many years and its importance is increasing due to various reasons.

Table 1.1. Direct use of geothermal energy in the world (World Geothermal Congress 2000)

COUNTRIES	INSTALLED CAPACITY MW	PRODUCTION GWh/year
CHINA	2814	8724
JAPAN	1159	7500
USA	5366	5640
ICELAND	1469	5603
TURKEY	820	4377
NEW ZELAND	308	1967
GEORGIA	250	1752

RUSSIA	307	1703
FRANCE	326	1360
HUNGARY	391	1328
SWEDEN	377	1147
MEXICAN	164	1089
ITALY	326	1048
ROMANIA	152	797
SWITZERLAND	547	663

Turkey, on the other hand, is in a rich position among the world countries in terms of geothermal, as it is located on an active tectonic belt due to its geological and geographical location. There are many geothermal resources at different temperatures in the form of around 1000 natural outlets spread all over Turkey. Turkey's geothermal potential is quite high, with 78% of the potential areas in Western Anatolia, 9% in Central Anatolia, 7% in the Marmara Region, 5% in Eastern Anatolia, and 1% in other located in the regions. 90% of Turkey's geothermal resources are low and medium temperature, suitable for direct applications (heating, thermal tourism, various industrial applications, etc.), and 10% is suitable for indirect applications (electricity generation).[6] For these reasons, although the use of geothermal energy in Turkey was not sufficient in the past, the rate of use of geothermal resources has gradually increased with the privatization of geothermal fields since 2008. [7] It is of great importance to benefit from geothermal resources such as Turkey's geological wealth, thermal capacity size, high efficiency, low investment cost, renewable, sustainable, regular, safe, and high quality, all of the products are consumed and the control of the country is under the control of the country. In addition, adverse environmental effects can be prevented and are at very low levels, making geothermal energy an important energy source for Turkey with its successful applications. Considering all these, it is very important to use geothermal energy both in the world and in Turkey.

As the need and demand for geothermal energy increase day by day, the number of geothermal field studies will increase. This will cause private companies and governments that want to do

research to allocate a high budget and time, work with experts in the field, and many other things. The aim of this project is to detect geothermal areas using Artificial Intelligence. Our main purpose is to help private mining companies and government agencies to detect the type of geothermal which also can be used for other metallic mines. Finding a mine or geothermal field for both private mining companies and government agencies entails enormous costs and time. However, thanks to this project, these people will save time and money. They will also be pleased with the result.

1.1 Literature Research

1.1.1. Methods of Geothermal exploration

1.1.1.1. Drilling

Drilling, which is one of the processes of geothermal energy exploration, provides the most accurate and sufficient information during exploration, besides it is one of the most expensive methods. This sufficient and accurate information is provided by Thermal gradient vents, exploration wells, and full-scale production boards. However, it can also be measured after drilling, and again, temperature gradients, thermal pockets and other geothermal properties are used for valuable information.

The underground materials that are generally associated with geothermal fields are types such as limestone chalet, volcanic rocks and granite, and the depth of geothermal exploration fields is approximately 4km maximum.

The overall success rate of a full-scale production dry is 25%, but with the increase of research and analysis of the information obtained, this rate is thought to increase between 60 and 80%. Of course, these studies have a serious cost, and in general the drilling costs correspond to 400 dollars per 0.3 meter. [8]

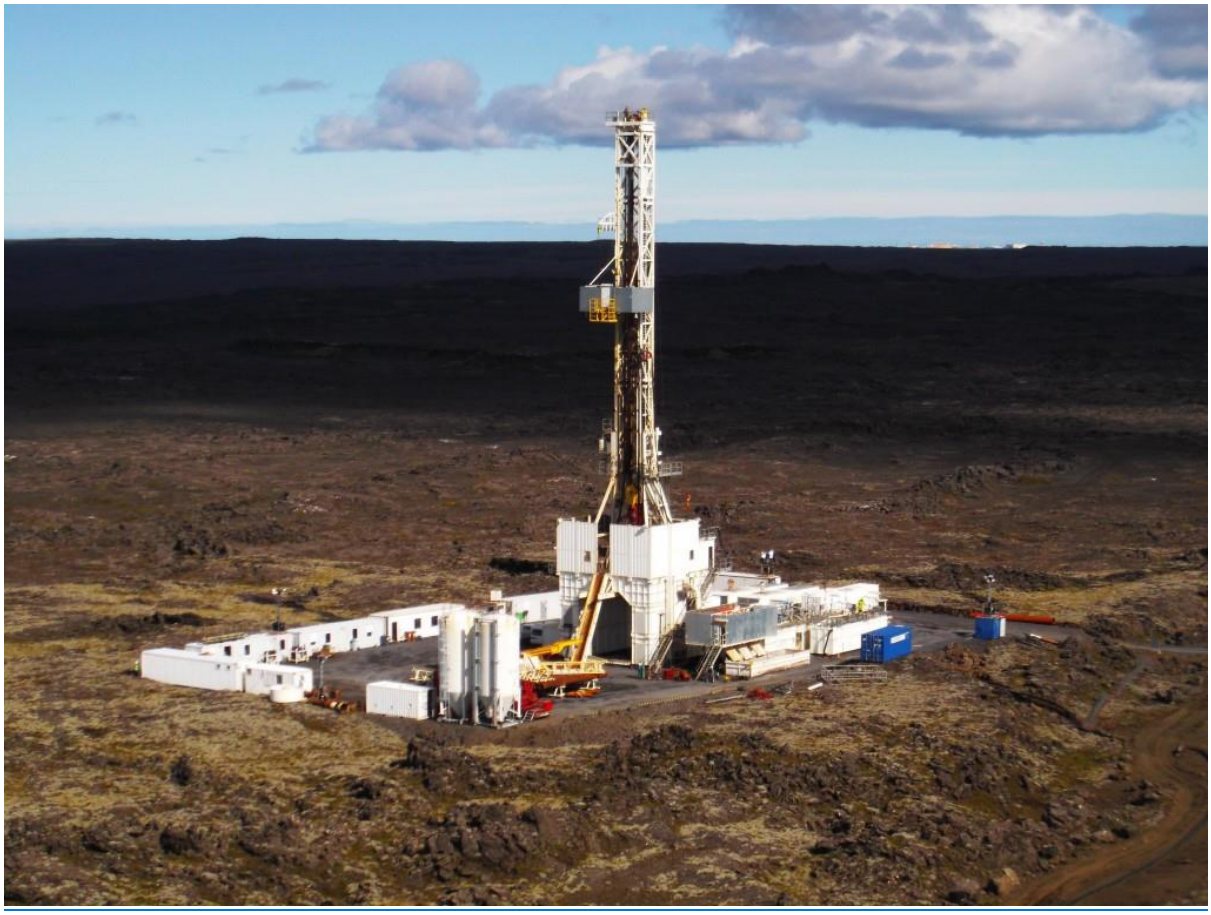


Figure 1.1. Geothermal Drilling

1.1.1.2. Seismology

Although seismology has generally been involved in the oil and gas industry from the past to this time, it is now also used in geothermal research. Seismic waves integrated into underground components react by interacting when propagated. This response and interaction can be divided into two, one is active seismology, where it is based on the use of man-made vibrations on the surface. The other is Passive seismology, where it uses tectonic activities, volcanic eruptions, etc.

However, as it is known, there are fault lines at the points where the areas are located in the geothermal fields, and with the seismicity level, it is understood whether these areas are geothermal. For example, there is geothermal energy in earthquake frequency and especially in areas below 2.0 magnitude and with a lot of earthquakes.

1.1.1.3. Magneto Telluric

As it is known, one of the most important factors in finding a geothermal resource is the temperature factor. Together with magneto telluric measurements, it helps to predict the geothermal reservoir temperatures. This first started in 1980 in countries such as America, Japan, New Zealand, and has been used all over the world and has been seriously helpful in finding and researching geothermal resources. These measuring devices are generally weak electrical conductors and are known to be resistant. [9] For example, flowing geothermal waters are thought to facilitate mapping geothermal areas, as they are of low resistance.

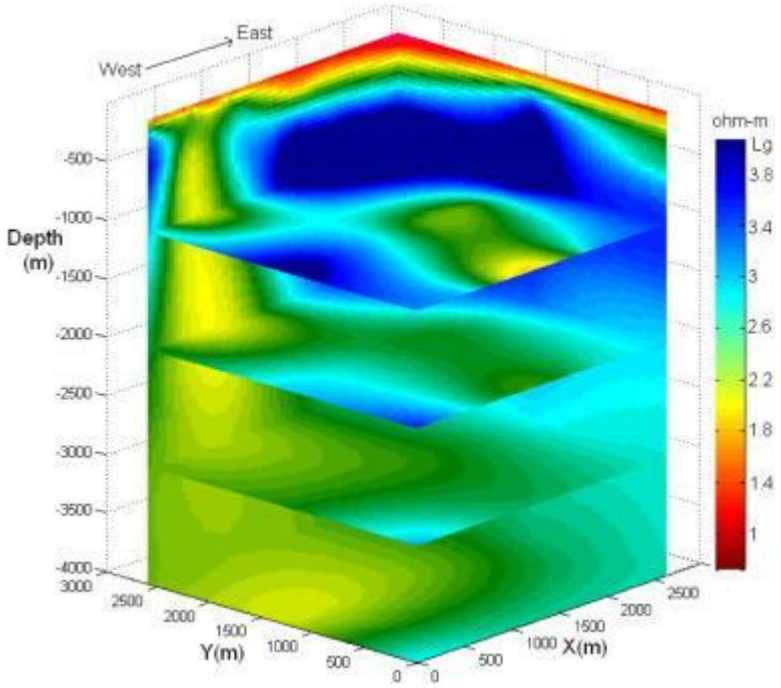


Figure 1.2. Magneto Telluric

1.1.1.4. Magnetics

Magnetism is one of the most well-known applications in the exploration and detection of geothermal fields. With this application, it learns the curie temperature, condensation or the depth of the curie point. Thus, finding the curie temperature of the materials underground is also an important parameter for future investments. As is generally known in geothermal fields, the curie temperature varies between approximately 200 and 570 degrees Celsius.

* Curie temperature: The curie temperature is the critical temperature at which a ferromagnetic material such as iron magnet loses its property and becomes paramagnetic.

1.1.1.5. Machine Learning

It is known how costly geothermal areas are, and besides this cost, it also takes a serious amount of time to explore. For this reason, various competitions have been organized for geothermal fields in recent years, for example a $ 5.5 million contest was organized by the US energy department for 10 projects using machine learning techniques. In fact, according to the Ministry, "If applied successfully, machine learning could lead to higher success rates in exploratory drilling, greater efficiency in plant operations, and ultimately lower costs for geothermal energy," the DOE said. We have also seen the reduction and attractiveness of the costs on it.[10] And it is known to significantly reduce costs. The data are generally composed of satellite images and these application images provide various information about the ground surface and the materials under it. Together with machine learning, it helps to map these geothermal fields by using the indicators (mineral) we have. Together with these maps, it helps to relate surface and underground features.

Artificial intelligence algorithms are used to detect whether geothermal fields exist, for example (SVM, CNN, DT) and it takes a lot of tags to make this prediction, which is normally a bit difficult. However, these difficulties are overcome by using machine learning algorithms. [11]

For example, as it is known, the fault is one of the important functions in geothermal space. Two people, researchers at Pennsylvania State University, are trying to learn induced seismicity for the geothermal field using machine learning. The reason for this is even a reference to geothermal areas where micro-earthquakes (2.0 and below) exist. These researchers who have been successful so far are to reach their predictions for the second phase of the Yang and Marone goals without any excavation on the site. [12]

1.1.2. Another Classification Method or Technique

1.1.2.1. Pixel Based classification

Pixel-based classification, which is one of the classification methods, is analyzed using individual image pixels with the spectral information they contain. One of the traditional approaches, this pixel-based classification is relatively easy as pixels are the smallest unit of an image. There are several types of pixel-based classifications, such as Maximum Lıkelihood, minimum-distance-to-means c, and a minimum of Mahalanobis distance. Shortly, If you use only spectral (such as pixel density) information for the training set, you will make a pixel-based classification. [13]

1.1.2.2. Maximum likelihood Method

In the maximum likelihood classification, consider the statistically normal distribution of each class in the bands and calculate the probability of whether the pixels belong to a certain class or not. All pixels are expected to be classified unless you set any probability criteria. As a result of the calculation, each pixel is assigned to the class that has the highest probability.[14] [15]

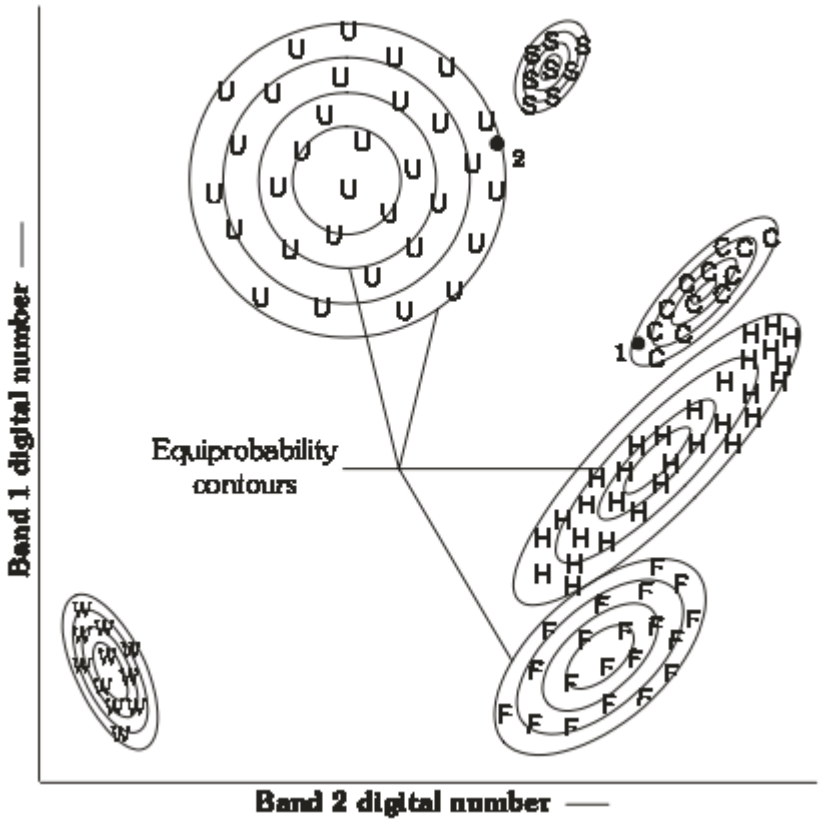


Figure 1.3. Maximum Likelihood Classification

1.1.2.3. Minimum-Distance-to-Means Classification Method

For the pixels of the already known classes, the average point of those in the numerical parameter space is calculated and for the unknown pixels, because there are different numerical number values, it is arithmetically assigned to the closest class. This discard is done through the remote sensing classification system. [16] [17]

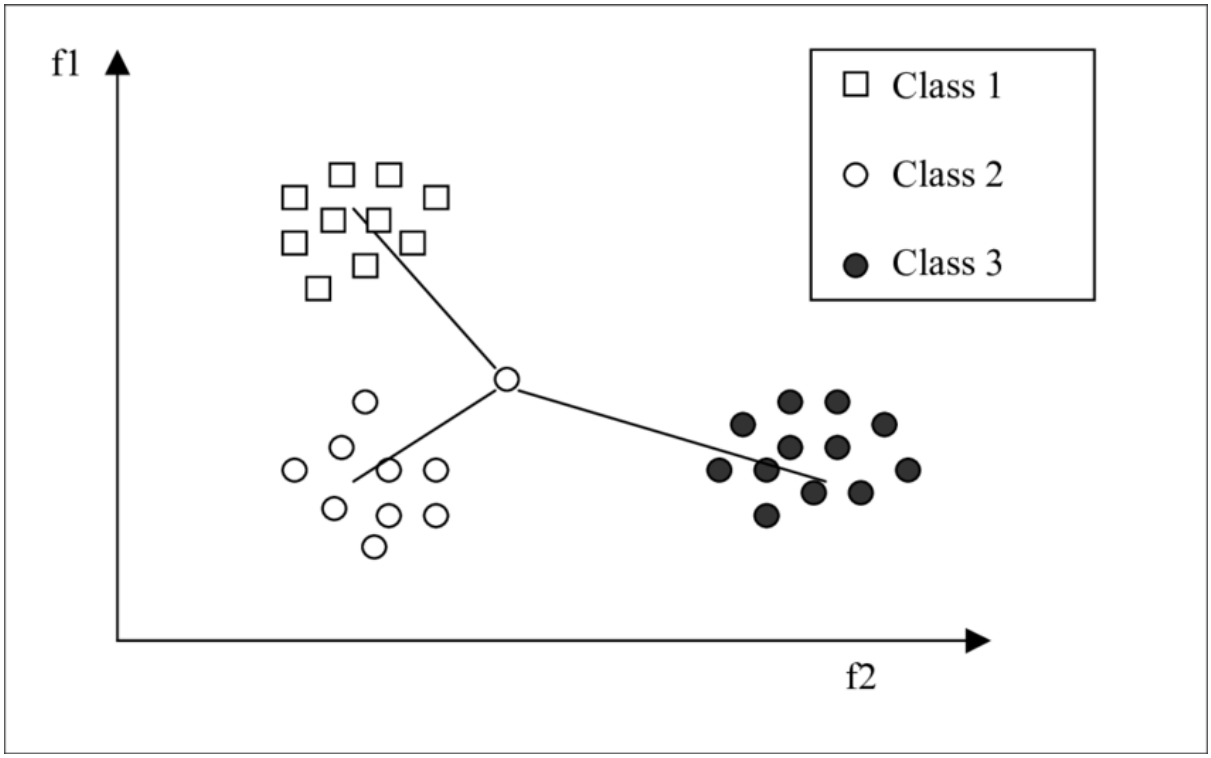


Figure 1.4. Minimum Distance-to-Means Classification

1.1.2.4. Mahalanobis Method

This distance-based classification known as Mahalanobis is known as a direction-sensitive classifier that uses statistics for each class. Although this Mahalanobis is similar to the maximum likelihood classification, he assumes that the covariances of all classes are equal and thinks this is the fastest method. When you do not specify any distance thresholds, some pixels that do not meet the threshold may be classified as. [18] [19]

