SUMMARY OF THESIS (Original Thesis: https://github.com/eozdemir24/thesis)

This study was carried out with two different data sets named Brandy and Desert Peak and it was tried to predict whether there is geothermal energy in the area by training the satellite image with classification algorithms.

As it is known on the business side, there are many types of energy in the world, they are divided into two types: fossil energy types and renewable energy types. For example, coal and oil are fossil resources, while wind, solar or geothermal energy are considered renewable energy sources.

From today until now, geothermal energy has been found in various ways, mainly by seismology and drilling.

Seismology: 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.

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.

As it can be seen, these old-style and traditional methods are very costly for energy companies and also take a serious amount of time. For this reason, energy companies and states that have invested enough in information systems have switched to computerized methods or are trying to pass. These methods are classification method and methods and machine learning

Popular ones of classification methods are *pixel based classification maximum likelihood method, Mahalanobis Method* and *object-based Methods*.

Machine Learning Solutions (WHAT WE IMPLEMENTED)

Classifcation methods:

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

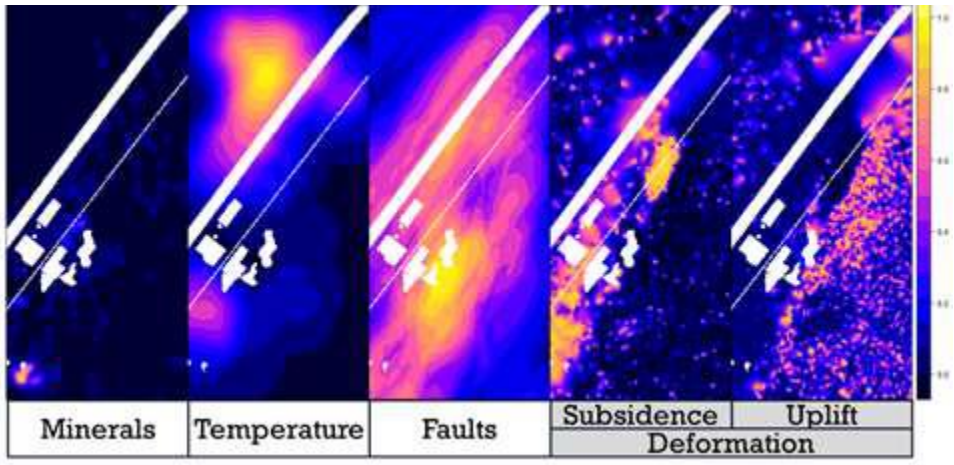
It uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). [1]

CNN (Deep Learning Solution): 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.

SVM: 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

Data

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]



Features of Data

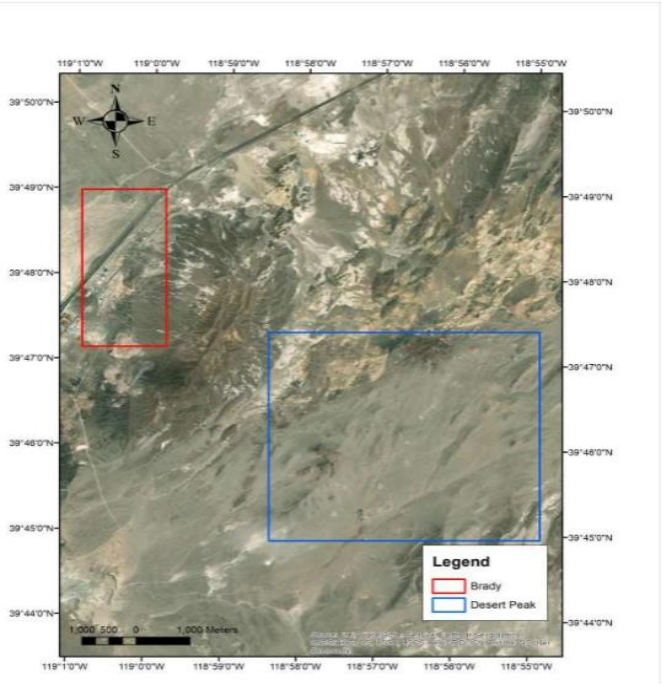
Minerals: In addition to the chemistry of brine that interacts with the bedrock in which liquids circulate, the reactivity of the minerals found in these bedrock changes significantly. This provides an important advantage in geothermal field detection.

Temperature: Why we choose temperature as an indicator since thermal manifestations such as fumaroles, steaming grounds, hydrothermally altered grounds, hot springs, volcanic gas vents, craters, and mud pools are typically found at sites of high-temperature geothermal energy resources.