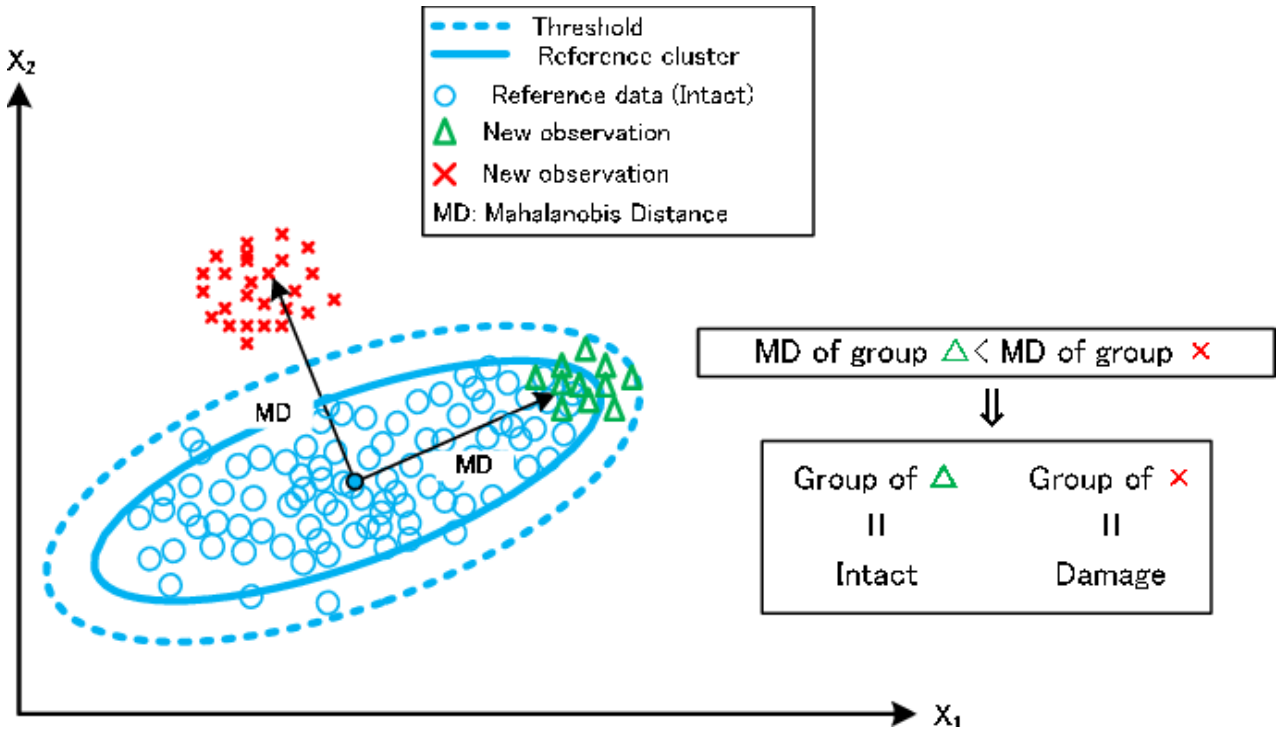


Figure 1.5. Mahalanobis Method

1.1.2.5. Object-based Method

This classification type, which is an object-oriented classification, uses both spatial information and spectral information. These object-based classification methods are a relatively later invented technique compared to traditional pixel-based (maximum likelihood) methods. Pixel-based classifications are based only on the spectral information found in pixels. They are groups of pixels that are similar to each other in the context of a neighboring neighborhood surrounding the image properties and image objects with spectral properties, shape, texture, and also the pixels. [20] Object-based classification, unlike pixel-based, here is a two-stage process, a cycle. In the first step, the image is divided into parts or divided into image features, after that, each image is classified as an object. Generally, this technique tries to interpret the image in hand and imitate the type of analysis made by the human. Shortly, If you include and use both spatial (neighboring pixels) and spectral information as a training set, you will have an object-based classification with this.

1.1.2.6. ANN Method

Artificial Neuron Network based on the structure and work of biological neural networks, the information flowing through the network affects the structure of the ANN because a neural network changes or learns in a sense according to this input and output.

Artificial Neuron Networks are considered and accepted as statistical data modeling tools of nonlinear ones where complex relationships between inputs and outputs are modeled or where models are found. Artificial neuron networks are also known as deep learning models capable of pattern recognition and machine learning. Finally, it is known as a part of artificial intelligence technology.

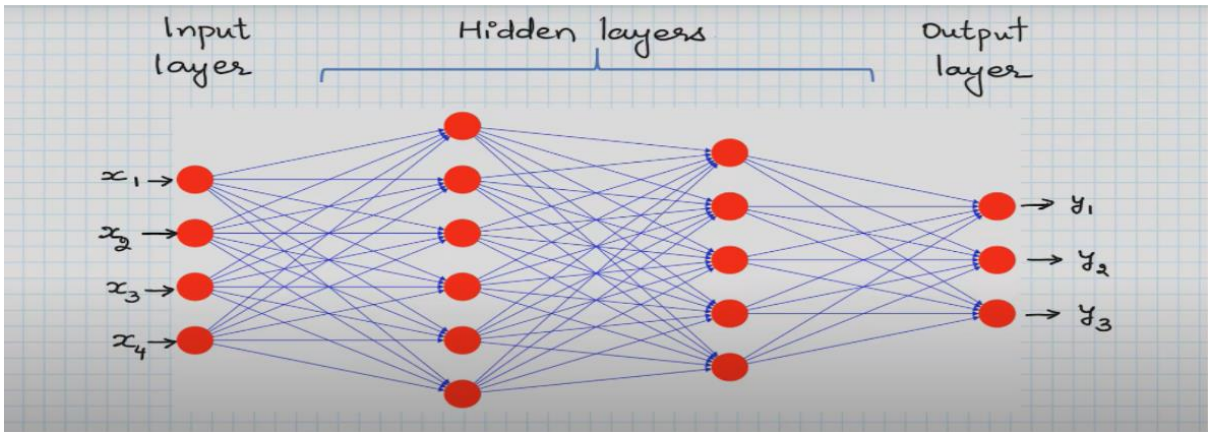


Figure 1.6. ANN Method

1.1.3. Logistic Regression

As is known, Logistic Regression is an appropriate regression analysis to be performed on two dependent variables. Logistic regression is used to describe its data and explain the relationship between a dependent binary variable and one or more independent variables at nominal, rank, range, or ratio level. [21]

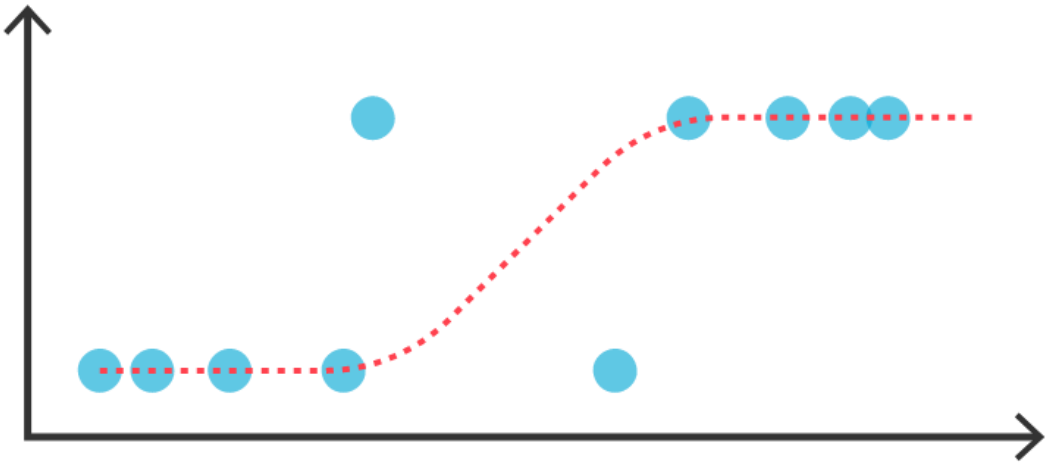


Figure 1.7. Logistic Regression

1.1.4. Raster Data

Rasters are made up of a matrix of pixels (also called cells), each containing a value that represents the conditions for the area covered by that cell (see figure 1). A raster dataset consists of rows (diagonally) and columns (flowing downward) of pixels (also known as cells). [22] In other words, in its simplest form, a raster consists of a matrix of cells organized into rows and columns where each cell contains a value representing information such as temperature or deformation. In reality, raster data is any pixel-based picture data (JPG, PNG, TIFF for example) which is loaded into the software. [23] Each pixel represents a geographic region, and the value in that pixel represents some feature of that region. Rasters are digital aerial photographs, imagery from satellites, digital pictures, or even scanned maps. [24] (Figure 2)

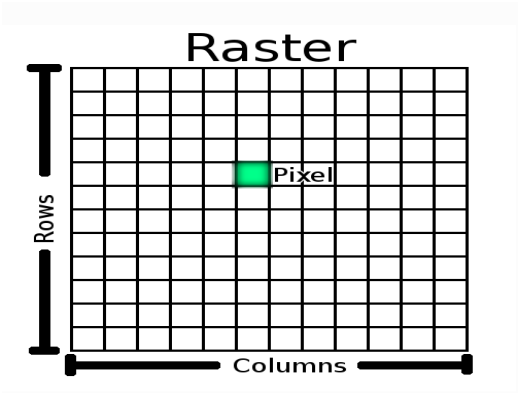


Figure 1.8. Raster data template

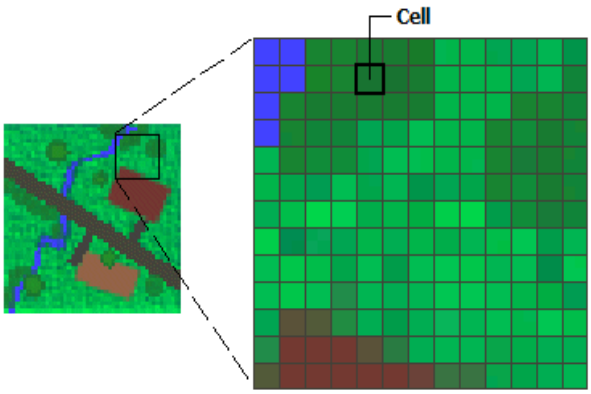


Figure 1.9. Raster data example

The primary use of raster data involves storing map information as digital images where cell values are associated with pixel colors of the image. To recreate the image, the computer reads

each of these cell values one by one and applies them to the pixels on the screen.[25] For instance, each pixel value in a satellite image has a red, green, and blue value. Alternatively, each value in an elevation map represents a specific height. It could represent anything from rainfall to land cover. Raster models are useful for storing data that varies continuously. For example, elevation surfaces, temperature, and lead contamination.[26] Raster data is useful when working with continuous data types: elevation, slope, satellite photos or good for large area analyses. Lastly it is good for surface analysis. [27] That is why raster data was used in this study. In this research, geothermal energy analysis was made by using multidimensional satellite images. These images are multidimensional raster data. Multidimensional data represents data captured at multiple times or multiple depths nor heights. Multidimensional raster data can be captured by satellite observations in which data is collected at certain time intervals, or generated from numerical models in which data is aggregated, interpolated, or simulated from other data sources. [28] In this study, there are 6 layers and 542897 pixels in the first data set named "brady_som" and in the second data set named "brady_desert" again there are 6 layers and 2533780 pixels. This dataset is raster data as each pixel has meaning. In other words, almost every pixel value in these satellite images has a different meaning. For example, some values indicate temperature, some values show deformation or faults.

1.1.5. Geothermal Field Indicators

Many indicators are used during the determination of geothermal areas. As a result of the research, when one or more of these indicators are encountered, it can be said that there is geothermal energy in the research area. In this study, temperature, deformation (subsidence and fall), fault, and mineral alteration indicators were used. Geothermal energy was determined in the area where the satellite images analyzed using these indicators were taken. In this study, different from the indicators used for geothermal field determination, different indicators were used in other studies. At some sites liquids and gases leak to the surface through faults and crevices or through permeable rocks, which are a clear indication of the existing geothermal reservoir in the ground. While these symptoms may be on top of a geothermal system, some may be discharged after flowing slightly downstream from a hydrothermal field [29]. Or, geothermal area estimation can be made by looking at the salinity rate in the region. Renner [30] defines the salinity of hot water systems as a salinity ratio ranging from 0.1% to 3%. Vapor-heavy reservoirs are good geothermal resources and occur where there is very high heat flow but low water charge. Near-surface gases filter out rocks in the source area as they condense to form acid. According to Fournier and Rowe [31], silica sinter beds are indicators

of the presence of hydrothermal reservoirs with temperatures higher than 175 ° C. On the other hand, parameters characterizing geothermal systems according to Macharia are: [32]

(1) High temperature

(2) High porosity

(3) High permeability

(4) Pressure and

(5) The chemical composition of the liquid [33].

A completely different perspective has been taken by geochemists as an indicator of the geothermal field. The geochemist searches for parameters that may give traces of geothermal fluids and prediction of subsurface temperature. Trace of metal in the fluid such as Mercury and Lithium will be an indicator of geothermal fields [34]. However, for this, a geochemist cost is added to the research as well as all costs. In this project, the results are obtained by using the indicators (temperature, deformation [collapse and fall], fault and mineral change) that are outside the indicators mentioned above.

1.1.5.1. Temperature

The "object" of the geothermal resource concept is “heat”. The entity whose existence is determined underground, and which is brought to the earth and transformed is "heat". This heat may be charged, circulating, accumulated in water, steam, gas, or hot dry rock, and emanates to the earth. [35] As a result of research, it is known that geothermal sources exhibit some characteristic surface temperature models. Why we choose temperature as an indicator because generally speaking sites of high-temperature geothermal energy resources are commonly marked by an area presenting thermal manifestation where fumaroles, steaming grounds, hydrothermally altered grounds, hot springs, volcanic gas vents, craters, and mud pools occur. [36] According to research conducted by MTA in 1996 and in Turkey in 2005. There are more than 110 different geothermal fields and found that they were between 22.5 and 220 degrees. These areas are generally located in Western Anatolia, Northern Anatolia, and Eastern Anatolia. The temperatures in 80 of the occurrences are above 60 degrees, in 13 of the above 100 degrees C, and in 8 occurrences above 140 degrees. All the data gave by MTA inventory and also the data available for the fields studied in the literature and as well as by our department in various projects are evaluated and used to estimate the identified apparent capacity. The identified apparent capacity is given in terms of maximum flow rate and the weighted flow rate temperatures. [37]

Table 1.1.5.1. Major geothermal fields of Turkey

Locality	Flow Rate (l/s)	Ave. Tem. (°C)	Max Temp. (°C)
Germencik/Aydin*	1515	220	232
Sultanhisar-Salavatli /Aydin°	731	163	171
Imamkoy/Aydin+	40	142	
Omer-Gecek /Afyon+	817.5	94	
Kizildere/Denizli°	250	217	242
Simav/Kutahya°	476	184	
Balcova/Izmir°	536	81	
Seferihisar/Izmir°	264	144	153
Diyadin /Agri+	561	72	
Sandikli/Afyon+	496	68	
Dikili/Izmir°	250	120	
Terme/Kirsehir°	688	102	
Kozakli/Nevsehir+	247	91	
Golemezli/Denizli+	340	70	
Kuzuluk/Sakarya+	271	81	
Tuzla/Canakkale+	120	160	174
Kula/Manisa+	140	135	
Salihli/Manisa°	150	104	
Caferbeyli/Manisa°	6.5	155	
Kavaklıdere/Manisa°	6.5	215	

In order for a geothermal field to be formed, a heat source, a porous reservoir rock, impermeable cover rock, and sufficient water supply are required. Measuring the underground temperature and determining the gradient increase is one of the healthy and reliable methods of geothermal energy exploration. The most striking feature of geothermal fields is that they contain an intense warm environment compared to their surroundings, so geophysical methods that examine the effects of heat on the physical properties of rocks form the basis of exploration activities [38]. On the other hand, remote sensing technology can play a role in geothermal exploration activity to map the distribution of land surface temperatures associated with geothermal manifestations. [39] From another point of view, it is very important to monitor thermal anomalies by measuring surface temperatures. Deviations in thermal flow (especially when associated with gas emissions and surface deformation) can indicate the presence of geothermal fields. [40] Because of these reasons, temperature is one of the most effective indicators used in this study.

1.1.5.2. Fault

The pattern and density of the faults present characteristic features for geothermal resources. Faults are another indicator used in this study. These are one of the most important indicators regarding geothermal fields. Knowledge of underground structure and error probabilities is essential to understanding the motion and degree of temperature. [41] Faults are important for understanding geothermal systems because of their impact on the stratigraphic structure and how this structure affects fluid flow. Fault networks influence not only conventional geothermal plays, but also hot sedimentary aquifers. Faults also create offsets in geological

layers, juxtaposing units with different hydraulic and thermal properties. Thus, faults can have a large impact on patterns of fluid flow and heat transport in sedimentary basins, which in turn has implications for geothermal and other resources. [42] The geothermal waters emerge along faults. This is another proof that the fault is an indicator for the geothermal field. Another proof is that geologic structure plays an important role in controlling fluid flow in geothermal systems. So in particular, very complex structural settings, consisting of many closely spaced and intersecting faults, host many geothermal systems. To elucidate the key geologic factors that affect fault-controlled geothermal circulation, it is critical to precisely characterize the structural and stratigraphic geometries in these complex settings. Faults and interconnected networks of faults and fractures can serve as pathways for upwelling in geothermal fields. This is particularly true in both magmatic and non-magmatic geothermal systems occupying extensional and transtensional domains in which normal and strike-slip faults serve as fluid conduits. [43] [44] There are a variety of techniques to find faults. The first is Standard geophysical techniques, but this technique is difficult to use in metro areas and large cities, instead Electromagnetic and magnetic techniques are used in living areas. For example, data were collected and analyzed in the Perth Basin, the southern part of Australia. According to the investigations, there is a serious fault network structure and heat transfer is also at a very high level. Therefore, faults have affected the probability of geothermal resources. Both pore pressure increase and fault slip contribute to the surface deformation. [45] Considering all this and researches, it can be said that geothermal resources exhibit some characteristic fault models.

1.1.5.3. Deformation (Subsidence & Uplift)

Deformation in the form of subsidence and uplift in geothermal fields is a highly curious and investigated subject in the literature. Subsidence is an indicator of the geothermal field because the geological rocks forming the underground layers change their physical-chemical properties by natural or artificial means and their vertical mobility is in question. As a result of these changes and mobility, surface deformation in the form of "subsidence" is observed on the earth with the formation of gaps in the lower layers. [46] Deformation is the unit deformation, deformity, and deformation of an object under the effect of various internal and external forces. These deformations are mainly due to pore pressure changes and thermal shrinkage. In this study, this indicator was used to determine the basic truth about the location and extent of the geothermal system utilized in the blind geothermal system at both Brady and Desert Peak. Besides, temperature, like pressure, can play a role in deformation in geothermal areas. If the

injected fluid is warmer than the rock, thermal expansion can cause deformation, or if, in contrast, water is injected into the rock that has a higher temperature, the water can boil and subsequent pressure increase can cause deformation [47]. There is deformation associated with both natural and man-made processes. Deformation is usually so great that tectonic processes such as plate boundary deformation can sometimes hide the deformation that occurs in the earth. Generally, geothermal energy, hydrocarbon production, and mining are carried out in such areas. For example, according to research, it was observed that deformation was noisy in the Reykjanes peninsula and as a result, the island shifted to the left and stretched slightly along the border region. [48] Subsequently, geothermal fields in the Reykjanes peninsula have been used since 1976, and in 2006 two new power plants were put into operation and there was subsidence due to man-made. On the other hand, the nature of geothermal operations causes a specific pattern of subsidence and uplift. The subsidence in a geothermal field stems from the interaction of various complex geological and production-related factors. Among these various factors, three of them can be considered as the leading factors. [49]

As a result of the gaps in the areas where the underground geothermal fluid is used, surface subsidence of 0.192 m occurred between 1977 and 1996. During this period, it was noticed that the current subsidence was larger compared to the subsidence that occurred at the place where many active vapor extractions were made. As a result, it was judged that more surface subsidence occurred in flooded areas than in vapor-dominated areas. [50]

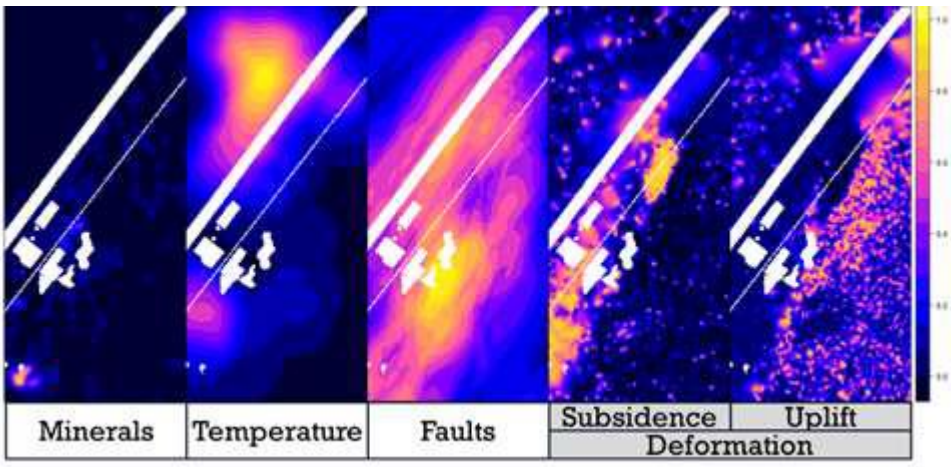
1.1.5.4. Mineral Alteration

The last indicator used in this research is "mineral alteration". Which is also a very clear indicator. [51] While some alteration minerals can be found in any alteration environment, some may occur under limited conditions; therefore, they are used to determine alteration conditions.[52] Hydrothermally altered minerals are generated as a result of minerals exposure to water at high temperatures. These minerals surface by means of faults, hot springs, and fumaroles, and their chemical characteristics become a signature of the probable underground system characteristics. [53][54] Geothermal areas generally have a separate signature in terms of surface mineralogy. In addition to the chemistry of brine that interacts with the bedrock in which liquids circulate, the reactivity of the minerals found in these bedrock changes significantly. This provides an important advantage in geothermal field detection. Additionally, surface mapping of mineralogy and rock type, vegetation stress, and thermal anomalies can be used in conjunction with structural and subsurface context to provide a more comprehensive picture of known geothermal source regions and can help identify new sources or expand

existing fields. [55] As another point of view, alteration mineral assemblages characterize the hydrothermal-chemical reactions that generally occur in geothermal areas. [56] According to mineral synthesis laboratories, chemical and temperature conditions affect their formation in general. Actually, alternation is the second type of change that results from interaction with atmospheric substances and earth conditions with organisms, generally called exposure to weather conditions. It is clear that changes that require high temperature or pressure cannot occur in the soil and therefore cannot be interpreted as pedogenic but are sometimes overlooked by micromorphologists. Tendency to change; It is a reflection of the mineral stability in soil environments determined by many factors such as temperature, humidity conditions, pH, redox conditions, the degree of infiltration, and the nature of the mineral, its composition, and grain size. Discussion of these factors is beyond the scope of this section but can be found in other publications. Finally, it is understood that the factors affecting mineral alteration, why it entered into this transformation, and its effect on geothermal fields. [57]

Figure (a) below shows the resulting layers for Brady Hot Springs, while Figure (b) shows the same layers for Desert Peak.

(a)



(b)

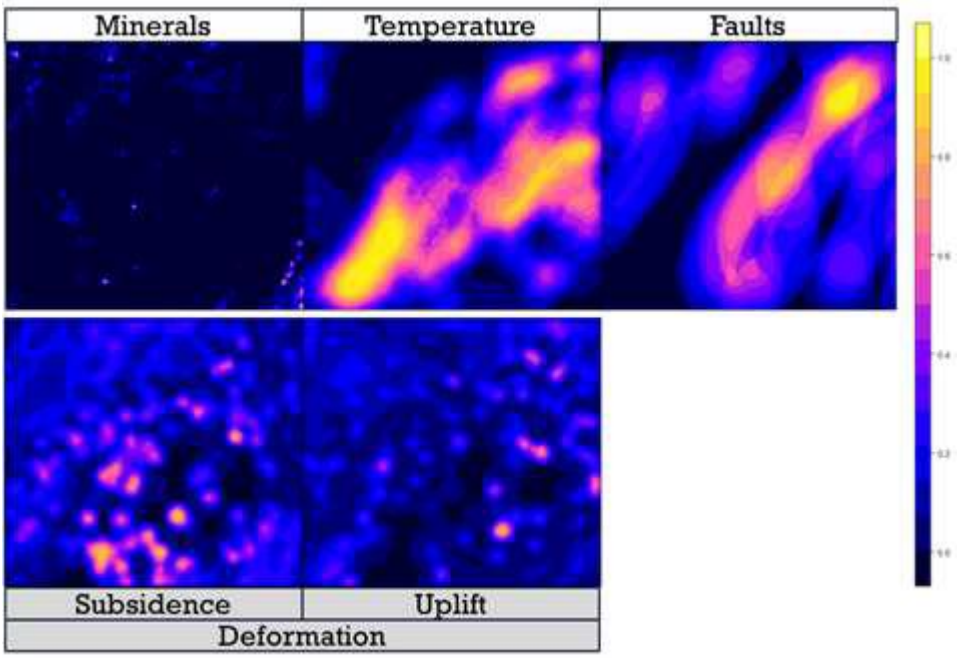


Figure 1.10 — Input layers for labeling algorithm: (a) Brady's hot springs; (b) Desert Peak

2 METHODOLOGY

In this study, different classification methods were used to predict the geothermal area of specific lands. These lands are generated from satellite images. Most of the work is done by using Python programming language using Google Colab environment. Python is one of most chosen programming languages for designing AI algorithms. It provides simple usage and thousands of open source libraries to use. According to the author Python libraries are "Python Libraries are a set of useful functions that eliminate the need for writing codes from scratch. There are over 137,000 python libraries present today. Python libraries play a vital role in developing machine learning, data science, data visualization, image and data manipulation applications and more." [58] As mentioned before Google Colab used in this study. which provides the ability to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. Which is a user could see the output of their code in every line that they created. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs.

2.1. Classification

Classification in machine learning and statistics is a supervised learning approach in which the computer program learns from the data given to it and makes new observations or classifications. In supervised learning, algorithms learn from labeled data. After understanding the data, the algorithm determines which label should be given to new data by associating patterns to the unlabeled new data. In this study we are performing three different supervised learning classification methods which are CNN, SVM and Decision Tree.

2.1.1. Decision Tree Algorithm

Decision Tree, one of the most well-known predictive approaches, is used in machine learning, data mining, and statistics.

It uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves).

Classification trees are meant to allow the target variable to take a separate set of values. In the tree structures found here, the leaves represent the class label, while the branches represent the combinations of features that lead to the class labels.

If the target value of the decision tree can have a continuous value, it is called a regression tree. As it is known, the decision tree is one of the most popular and known machine learning algorithms.

A decision tree or a classification tree is a tree in which each internal (non-leaf) node is labeled with an input feature. The arcs coming from a node labeled with an input feature are labeled with each of the possible values of the target feature or the arc leads to a subordinate decision node on a different input feature. Each leaf of the tree is labeled with a class or a probability distribution over the classes, signifying that the data set has been classified by the tree into either a specific class or into a particular probability distribution (which, if the decision tree is well -constructed , is skewed towards certain subsets of classes.

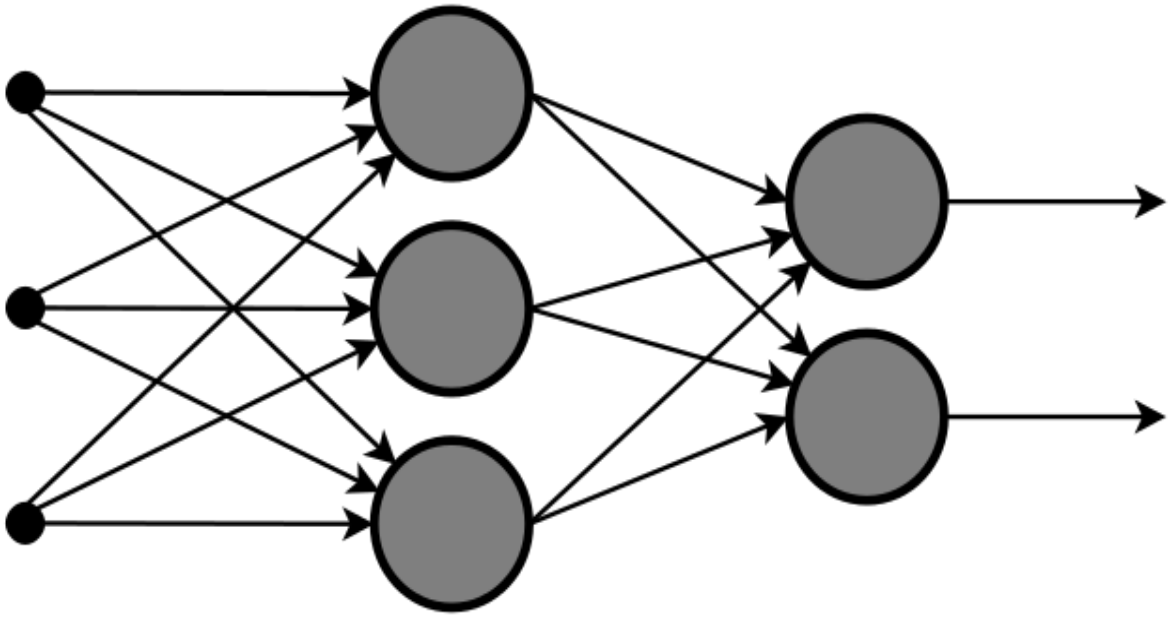


Figure 2.1. Decision Tree Topology

The decision of making strategic splits heavily affects a tree's accuracy. The decision criteria are different for classification and regression trees.

Decision trees use multiple algorithms to decide to split a node into two or more sub-nodes. The creation of sub-nodes increases the homogeneity of resultant sub-nodes. In other words, we can say that the purity of the node increases with respect to the target variable. The decision tree splits the nodes on all available variables and then selects the split which results in most homogeneous sub-nodes. [59]

2.1.2. CNN Algorithm

Convolutional neural networks were discovered by Japanese Yann Lecun, a computer research researcher.

The behavior of neurons depends on the weights of the pixel values, and when we add the pixel values, the various images are chosen by the neurons of a CNN. The interactions at the layers are created by entering an image into ConvNet, for example, to examine the activations related to the entered image. One of the pixels in the pictures entered is taken as a patch by the neurons and this input is counted. This input is multiplied by color weights, balls, and color is run at activation.

However, horizontal, vertical, and cross edges are detected in the first or lowest layer of CNN. The resulting output is input to layers with more complex corners. In short, we remove the complexity in the corners of the next layers with the first result we get, for example, edge combinations. Subsequently, faces, layers become perceived as objects that go deeper in CNN. Addition and multiplication with pixel weights such as the above operations are called convolution for short. So the name Convolution Neural Network also comes from here. As is known, many layers make up CNN. The last layer is known as the classification layer and here it is a classification layer that takes as input to the output of the last convolution layer. [60] According to the activation area of the last convolution layer, the classification layer yields a set of confidence points (values between 0 and 1) determining the likelihood that the image belongs to a "class". For example, if you have a ConvNet that detects cats, dogs, and horses, the output of the last layer is that the entrance review includes any of these animals.[61]

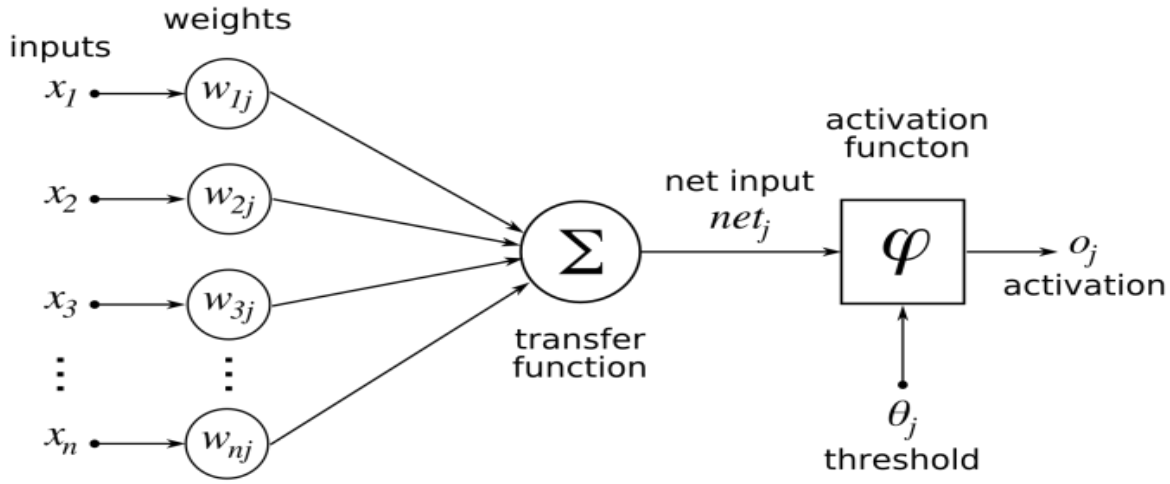


Figure 2.2. Neural Network Architecture

2.1.3. SVM Algorithm

If we talk about the general purpose of SVM, it is to find the hyperplane in N-Dimensional Space and separate the data points there. And your goal here is to find the maximum margin on the plane. For example, the distance of maximum data points in two classes. Maximizing the margin allows more secure classification of future data points.

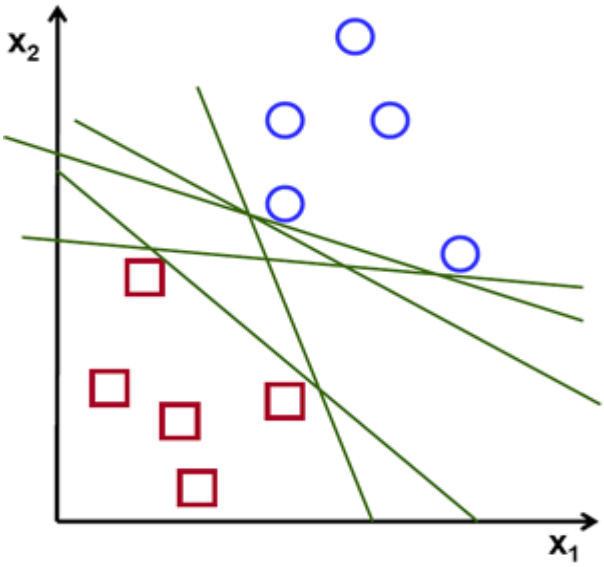


Figure 2.3. SVM Graphic

Take a look at the example below to better understand SVM. Meanwhile we are trying to find the best hyperplane to divide both classes. (Circle A class, Triagle B Class)

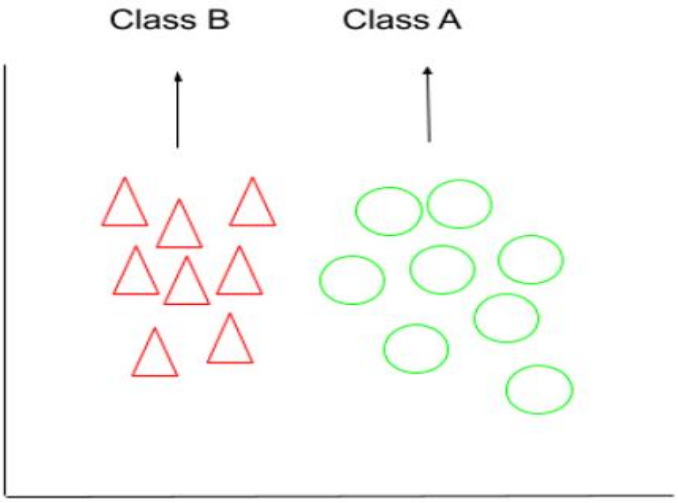


Figure 2.4. SVM Hyperplane

All data points are received by SVM, and a line called Hyperplane that divides the class is created. This line is called "decision boundary", and as given in the example, everything entered into the triangle class will be in the triangle, and the rest will be in the circle class.