Faults: 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.

Subsidence: 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.

Uplift: Uplift movements are also strongly related to the structural geology of the site. To this end, identifying deformation patterns and understanding their temporal and spatial changes provides invaluable information that can be related to the operational activities in the geothermal site to characterize the geothermal reservoir and understand the operational impacts in the region. [2]



Desert Peak 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

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		

The project was made using Python and R language and the libraries used are as follows

Gdal,Numpy,Keras,Sklearn,Pandas,Rasterio,Geopandas

Implementation:

CNN

In order to implement CNN, images must first be reshaped, and an image patch must be made for this process. In the image patching process, some parameters are used, these are the kernel size "17x17" is the size of the patches and the kernel channels with "5", these channels represent the dimensions of the image used. The image used for this study has 6 layers and 5 of them are used as attributes and 1 of them is used as a label. Then, it was divided into 30% and 70%, and it was trained and tested.

Brandy data consists of 542897 pixels, while Desert Peak data consists of 2400 pixels.

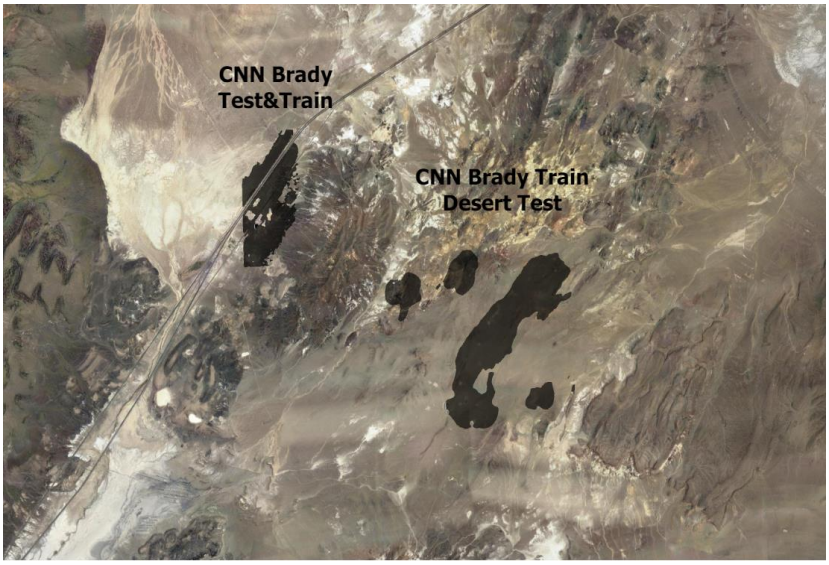
Decision Tree and SVM

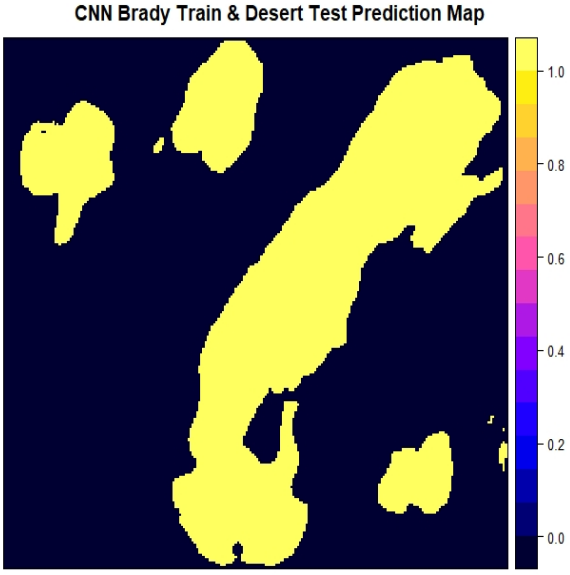
Two known classification algorithms SVM and Decision Tree were used to compare and contrast the CNN algorithm. Firstly, with the help of R Studio, the dataset, which was Unstructured, was transformed into a Structure to use SVM and Decision Tree on the data. Since our data is a raster, this process was done with "Raster", one of R Studio's libraries, and a data frame was obtained. Afterwards, the process was continued with Python on Jupiter again, and SVM and Decision Tree algorithms were trained and tested with both data sets. I should also say that 30 percent train is reserved as 70 percent test here.

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

Example of The mapped version of the estimate.





Reference:

[1] https://en.wikipedia.org/wiki/Decision_tree_learning ,14 April 2022

[2] 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.