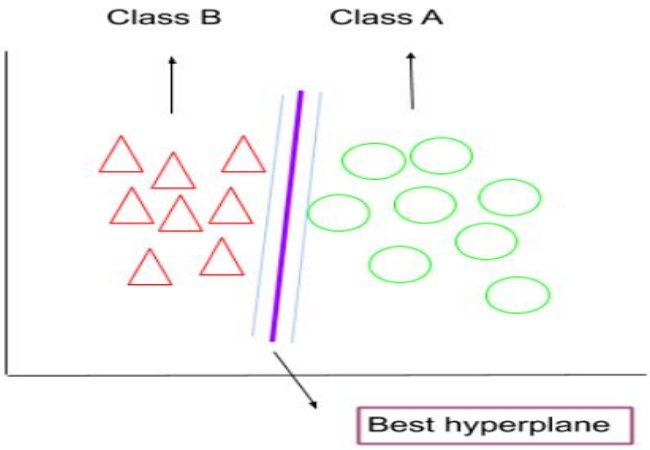


Figure 2.5. SVM best hyperlane

Many hyperplanes can be made that separate classes such as those in the example, and at the end of the job one of the hyperplanes with the largest distance is considered the best, most reliable.

As you can see in the example below, consider two classes, yellow and red, and here we try to find the hyperplane that separates the class best.

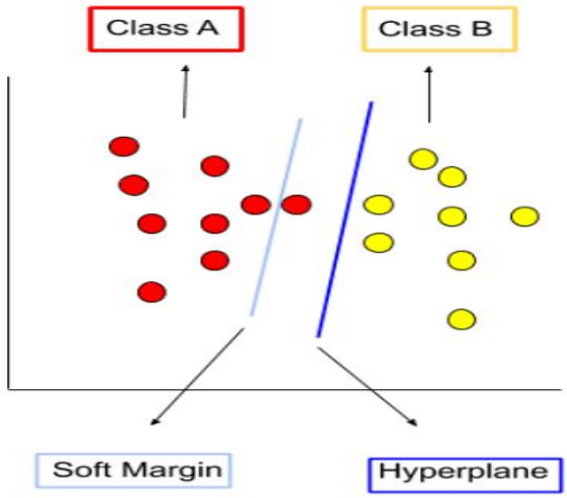


Figure 2.6. SVM Classes

Lastly, as you can see in the example, soft margin makes the wrong classification and ignores one of the data points, but as you can see on the hyperplane side, it minimizes the classification errors by constantly trying to find the best separation. [62] [63]

3 IMPLEMENTATION

3.1. Dataset

In this project, multispectral raster images are used for data meaning that [64] “A multispectral image comprises a set of co-registered images, each of which captures the spatially varying brightness of a scene in a specific spectral band, or electromagnetic wavelength region. An image is structured as a raster, or grid, of pixels”. Two different datasets used in this project which are in “.grd” raster image format, and these images are stacked with six different layers. These datasets are called “brady_som” and “brady_desert” which contains same features which are the layers of the images and these features are Geothermal, Minerals, Temperature, Faults, Subsidence and Uplift each of these features were normalized 0 to 1 which is helpful for the implementation of the classification algorithms. In order to perform classification methods each pixel of these figures needed to be extracted. The images are collected from The Brady Hot Springs Geothermal area is located North East from Fernley, NV, and is part of Nevada’s Northwest Basin and Range Geothermal Region. In this area, there are geothermal operations in two sites, Brady (39.79°N, 119.02°W) and Desert Peak (39.75°N, 118.95°W). Map view of the area is given in the figure below. Figure[1]

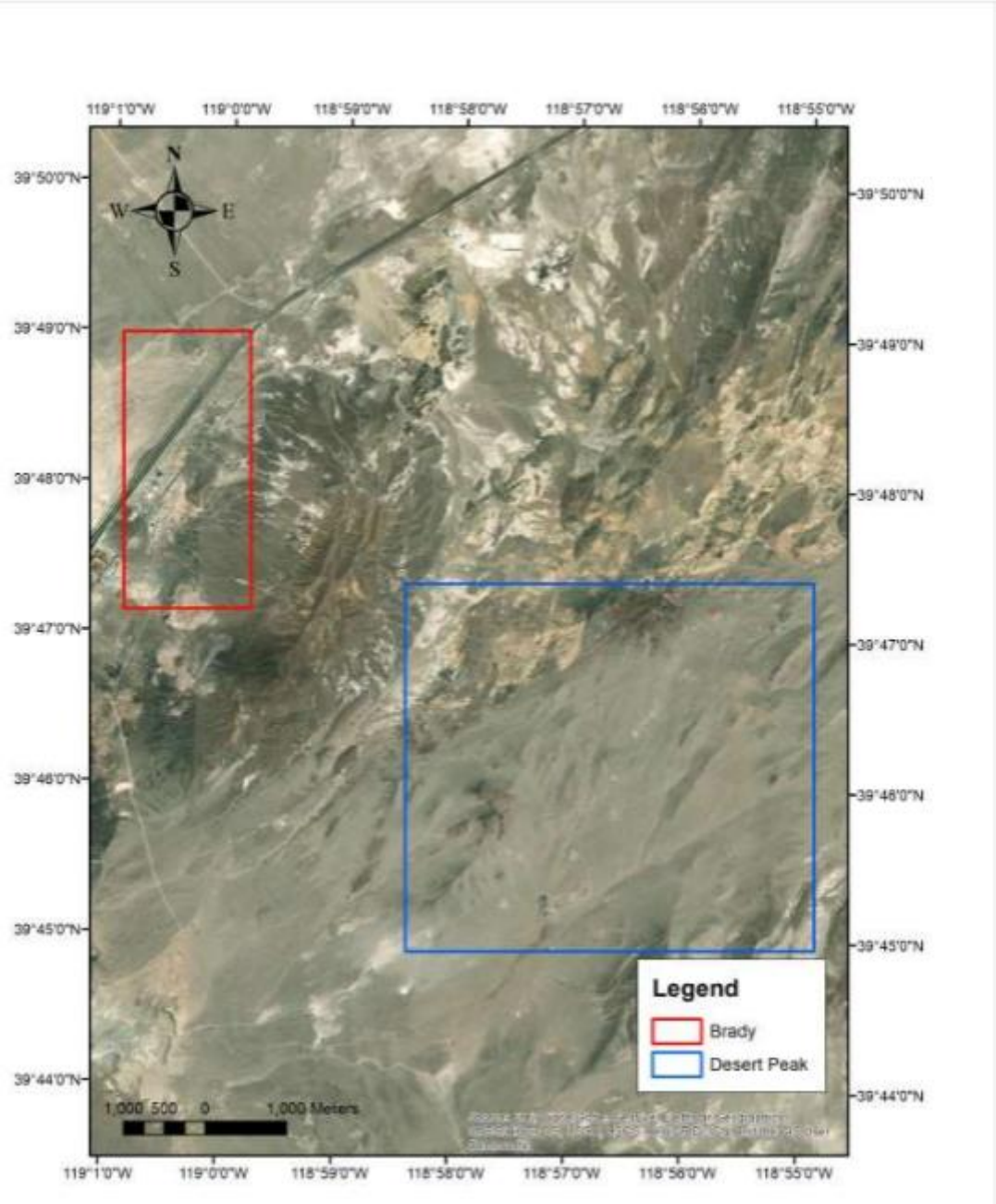


Figure 3.1. Satellite Map View of the Area

The first figure (dataset) which is called “brady_som” contains 542897 pixels and the other figure (dataset) which is called “brady_desert” contains 2533780 pixels. In the figures down below descriptive values of the datasets are given.Figure[2][3]

Brady Desert Dataset Summary

Geothermal	Minerals	Temperature	Faults	Subsidence	Uplift
Min. :0.000	Min. :0.000000	Min. :0.00000	Min. :0.00000	Min. :0.0000	Min. :0.00000
1st Qu.:0.000	1st Qu.:0.000000	1st Qu.:0.03004	1st Qu.:0.09649	1st Qu.:0.0715	1st Qu.:0.03489
Median :0.000	Median :0.000000	Median :0.22478	Median :0.21921	Median :0.1398	Median :0.08555
Mean :0.486	Mean :0.007306	Mean :0.29898	Mean :0.25662	Mean :0.1669	Mean :0.10180
3rd Qu.:1.000	3rd Qu.:0.000000	3rd Qu.:0.51057	3rd Qu.:0.36384	3rd Qu.:0.2232	3rd Qu.:0.14137
Max. :1.000	Max. :1.000000	Max. :1.00000	Max. :1.00000	Max. :1.0000	Max. :1.00000

Figure 3.2. Brady Desert Dataset Summary

Brady Som Dataset Summary

```
   Geothermal        Minerals          Temperature         Faults          Subsidence          Uplift
 Min.   :0.0000   Min.   :0.000000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000000   Min.   :0.00000
 1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:0.0165   1st Qu.:0.2842   1st Qu.:0.003309   1st Qu.:0.03948
 Median :1.0000   Median :0.000000   Median :0.1131   Median :0.4839   Median :0.044583   Median :0.14672
 Mean   :0.5934   Mean   :0.007948   Mean   :0.2092   Mean   :0.4576   Mean   :0.166039   Mean   :0.22447
 3rd Qu.:1.0000   3rd Qu.:0.000000   3rd Qu.:0.3399   3rd Qu.:0.5921   3rd Qu.:0.247692   3rd Qu.:0.36455
 Max.   :1.0000   Max.   :1.000000   Max.   :1.0000   Max.   :1.0000   Max.   :1.000000   Max.   :1.00000
 NA's   :60                                           NA's   :60
```

Figure 3.3. Brady Som Dataset Summary

3.2 Convolutional Neural Network Classifier

First, the classification algorithm used in this project is Convolutional Neural Network Classifier which is commonly used in image classification. Creating a code structure must be done to be able to implement the algorithm. To perform classification processes 2 different classes used in this and the other algorithms which are “geothermal” as the value of 0 and “non-geothermal” as the value of 1 each of the pixels classified according to these values by a classification using all the features contained in the datasets. In order to perform a CNN algorithm performing data pre-processing is needed which is called image patching. An image patch is a container of pixels in larger form. A patch is a small piece of an image in this project shape of pieces is rectangular.[69]

In this part of the project Python programming language is used in the Google Colab environment. Python is one of the most popular programming languages for AI algorithms. Also, the Colab environment provides a cloud-based implementation for Python. All the libraries that we used in this project are provided in the table below. [Table 3.2.1]

Table 3.2.1. Python Library List

Library Name	Description
Gdal	GDAL is a translator library for raster and vector geospatial data formats that is released under an X/MIT style Open Source License by the Open Source Geospatial Foundation. As a library, it presents a single raster abstract data model and single vector abstract data model to the calling application for all supported formats. It also comes with a variety of useful command line utilities for data translation and processing. The NEWS page describes the May 2021 GDAL/OGR 3.3.0 release. [70]
Numpy	NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python. [71]
Keras	Keras is a powerful and easy-to-use free open source Python library for developing and evaluating deep learning models. It wraps the efficient numerical computation libraries Theano and TensorFlow and allows you to define and train neural network models in just a few lines of code. In this tutorial, you will discover how to create your first deep learning neural network model in Python using Keras. [72]

Sklearn	Scikit-learn (formerly scikits.learn and also known as sklearn) is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy. [73]
Pandas	Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data analysis" itself. Wes McKinney started building what would become pandas at AQR Capital while he was a researcher there from 2007 to 2010.[74]
Rasterio	Geographic information systems use GeoTIFF and other formats to organize and store gridded raster datasets such as satellite imagery and terrain models. Rasterio reads and writes these formats and provides a Python API based on Numpy N-dimensional arrays and GeoJSON.[75]
Geopandas	The goal of GeoPandas is to make working with geospatial data in python easier. It combines the capabilities of pandas and shapely, providing geospatial operations in pandas and a high-level interface to multiple geometries to shapely. GeoPandas enables you to easily do operations in python that

	would otherwise require a spatial database such as PostGIS.[76]

First, data that is used as input should be in a format that a computer could understand for a Python package named “Osgeo” used from that package “Gdal_array” module used in this study. This module is for reading raster image data as three dimensional which are width, height, channels. In here these three dimensions refer to the shape of the image. After reading the step of the image as a return a “numpy” array is obtained. The term “numpy” comes from a python library called “NumPy" which refers to [65] “NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.”. Then as mentioned before the images that will be used in CNN algorithm should be reshaped in order to fit into a model that will be designed. For this operation image patching should be performed. In the image patching process some parameters are used which are, kernel size is the size of the patches in this work “17x17 " is used and kernel channels which are “5” these channels are representative of the dimensions of the image that is used. The image that is used for this study has six dimensions which five of them are used as features and one of them is for labels since a supervised model is used for this project. With these parameters three dimensional image patches are extracted to fit into the CNN model. After performing image patching each patch that is generated classified according to their labelling values which are 0 and 1, and created new databases. After preprocessing steps are done, the modelling part is needed. As mentioned before CNN algorithm is used for this part of the study, for this model a specific Python library is needed called “Keras”, there are several optimizers for CNN algorithm in this project for an optimizer Adadelta is used from “Keras”. The CNN model created consists of several layers called; Convolutional Layer, Max Pooling, Concatenate, Average Pooling, Flatten and Dense at the last Activation Function is performed with a total of 3,043,410 parameters. Each of the layers are given in the table below with parameters and the layers that are connected to [Table 3.2.].

Table 3.2.2. CNN layers and parameters

Layer	**Output Shape**	**Parameter**	**Connected to**
input_1 (InputLayer)	(None, 17, 17, 5)	0	
conv2d (Conv2D)	(None, 17, 17, 128)	768	input_1[0][0]
conv2d_2 (Conv2D)	(None, 17, 17, 96)	12384	conv2d[0][0]
conv2d_3 (Conv2D)	(None, 17, 17, 16)	2064	conv2d[0][0]
conv2d_4 (Conv2D)	(None, 17, 17, 16)	2064	conv2d[0][0]
conv2d_5 (Conv2D)	(None, 17, 17, 16)	2064	conv2d[0][0]
max_pooling2d (MaxPooling2D)	(None, 17, 17, 128)	0	conv2d[0][0]
conv2d_1 (Conv2D)	(None, 17, 17, 256)	33024	conv2d[0][0]
i_3x3 (Conv2D)	(None, 17, 17, 128)	110720	conv2d_2[0][0]
i_5x5 (Conv2D)	(None, 17, 17, 128)	51328	conv2d_3[0][0]
i_7x7 (Conv2D)	(None, 17, 17, 128)	100480	conv2d_4[0][0]
i_9x9 (Conv2D)	(None, 17, 17, 64)	50240	conv2d_5[0][0]
conv2d_6 (Conv2D)	(None, 17, 17, 32)	4128	max_pooling2d[0][0]
flatten_1 (Flatten)	(None, 1445)	0	input_1[0][0]
concatenate (Concatenate)	(None, 17, 17, 736)	0	conv2d_1[0][0] i_3x3[0][0] i_5x5[0][0] i_7x7[0][0] i_9x9[0][0] conv2d_6[0][0]
dense_1 (Dense)	(None, 16)	23136	flatten_1[0][0]

avg_pooling (AveragePooling2D)	(None, 15, 15, 736)	0	concatenate[0][0]
dropout_1 (Dropout)	(None, 16)	0	dense_1[0][0]
flatten (Flatten)	(None, 165600)	0	avg_pooling[0][0]
dense_2 (Dense)	(None, 16)	272	dropout_1[0][0]
dense (Dense)	(None, 16)	2649616	flatten[0][0]
dropout_2 (Dropout)	(None, 16)	0	dense_2[0][0]
dropout (Dropout)	(None, 16)	0	dense[0][0]
concatenate_1 (Concatenate)	(None, 32)	0	dropout_2[0][0] dropout[0][0]
dense_3 (Dense)	(None, 32)	1056	concatenate_1[0][0]
dropout_3 (Dropout)	(None, 32)	0	dense_3[0][0]
dense_4 (Dense)	(None, 2)	66	dropout_3[0][0]
prob (Activation)	(None, 2)	0	dense_4[0][0]

3.2.1 Convolutional Layer

The authors stated that; [68] “Convolutional layers are the major building blocks used in convolutional neural networks. A convolution is the simple application of a filter to an input that results in an activation. Repeated application of the same filter to an input results in a map of activations called a feature map, indicating the locations and strength of a detected feature in an input, such as an image.”. This layer consists of 2D convolutions with the defined kernel

sizes. In our work 3x3, 5x5, 7x7 and 9x9 nodes are implemented. In this layer filtering task is applied and best filter matches to the relevant patch.

3.2.2 Max Pooling Layer

In this layer 3x3 patches (matrices) are used and in each patch maximum is obtained later on will be used. In the figure below (Figure 4) representative of 3x3 patches with stride 1x1 is given.

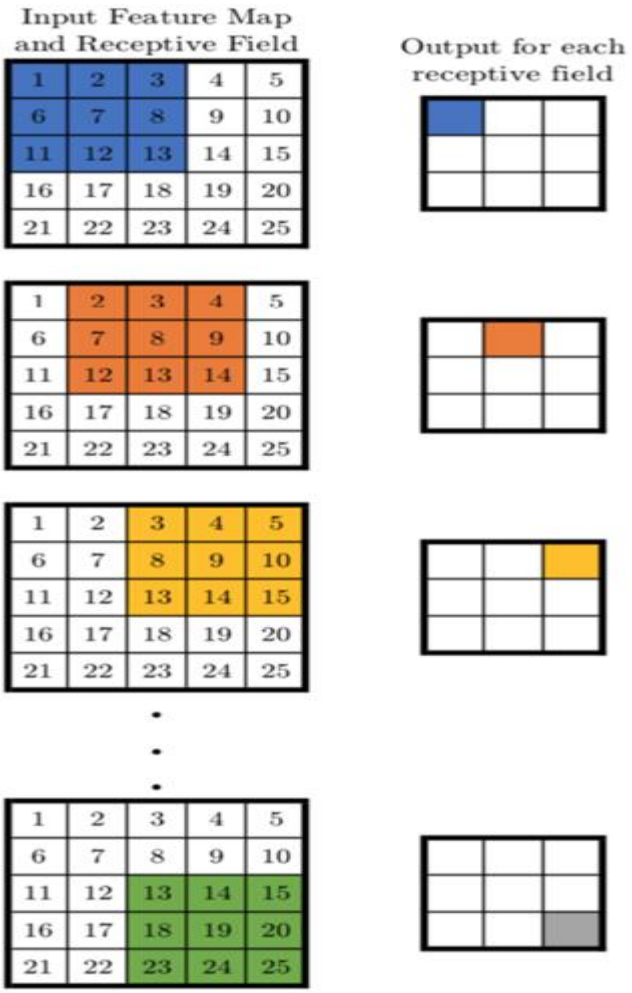


Figure 3.4. Max Pooling Architecture

3.2.3 Concatenate

In this layer the data is used to merge into a one single input.

3.2.4 Average Pooling

As in Max Pooling Layer we used 3x3 matrices but instead of getting the maximum value this time average value is obtained.

3.2.5 Flatten

In this layer the pooled layers are converted one dimensional into a single column.

3.2.6 Dense

The dense layer is a fully connected layer, meaning all the neurons in a layer are connected to those in the next layer.

3.2.7 Activation Function

For the last part Activation Function is performed which is a function that you use to get the output of a node it differs based on the selected function. In this project SoftMax is used, which turns a continuous variable into a label (“Geothermal” or “Non-geothermal”)

First, we trained our model using “Brady” data which contains 542897 pixel since our computational is not enough for the all pixels to be used in CNN algorithm we could use very small partition of the dataset which is 2400 pixels 20 percent of the pixels were used for testing and the same methods are applied to “Desert Peak” data.

In the figure below the topology of the CNN algorithm is provided. Figure[5]

3.2.8. CNN Topology

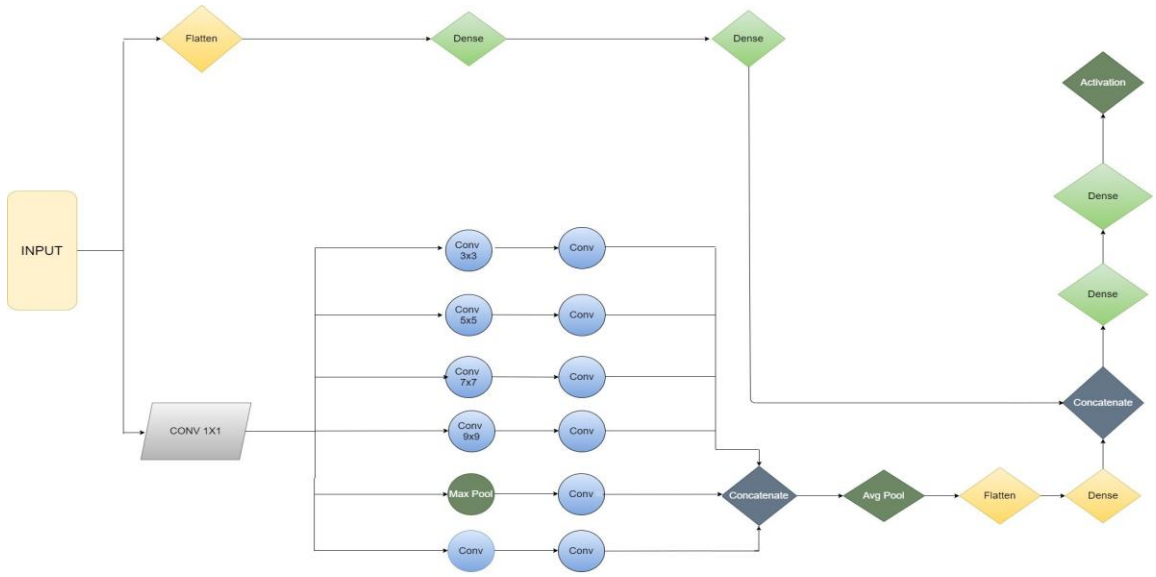


Figure 3.5. CNN Topology

3.3 Support Vector Machine & Decision Tree Classifiers

In this part of the study two different classifiers methods are used for comparing the result that is obtained by CNN algorithm. Before designing the models the data that is used for input for the models should be in a format that the models are designed could take as an input. For that "R" programming language is used in "Rstudio " environment. The image that is used as data is an unstructured data meaning that, [66]"Unstructured data (also known as qualitative data) is information that has not been structured in a predefined manner. Unstructured data is typically text-heavy, like open-ended survey responses and social media conversations, but also includes images, video, and audio.". So the data that is used needs to be converted into structured data which consists of rows and columns. Like mentioned before, R programming language is used for this purpose. For this need a R library called "Raster" is used which is suitable for converting raster data into data frames. As mentioned before, images that are used in this study consist of six different layers. So each layer's values are extracted with their corresponding coordinates as "x" and "y" values are extracted in the end a data frame is obtained for the two images that are used. The modelling part is done by using Python

programming language. To import these models, “SVM " and “Decision Tree” a library called “sklearn” is used which is an efficient tool for predictive data analysis and contains several AI models. In the first part which the “brady_som” image is used for both training and testing is used for, in each model 70 percent of the data is used as training and 30 percent of the data is used for testing. In the following attempts. The other image called “desert_som” is used for testing while “brady_som” data is used for training as a whole and “desert_som” was used for testing as a whole. The same steps are done as “desert_som” used as the train set and “brady_som” as the test set.

3.4 Prediction Maps

The prediction map obtained by CNN algorithms is done by getting the shape of the original image’s first layer which is called “Geothermal” in the model. The data is not used as a whole, since the computural power is not enough for all of the image. For this masking process is performed. The meaning of “Masking” is [67] “Masking is a way to tell sequence-processing layers that certain timesteps in an input are missing, and thus should be skipped when processing the data.”. After performing CNN algorithm the shape of the original data is saved as a mask and created an empty map with value of zeros later on replaced with the corresponding prediction values. With the obtained array the array turned into a raster “.gri” file using “osgeo" library. The file that is obtained could be plotted by using Python programming language, GIS tools or R programming language. The path that is followed for SVM and Decision Tree algorithm’s prediction map is less complicated since the data is used as a whole. Each coordinate value “x” and “y” values are saved as another data frame before designing the algorithms for each model. Later on the prediction value of corresponding coordinates is added to that data frame. In the end a map data frame is obtained. The data frame could be saved as an “csv " file and later on visualized using Python programming language, GIS tools or R programming language. To visualize the data frame using Python programming language a library called “geopandas” is used which is a useful tool for visualizing data frames. In the data frame that we obtained had 3 columns which are “x” and “y” values and “Geothermal” prediction values as 0 and 1 the pixel is predicted as geothermal has value of 1 and non-geothermal has value of 0 so the map is visualized consists of 2 different colors.

4 RESULT & DISCUSSION

The need for geothermal energy is increasing day by day all over the world. As a result, the studies for the determination of the area containing geothermal energy are also increasing. This creates a need for large-scale investment. Roads such as Drilling, Seismology, Magneto Telluric, Magnetism, and Machine Learning are methods used to identify areas containing geothermal energy. Most of these methods are traditional methods and have consequences such as high cost and environmental damage. On the other hand, it does not guarantee 100% that geothermal energy will be detected in the area where the study is conducted.

Although the need for renewable energy has increased in the world, geothermal energy is a more costly energy type compared to others. This resulted in geothermal energy not being adopted at the same pace with other renewable energy technologies due to geological, economic and technological uncertainties that led to a decrease in investments. More technological methods are needed to reduce uncertainty in exploration of geothermal resources in order to reduce the cost and risk of projects carried out in this regard. Recent developments in RS, ML and AI offer highly meaningful opportunities to exploit geothermal energy use. Current research using these methods mainly takes into account large regions and also does not significantly benefit from RS, ML, and AI integration.

In this study, an artificial intelligence algorithm has been developed using Remote Sensing technology (with satellite images). Our method combines RS, ML, and AI to forecast possible geothermal sites, which can be used not only for new field discovery but also for existing site expansions. As mentioned before we implemented three different AI algorithms using two different datasets. In this part of the study we will provide information about the results of these algorithms and the prediction map of the geothermal areas that we visualize using those algorithms. In the table below classify the results according to the data that we used. As prediction scoring we used Area Under the Receiver Operating Characteristic Curve (ROC AUC).

Table 4.1. Results of Algorithms Scores

Brady Train & Test	
CNN	0.93
SVM	0.96
Decision Tree	0.99
Brady Train & Desert Peak Test	
CNN	0.63
SVM	0.73
Decision Tree	0.72
Desert Peak Train & Test	
CNN	0.91
SVM	0.92
Decision Tree	0.99
Desert Peak Train & Brady Test	
CNN	0.68
SVM	0.72
Decision Tree	0.66

Since our computer that we used to implement these algorithms could not see the optimum results of CNN algorithm so we used only a small partition of the data which is 2400 pixels but overall these three algorithms showed significant results. We want to use different data for testing our models which showed us the real scoring when exploration of geothermal areas. In

the figures below we provided all the prediction maps that we visualize. These maps are showing us the geothermal areas in the two different places Brady and Desert Peak.

4.1. Brady Train & Test Prediction Maps

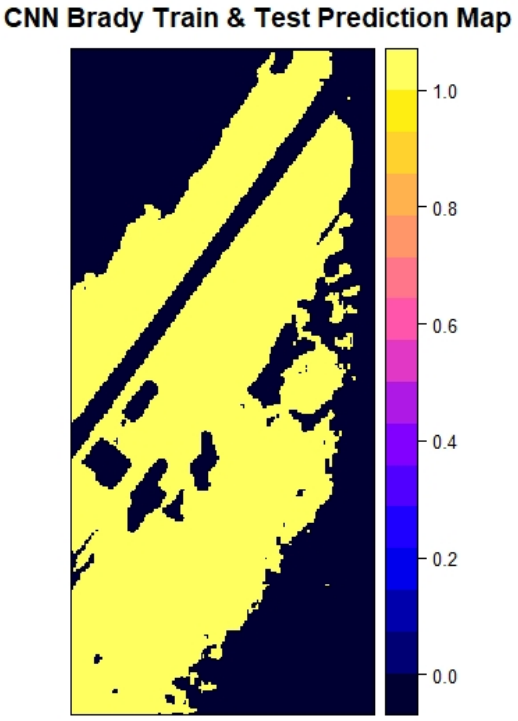


Figure 4.1. CNN prediction map

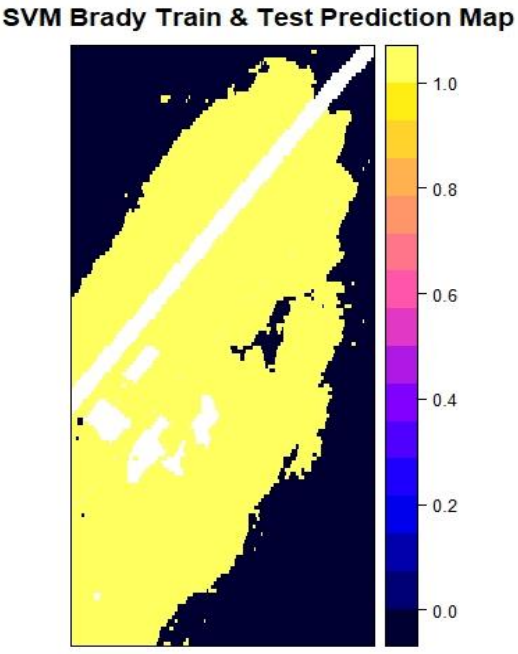


Figure 4.2. SVM prediction map

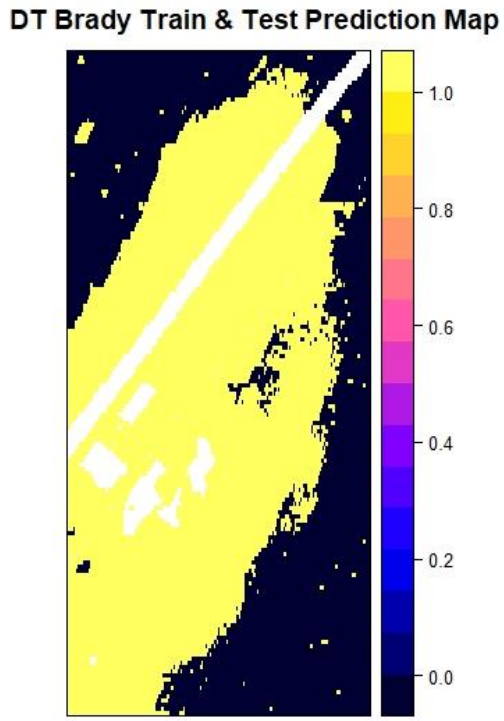


Figure 4.3. Decision Tree prediction map

4.2. Brady Train & Desert Peak Test Prediction Maps

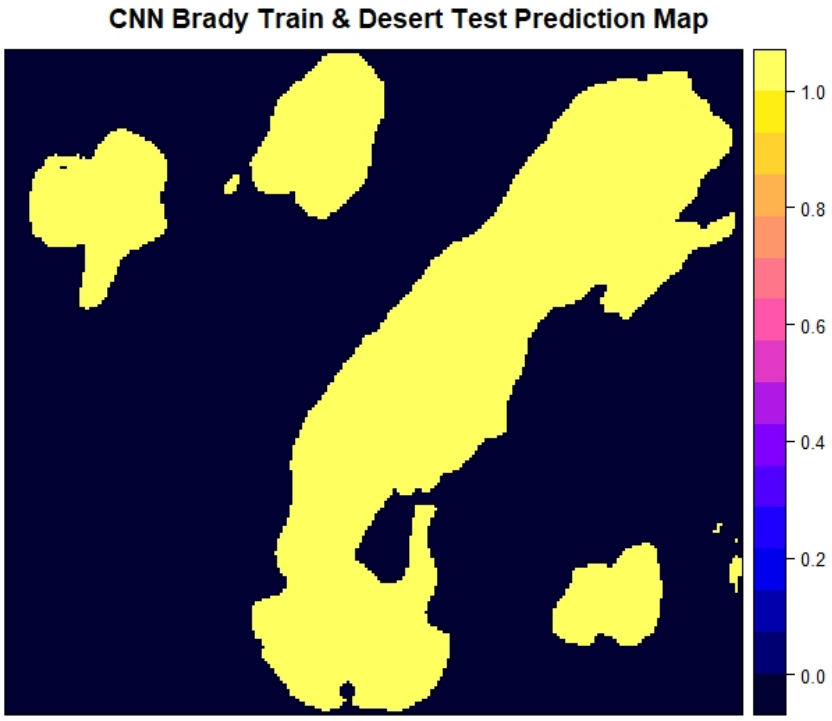


Figure 4.4. CNN prediction map

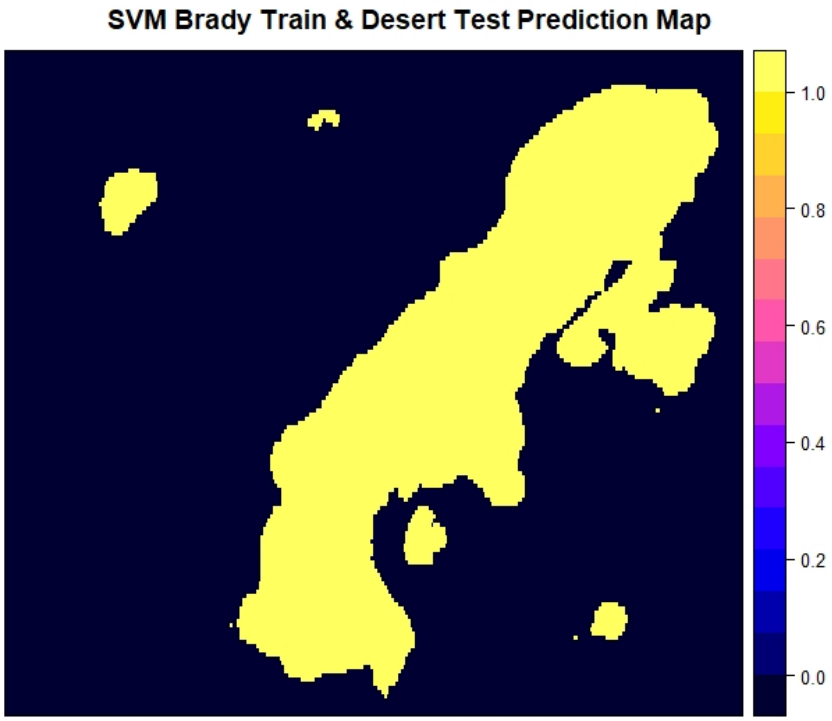


Figure 4.5 SVM prediction map

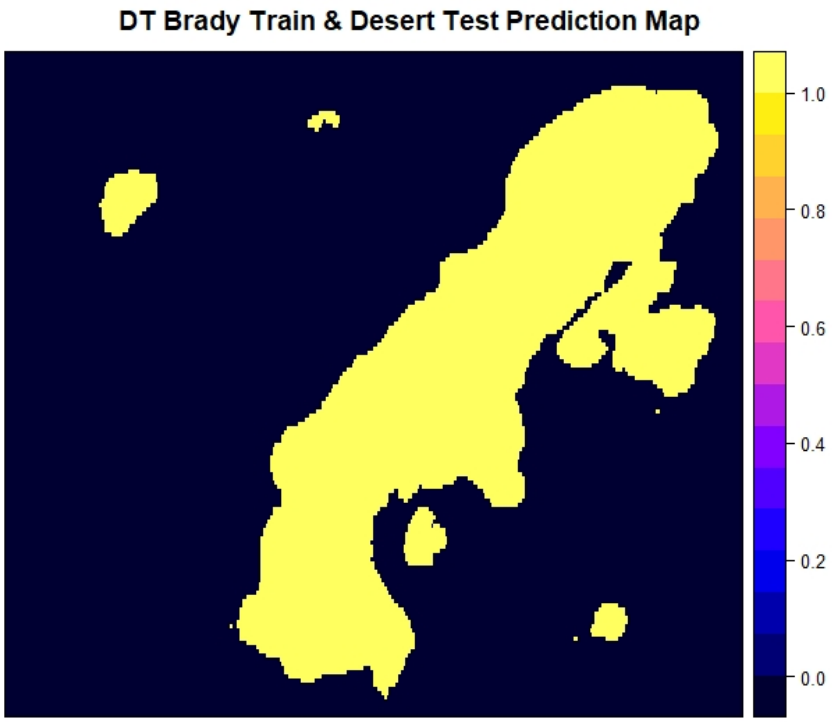


Figure 4.6. Decision Tree prediction map

4.3. Desert Peak Train & Test Prediction Maps

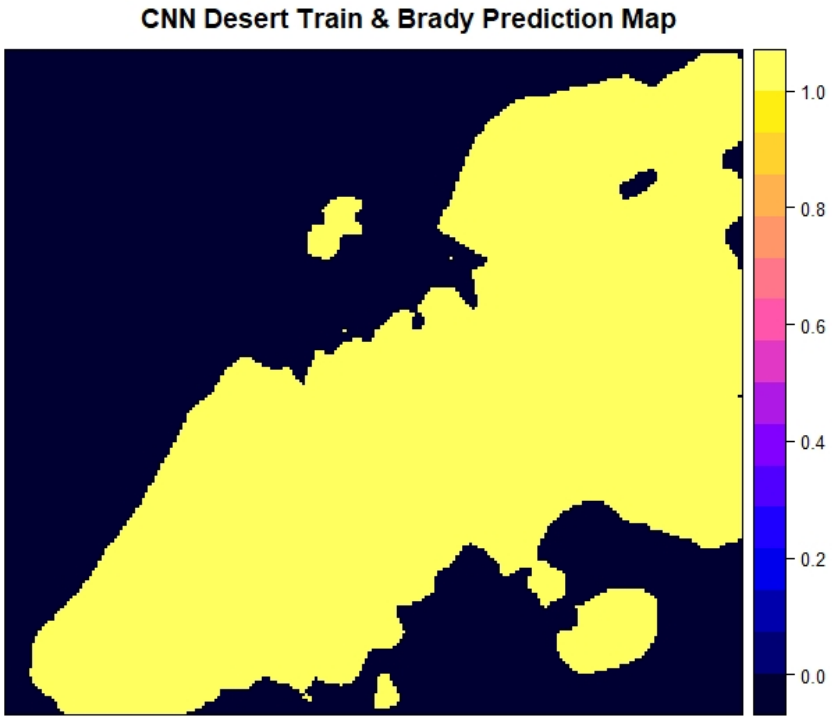


Figure 4.7. CNN prediction map

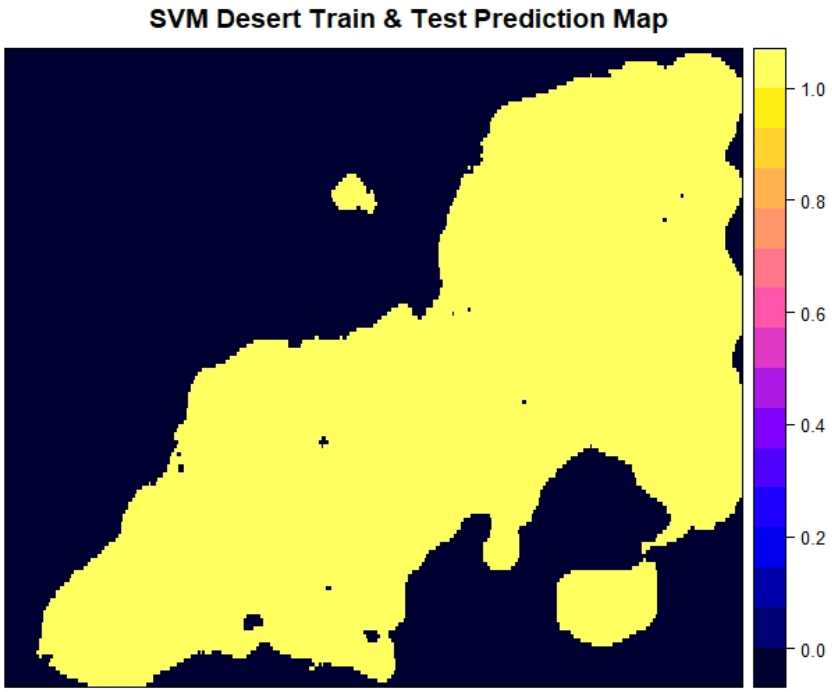


Figure 4.8. SVM prediction map

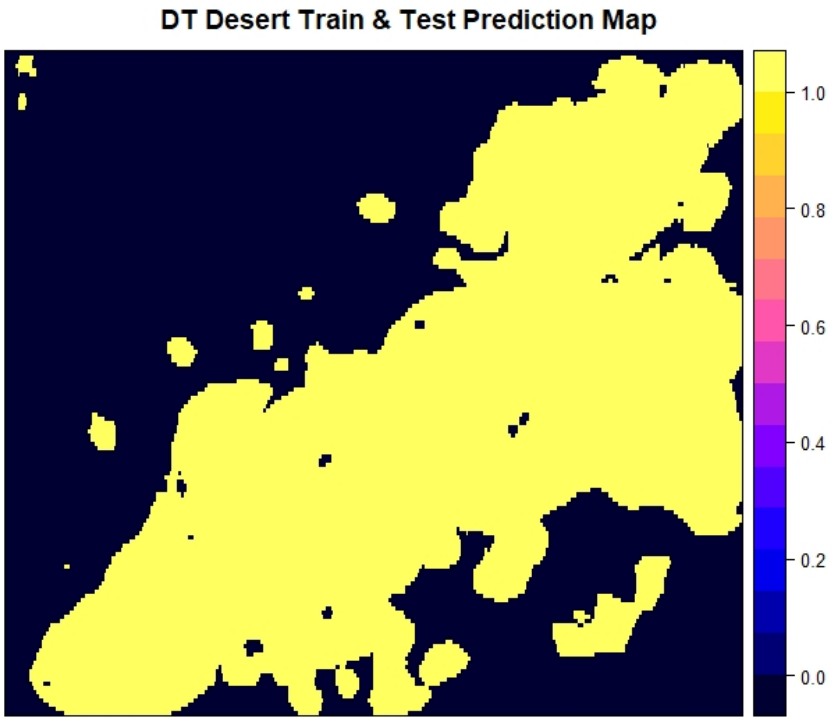


Figure 4.9. Decision Tree prediction map

4.4. Desert Peak Train & Brady Test Prediction Maps

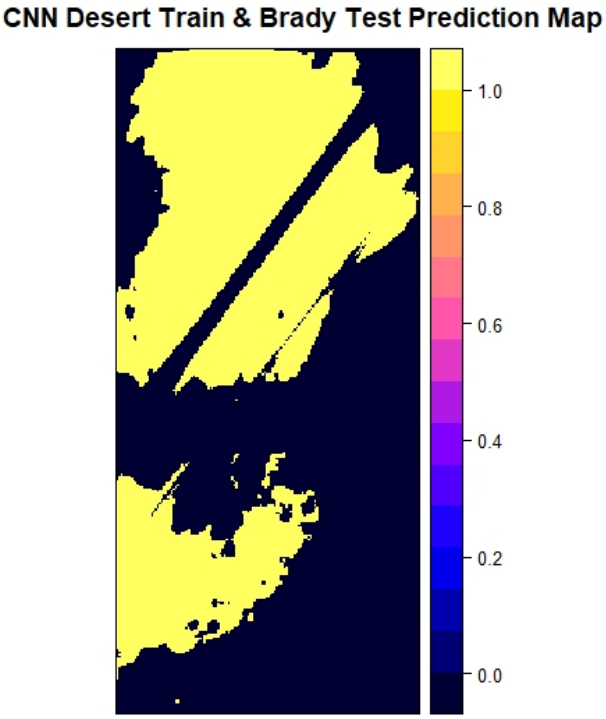


Figure 4.10. CNN prediction map.

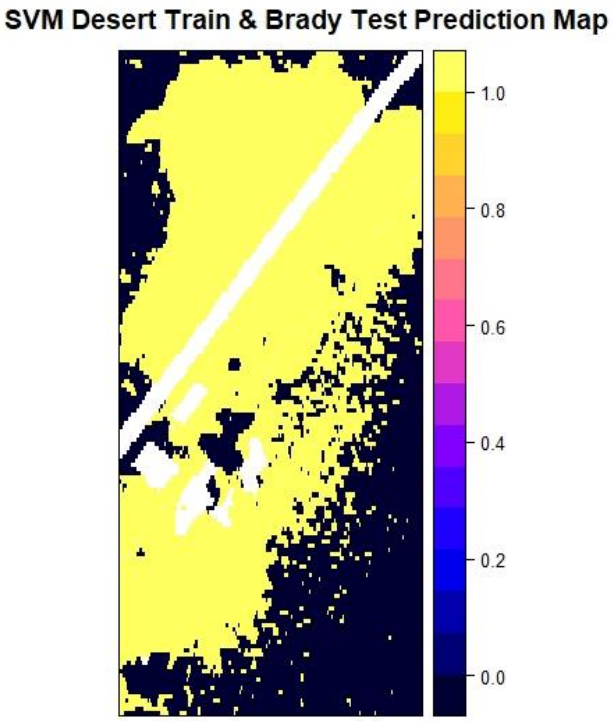


Figure 4.11. SVM prediction map

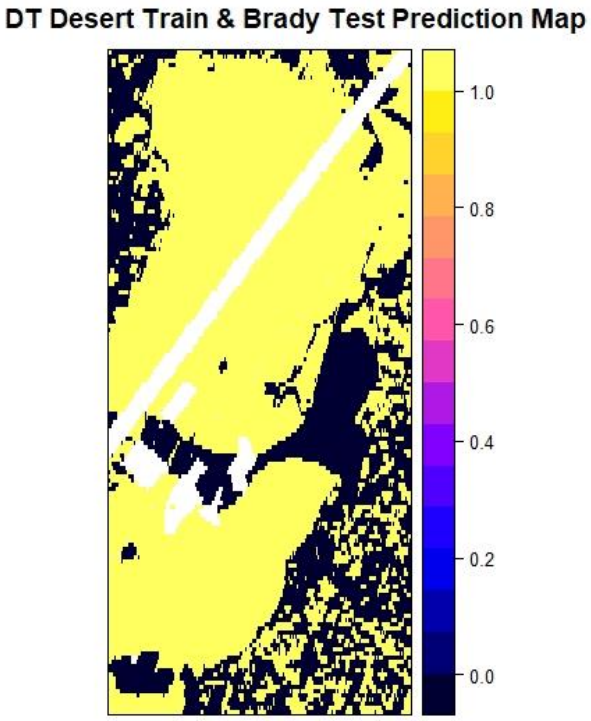


Figure 4.12. Decision Tree prediction map

Also we visualize these maps using QGIS environment in this step we plot the prediction maps on to the satellite view of the The Brady Hot Springs Geothermal area.

CNN Brady Train Maps

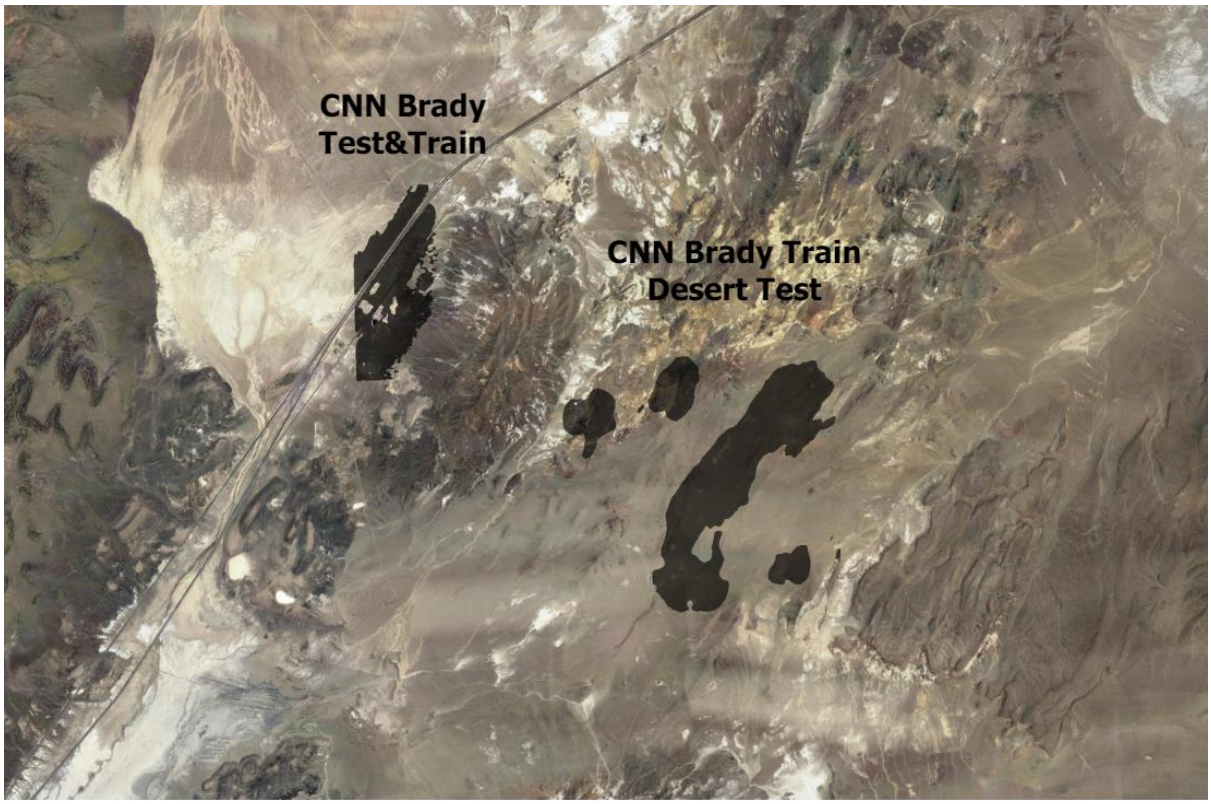


Figure 4.13. CNN Brady Train Maps

SVM Brady Train Maps

Figure 4.14. SVM Brady Train Maps

Decision Tree Brady Train Maps

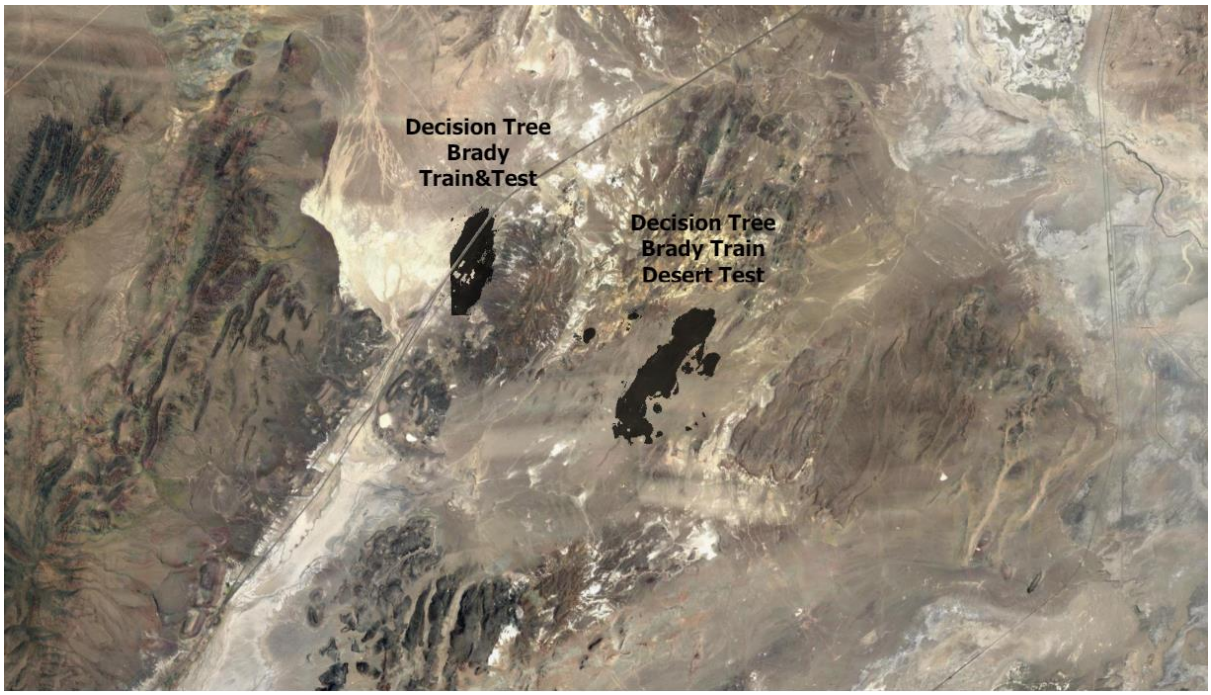


Figure 4.15. Decision Tree Brady Train Maps

CNN Desert Peak Train Maps

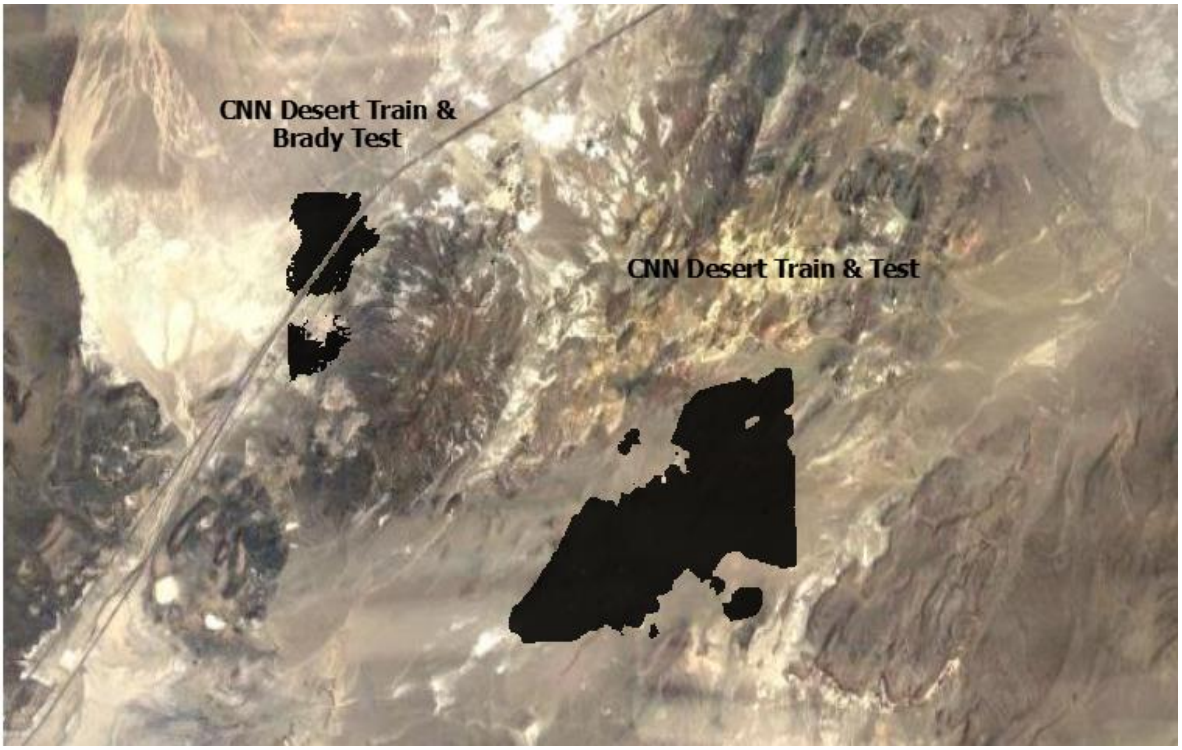


Figure 4.16. CNN Desert Peak Train Maps

SVM Desert Peak Train Maps

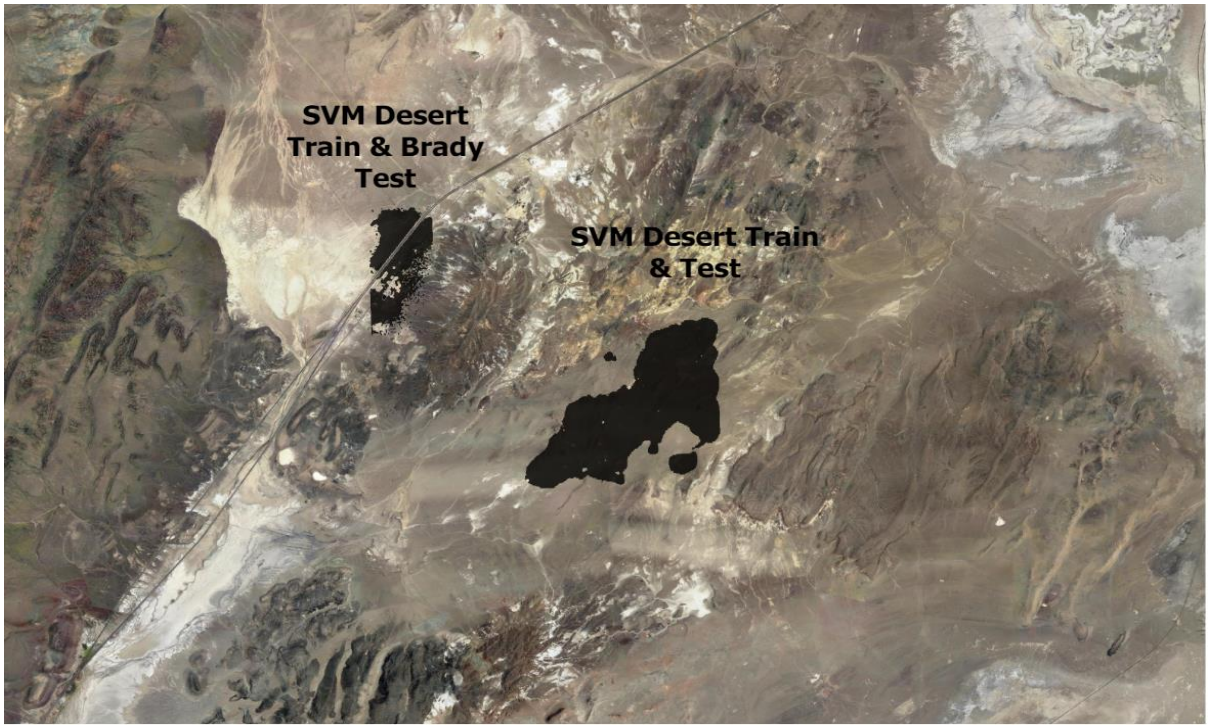


Figure 4.17. SVM Desert Peak Train Maps

DT Desert Peak Train Maps

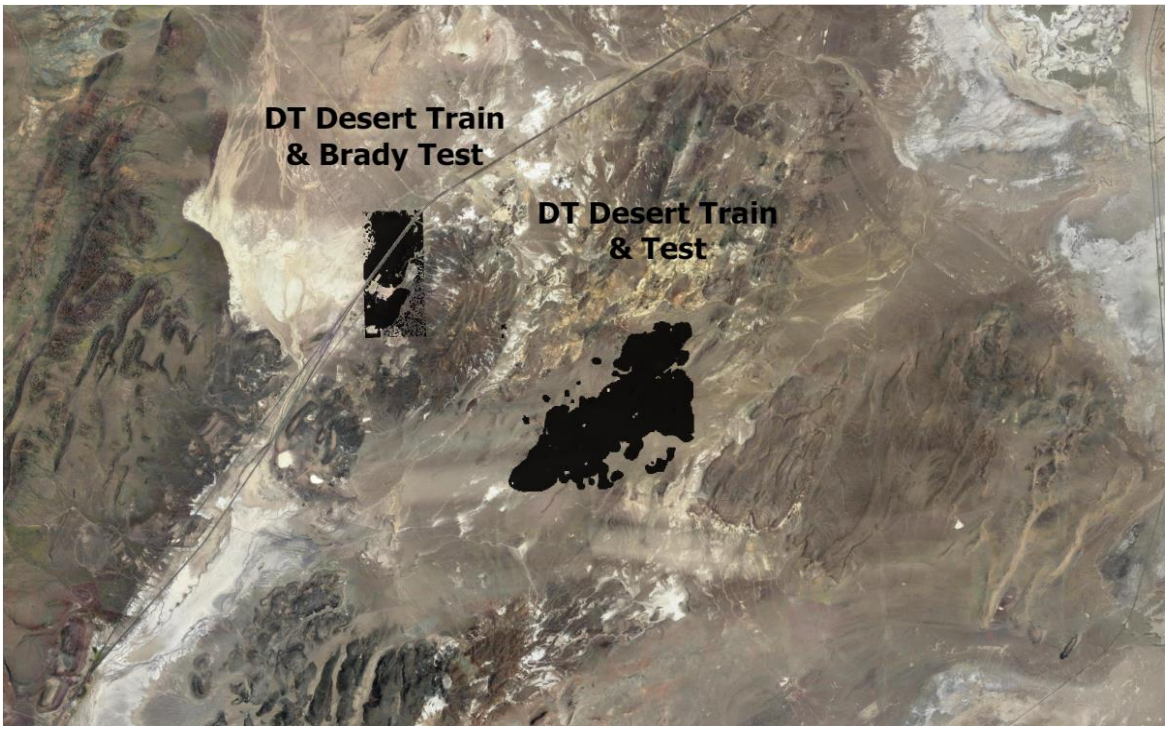


Figure 4.18. Decision Tree Peak Train Maps

Decision Tree & CNN Comparison Brady Train & Test

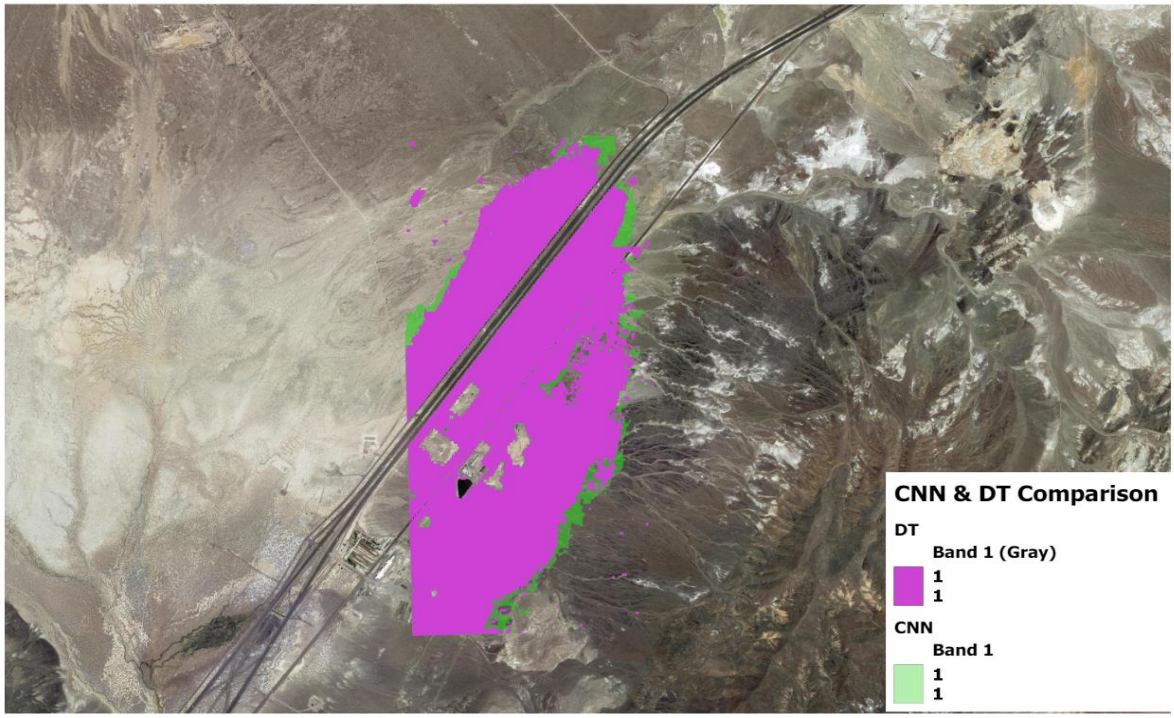


Figure 4.19. Decision Tree & CNN Comparison Brady Train & Test

SVM & CNN Comparison Brady Train & Test

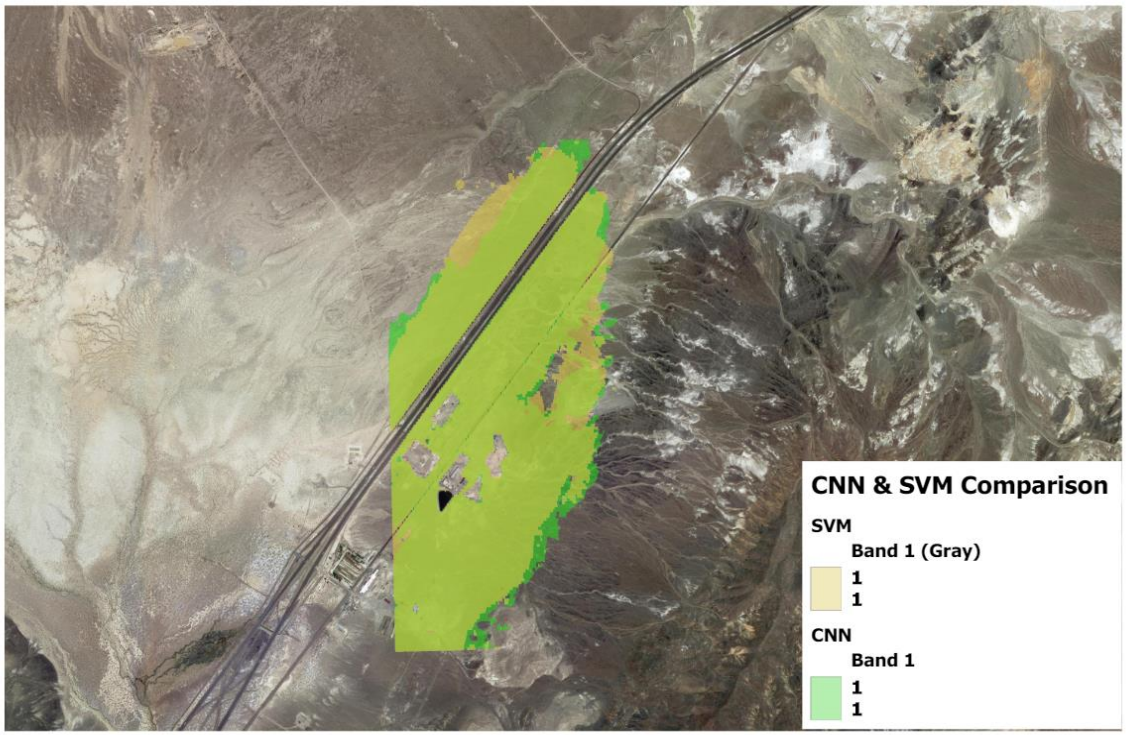


Figure 4.20. SVM & CNN Comparison Brady Train & Test

‚Brady Train & Desert Peak Test Comparison

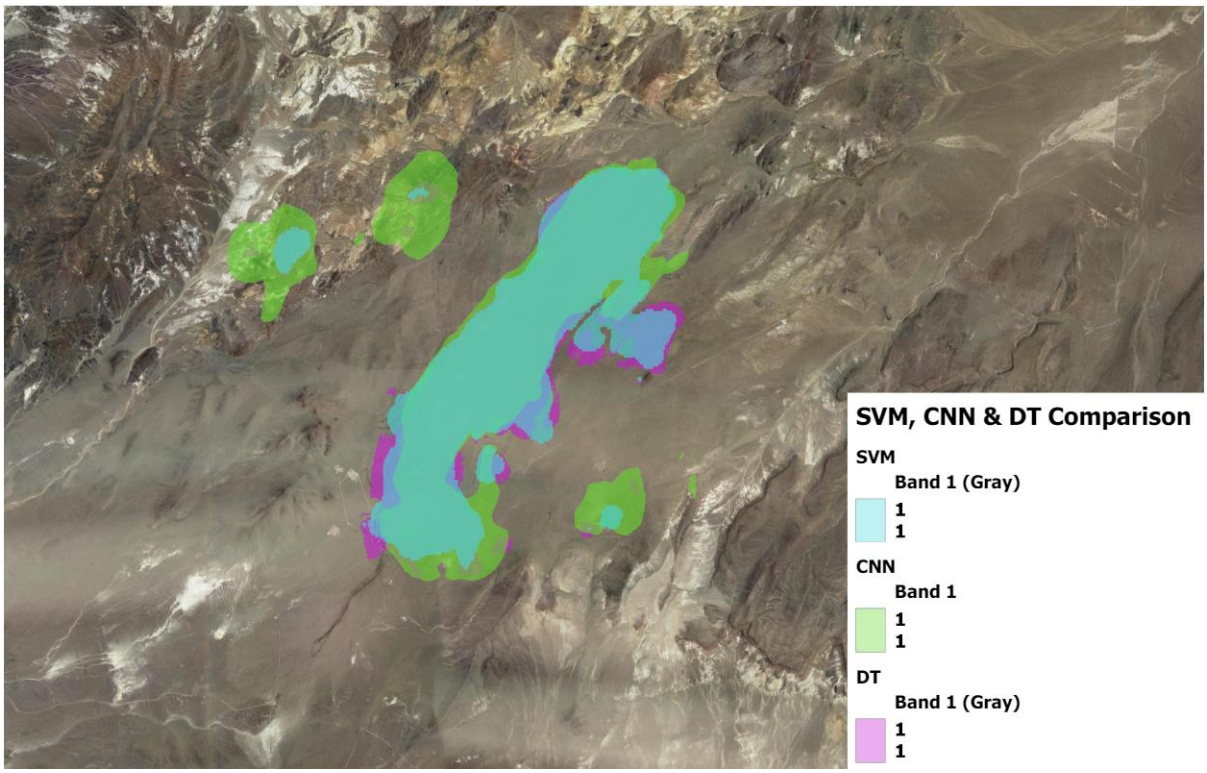


Figure 4.21. Brady Train & Desert Peak Test Comparison

Decision Tree & CNN Comparison Desert Peak Train & Test

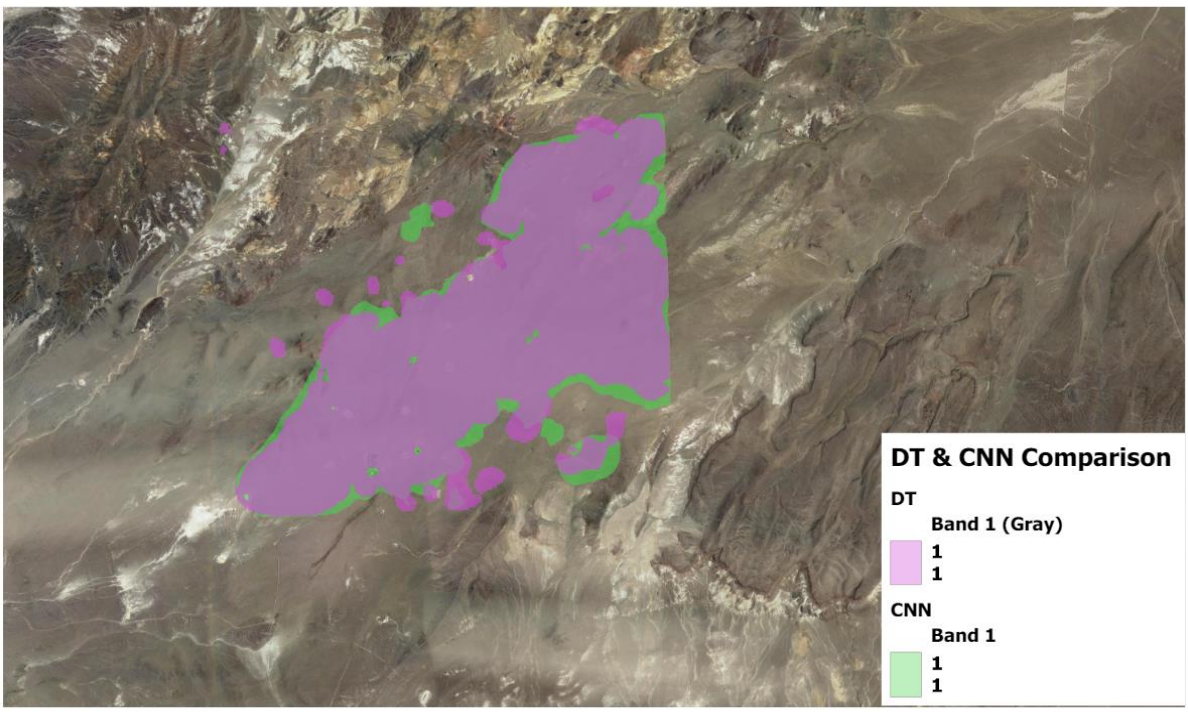


Figure 4.22. Decision Tree & CNN Comparison Desert Peak Train & Test

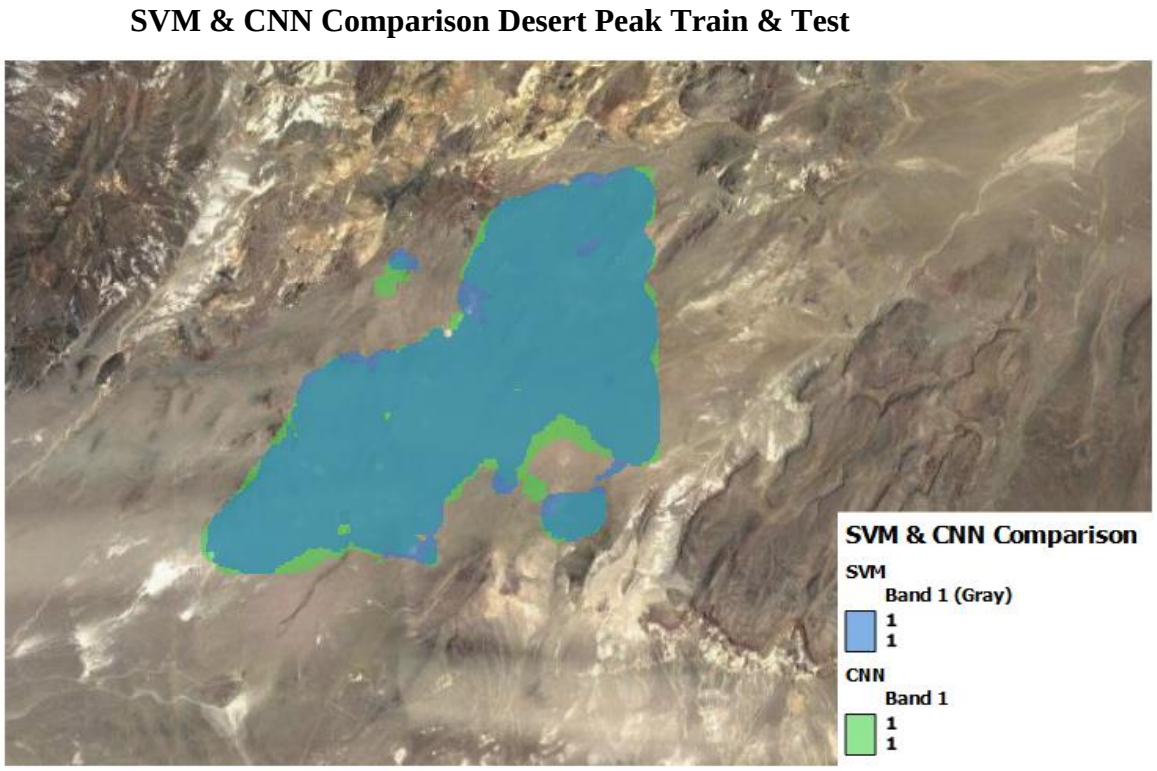


Figure 4.23. SVM & CNN Comparison Desert Peak Train & Test

5 CONCLUSION

Geothermal energy, one of the most important and endless energy resources known in our country and in the world, is used in various fields in the industry. Although this geothermal energy has an infinite resource and although it is too much to be found, it was a very time-consuming and costly process. Conventional companies have been trying to discover and find geothermal fields with techniques such as drilling and seismology, which are old techniques until recently, but recently this problem has begun to be solved with machine learning. In this study, first, we trained and tested it with Brady data with 5 stacks of satellite images. Then we have done these operations with Desert data. In this study, 3 different machine learning algorithms were created with a Brandy dataset and we got scores of 0.96 from SVM, 0.99 from Decision Tree, and 0.93 from CNN, respectively. Subsequently, the Desert dataset was also trained and tested with the same 3 machine learning algorithms. Coming to the results, we have 0.73 Accuracy values from SVM, 0.66 from CNN, and 0.72 from the Decision Tree. In the latest work, we have tried to train with the Desert dataset and test with the Brandy dataset and we have gotten results of 0.68 Accuracy from CNN, 0.72 Accuracy from SVM and lastly, 0.66 Accuracy score comes from the Decision Tree Algorithm As a result, we have obtained verbatim images of these and we have outputs even where geothermal and not with the test data we have. We believe that state or private companies will make a significant contribution to geothermal research and significantly reduce their costs with the work done.

If we come to how many beats we have covered in the project, we can say that we have reached all of our goals. At the beginning of the project, our goal was to apply a machine learning model to the Brady data we have, to map and map the area given to it.

The next step in this project would be trying to use other machine learning algorithms. It may help to increase the score which we had. With the improved version of the project, the private firms and companies can get a better solution. Also, in the future, the number of the sample could be increased and it also may help to detect geothermal areas on the test dataset. Lastly, this work will help future researchers, engineers, and mining companies that want to decrease the cost of exploring and the time.

REFERENCES

[1] Lund, W. John, "Geothermal Energy", https://www.britannica.com/science/geothermal-energy/History, 2020

[2] "What Is Geothermal Energy? How Does It Work?", https://www.twi-global.com/technical-knowledge/faqs/geothermal-energy, 2021

[3][4] Arslan S., Darıcı M., Karahan Ç, (2001), "Türkiye'nin Jeotermal Enerji Potansiyeli", Jeotermal Enerji Semineri , pp. 19-29.

[5] FRIDLEIFSSON, Ingvar, RAGNARSSON, Arni (2011), "Geothermal Conclusions", http://www.worldenergy.org

[6] "Türkiye Jeotermal Enerji Potansiyeli ve Arama Çalışmaları", https://www.mta.gov.tr/v3.0/arastirmalar/jeotermal-enerji-arastirmalari,2020

[7] Erkul, H., (2012), "Jeotermal Enerjinin Ekonomik Katkıları ve Çevresel Etkileri: Denizli-Kızıldere Jeotermal Örneği", pp. 14-18.

[8] "Geothermal Drilling Rigs", https://www.massenzarigs.com/geothermal-drilling-rigs/, May 2021

[9] Zhang L., Tianyao H., Qibin X., Liang Z., Xiangpan C., Magnetotelluric investigation of the geothermal anomaly in Hailin, Mudanjiang, northeastern China. J. Appl. Geophys.,vol 118, 2015

[10] "US awards $5.5 million for geothermal machine learning", https://www.aa.com.tr/en/energy/finance/us-awards-55-million-for-geothermal-machine-learning/25385, 2019

[11] "Harnessing heat: Applying machine learning to geothermal exploration", https://www.minesnewsroom.com/news/harnessing-heat-applying-machine-learning-geothermal-exploration , 13 Jan 2020

[12] "Machine learning seen as a means to enhance geothermal exploration and production", https://www.thinkgeoenergy.com/machine-learning-seen-as-a-means-to-enhance-geothermal-exploration-and-production/ , 2 May 2019

[13] "Pixel-based Classification", https://www.stars-project.org/en/knowledgeportal/magazine/image-analysis/algorithmic-approaches/classification-approaches/pixel-based-classification/, 2021

[14] "Maximum Likelihood", https://www.l3harrisgeospatial.com/docs/maximumlikelihood.html#:~:text=Maximum%20likelihood%20classification%20assumes%20that,threshold%2C%20all%20pixels%20are%20classified. , 2021

[15] “Maximum Likelihood Classification”, http://www.geo-informatie.nl/courses/grs20306/lectures/08imageprocessingparta/08imageprocessingparta27.htm, 2021

[16] “Minimum-Distance-To-Means Classification”, https://www.encyclopedia.com/science/dictionaries-thesauruses-pictures-and-press-releases/minimum-distance-means-classification, 2021

[17] “Minimum Distance To Means Classification”, http://www.geo-informatie.nl/courses/grs20306/lectures/08imageprocessingparta/08imageprocessingparta25.htm, 2021

[18] C.W. Kim, R. Isemoto, K. Sugiura, M. Kawatani, Structural fault detection of bridges based on linear system parameter and MTS method, J. JSCE 1 (2013) pp. 32–43.

[19] “Mahalanobis Distance”, https://www.l3harrisgeospatial.com/docs/mahalanobis.html, 2021

[20] “Object-based Classification”, http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson6-1/object.html, 2021

[21] “What is Logistic Regression?”, https://www.statisticssolutions.com/what-is-logistic-regression/, 2021

[22] https://docs.qgis.org/2.8/en/docs/gentle_gis_introduction/raster_data.html, 15 May 2021

[23] https://learngis.org/textbook/section-three-raster-data, 16 May 2021

[24]https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm, 15 May 2021

[25] https://www.ordnancesurvey.co.uk/business-government/tools-support/gis/raster-vector, 15 May 2020

[26]https://gisgeography.com/spatial-data-types-vector-raster/, 14 May 2021

[27]http://gsp.humboldt.edu/OLM_2017/Lessons/GIS/08%20Rasters/RasterToVector.html, 16 May 2021

[28]https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/an-overview-of-multidimensional-raster-data.htm, 16 May 2021,

[29] Muhagaze L (1984). Geological mapping and borehole geology in geothermal exploration. Report 5, 38 pp. Li.

[30]Renner, J. L., White, D. E., and Williams, R. L., 1975. Hydrothermal convection systems. U.S. Geol. Surv. Circ. 726, pp. 5–57.

[31]Fournier RO, Rowe JJ (1966). Estimation of underground temperatures from the silica content of water from hot springs and wet steam wells.

Am. J. Sci. (264):685-697.
[32]Macharia M.W., Gachari M. K., Kuria D. N., and Mariita N. O. (2017) Low cost geothermal energy indicators and exploration methods in Kenya. pp. 5-10.
[33]Wanjohi A (2012). Geophysical Mapping - Introduction. Short course VIII on exploration for geothermal resources.
[34]Bowen R (1989). Geothermal Exploration. In R. Bowen (Ed.), Geothermal Resources. Dordrecht: Springer Netherlands. pp. 117-167.
[35]Öngür , H., "JEOTERMAL SAHALARDA JEOLOJİK VE JEOFİZİK ARAMA İLKE VE STRATEJİLERİ", pp. 2-14, 2005.
[36]Silvestri M., Marotta E., Buongiorno M.F., Avvisati G., Belviso P., Sessa E.B., Caputo T., Longo V., De Leo V. and Teggi S., "Monitoring of Surface Temperature on Parco delle Biancane (Italian Geothermal Area) Using Optical Satellite Data, UAV and Field Campaigns", pp. , 2020.
[37]Korkmaz Basel E.D., Serpen U., Satman A., "TURKEY'S GEOTHERMAL ENERGY POTENTIAL:" pp. 1-10, 2010.
[38]Şener, Ç., Erdoğan A.R., Özgüler, M.E., 1986. Türkiye'deki jeotermal alanların araştırılmasında jeofizik çalışmalar. MTA Dergisi. Sayı:107. 152- 169.
[39] [40] Nugroho U. C., Domiri D. D., "IDENTIFICATION OF LAND SURFACE TEMPERATURE DISTRIBUTION OF GEOTHERMAL AREA IN UNGARAN MOUNT BY USING LANDSAT 8 IMAGERY", pp. 17-20, 2015.
[41] Corbel, S., Schilling, O., Horowitz, F.G., Reid, L.B., Sheldon, H.A., Timms, N.E.,Wilkes, P., "IDENTIFICATION AND GEOTHERMAL INFLUENCE OF FAULTS IN THE PERTH METROPOLITAN AREA, AUSTRALIA", pp. 2-5, 2012.
[42]Corbel, S., Schilling, O., Horowitz, F.G., Reid, L.B., Sheldon, H.A., Timms, N.E.,Wilkes, P., "IDENTIFICATION AND GEOTHERMAL INFLUENCE OF FAULTS IN THE PERTH METROPOLITAN AREA, AUSTRALIA", pp. 4-8, 2012.
[43]Moeck IS, Beardsmore G. A new 'geothermal play type' catalog: streamlining exploration decision making. In: Proceedings, thirty-ninth workshop on geothermal reservoir engineering, Stanford University, vol. 39. 2014. p. 1–8.
[44] Siler, D.L., Faulds, J.E., Hinz, N.H. *et al.* Three-dimensional geologic mapping to assess geothermal potential: examples from Nevada and Oregon. *Geotherm Energy* 7, 2 (2019). https://doi.org/10.1186/s40517-018-0117-0

[45]D. Juncu, Th. Árnadóttir, H. Geirsson, G.B. Guðmundsson, B. Lund, G. Gunnarsson, A. Hooper, S. Hreinsdóttir, K. Michalczewska, “Injection-induced surface deformation and seismicity at the Hellisheidi geothermal field, Iceland”, pp. 1, 2018

[46]Badur Ö., “JEOTERMAL ALANLARDA ÇÖKME - DEPREMSELLİK VE SAR GİRİŞİMÖLÇER ÇALIŞMALARI”, pp. 3-4 ,2011.

[47]Hutnak, M., Hurwitz, S., Ingebritsen, S.E., Hsieh, P.A., 2009. Numerical models of caldera deformation: effects of multiphase and multicomponent hydrothermal fluid flow. J. Geophys. Res. Solid Earth 114,

[48]Keiding M., Hooper A., Arnadottir T., Jonsson S., Decriem J., “NATURAL AND MAN-MADE DEFORMATION AROUND GEOTHERMAL FIELDS ON THE REYKJANES PENINSULA, SW ICELAND” pp. 1-4, 2010.

[49] Cavur M., Moraga J., Duzgun S.H., Soydan H., Jin G., “Displacement Analysis of Geothermal Field Based on PSInSAR And SOM Clustering Algorithms A Case Study of Brady Field, Nevada—USA”, pp. ,2021.

[50]Badur Ö., “JEOTERMAL ALANLARDA ÇÖKME - DEPREMSELLİK VE SAR GİRİŞİMÖLÇER ÇALIŞMALARI”, pp.11-16 ,2011.

[51]Moraga, J., Duzgun, H.S., The Geothermal Artificial Intelligence for geothermal exploration, pp 7-9., Nov 2021

[52]Öngür , H., “JEOTERMAL SAHALARDA JEOLOJİK VE JEOFİZİK ARAMA İLKE VE STRATEJİLERİ”, pp. 2-14, 2005.

[53]Lynne BY, Campbell KA. Diagenetic transformations (opal-A to quartz) of low- and midtemperature microbial textures in siliceous hot-spring deposits, Taupo Volcanic Zone, New Zealand. Canadian Journal of Earth Sciences. 2003 Nov 1;40(11):1679–96.

[54]Huntington JF. The Role of Remote Sensing in Finding Hydrothermal Mineral Deposits on Earth. Ciba Foundation symposium. 1996 Feb 1;202:214–31; discussion 231.

[55] Calvin, W., Coolbaugh, M., and Vaughan, R.G., 2002. “Geothermal Site Characterization Using Multi- and Hyperspectral Imagery.” Geothermal Resources Council Transactions , v. 26, p. 483-485.

[56] Junko H., Noriyoshi T., “Coupled Thermo-Hydro-Mechanical-Chemical Processes in Geo-Systems" in Elsevier Geo-Engineering Book Series, 2004.

[57]Vaughan, R.G., W.M. Calvin and J.V. Taranik, SEBASS hyperspectral thermal infrared data: Calibrated surface emissivity and mineral mapping, Remote Sensing of Environment, submitted, April 2002.

[58] https://www.mygreatlearning.com/blog/open-source-python-libraries/, 11 Sep 2020

[59] https://en.wikipedia.org/wiki/Decision_tree_learning, 19 May 2021

[60]https://bdtechtalks.com/2020/01/06/convolutional-neural-networks-cnn-convnets/#:~:text=Convolutional%20neural%20networks%2C%20also%20called,a%20postdoctoral%20computer%20science%20researcher , 6 Jan 2020

[61] https://en.wikipedia.org/wiki/Convolutional_neural_network, 18 May 2021

[62]https://www.analyticssteps.com/blogs/how-does-support-vector-machine-algorithm-works-machine-learning , 4 May 2020

[63]https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47, 7 Jun 2018

[64]Warner, T. A. (2017). Nature of Multispectral Image Data. The Geographic Information Science & Technology Body of Knowledge (3rd Quarter 2017 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2017.3.1

[65] https://numpy.org/doc/stable/user/whatisnumpy.html, 15 January 2021

[66]https://monkeylearn.com/unstructured-data/#:~:text=Unstructured%20data%20(also%20known%20as,images%2C%20video%2C%20and%20audio., 25 Feb 2021

[67] https://www.tensorflow.org/guide/keras/masking_and_padding , 18 April 2021

[68]https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/ , 17 April 2020

[69]https://heartbeat.fritz.ai/classification-with-tensorflow-and-dense-neural-networks-8299327a818a#:~:text=What%20is%20a%20dense%20neural%20network%3F&text=Each%20neuron%20in%20a%20layer,those%20in%20the%20next%20layer., 8 Feb 2021

[70] https://gdal.org/, 18 Apr 2021

[71] https://www.w3schools.com/python/numpy/numpy_intro.asp, 11 March 2021

[72] https://machinelearningmastery.com/tutorial-first-neural-network-python-keras/, 15 September 2021

[73] https://en.wikipedia.org/wiki/Scikit-learn, 16 May 2021

[74] https://en.wikipedia.org/wiki/Pandas_(software), 19 May 21

[75] https://rasterio.readthedocs.io/en/latest/, 18 May 2021

[76] https://geopandas.org/, 17 May 2021