International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Performance Evaluation of Machine Learning on Forest Fire

Sonia Christy[1], Dr. Clara Shanthi D.[2]

[1]Department of Computer Science, Mount Carmel College Autonomous, Bengaluru, India

[2]Assistant Professor, Department of Computer Science, Mount Carmel College Autonomous, Bengaluru, India

Abstract: ***As the technology is developing, Forest fires are one of the major environmental concerns, each year millions of hectares are destroyed over the world, causing economic and ecological damage as well as human lives. It is not always possible to cope with a fire and get to the forest in a timely manner. These disasters can cause widespread damage and destruction, so they must be identified quickly and accurately. Which is why, an early warning system can provide effective strategic data and accurate prediction results tailored for interference and fire control. This success comes with challenges and problems that are comparable with limited resources. At the moment, they have little or nothing to do with the level of risk analysis mechanisms involved. In general, the fire prediction models used today offer neither satisfaction nor accuracy. Factors that affect the prevalence of fires are the frequency and variety of fires, which are shockingly rare. This research provides an insight into the use of Machine Learning and Deep Learning models towards the occurrence of forest fire. Techniques such Random Forest, K - nearest neighbors, Artificial Neural network, Convolutional neural network has been used for prediction of forest fires. Random Forest model showed the highest forest fire prediction accuracy.***

Keywords: Forest Fire, Machine Learning (ML), Deep Learning (DL), Random Forest (RF), K - nearest neighbors (KNN), Artificial Neural network (ANN), Convolutional neural network (CNN)

1. Introduction

Forest fires are one of the environmental problems that cause economic and ecological damage. They are a concern because they cause a lot of damage to the environment, property and human lives, the impact of wildfires, countries impact economic and social factors. Several causes are the source of this problem such as ignorance of people who are exposed to forests, global warming and natural factors. Therefore, it is important to detect forest fires at an earlier stage. This can help save the flora and fauna in the area as well as resources [1]. In addition, it can help control the spread of fires in the early stages. The task of forest monitoring faces many difficulties due to its large territory and forest density. Today, there are various fire modeling technologies to predict the spread of a fire, such as physical modeling and mathematical modeling. These models rely on data collection during wildfires, simulations, and laboratory experiments to identify and predict fire development in multiple regions.

Although there are many mature forecasting systems around the world, there are variations in climate in different regions. Even within the same forest area, in different districts there are differences in fuel, topography and other factors. Therefore, the parameterization of the model depends on the assistance of experts when applying regionalization to the predictive model. The complexity of the technology limits its wide applications. In this paper, ML and DL algorithms are applied to predict forest fire risk and comparative studies between algorithms are performed.

2. Background

The ability to accurately predict areas likely to be affected by wildfires can help optimize fire management efforts. Wildfires exhibit different behaviors depending on their origin and the factors that affect it with combustible materials. Fossil coals indicate that wildfires began shortly after terrestrial vegetation appeared about 419 million years ago during the Silurian period. The occurrence of wildfires throughout the history of life on Earth suggests that fire must have had marked evolutionary effects on the flora and fauna of most ecosystems. The Earth's carbon - rich vegetation, seasonally dry climate, atmospheric oxygen, along with the spread of lightning and volcanic eruptions create favorable conditions for fires.

Scientists are creating computer models to predict the likelihood of wildfires in a range of potential climate futures. Using different temperature and precipitation projections, scientists predict where and when wildfires are most likely to occur. Scientists also predict that the bushfire season will be longer, extending from spring through summer and into the fall months. Using computer models and geographic information systems (GIS), scientists are creating maps that show which areas will be more vulnerable to wildfires. These areas can therefore be specifically managed to reduce the risk of wildfires.

A. Problem Statement

- The data set for fire occurrence is sparser since most of its data sets' values revolve around zero
- Early prediction measures varies from Satellite Image Processing to Wireless Sensor Networks to some Prediction models

B. Objective

- To make a comparative study of algorithms and analyze which algorithm predicts better and is efficient with respect to time and space complexity
- To understand whether machine learning or deep learning algorithms perform better

3. Literature Survey

Initial predictive work ranges from satellite imagery to wireless sensor networks and several predictive models. Two - class machine learning algorithms are compared and evaluated based on precision, recall, Fscore, accuracy, and subcurve area for early fire detection. The proposed framework aims to analyze images of forest areas and simultaneously and continuously detect various parameters such as temperature, relative humidity, and carbon monoxide (CO) levels throughout the day [1]. Traditional approaches to fire monitoring, nevertheless, frequently fail to identify a fire in the moment [2].

Forecasting models based on geographic and meteorological explanatory variables help predict the area of fire spread from local fires. This deep learning goes beyond using images to achieve object recognition, clear audio will help optimize the predictions made for our wildfire problem [3]. All models and hyper parameters from the Bayesian optimization process are preserved, and fast integral learning is performed at a later stage to reduce model error rates and improve overall performance of the model [6]. The systems use charge - integrated cameras and infrared detectors, satellite systems, wireless sensor networks (WSNs) and unmanned aerial vehicles (UAVs).

The most widely used ANN model is a multi - layer ANN that is monitored as a single unit or integrated into WSN - based fire monitoring systems [5]. A machine learning approach for estimating time - resolved spatial evolution of wildfire fronts using the Deep Convolution Inverse Graph Network (DCIGN). It turns out that the complexity of this method is significantly better than the computational model of non - uniform spatial conditions, which shows a reduction in simulation time. The DCIGN method helps predict the spread of fires across homogeneous and heterogeneous landscapes with simple and complex terrain [12].

4. Methodology

A. Process Flow

The below flowchart shows the complete flow of the whole project as per how it was carried out. A total of 12 research papers were studied and then the relevant machine learning models were studied. Then the data was cleaned, pre - processed, and used for model building. A comparative analysis of the models was done with respect to accuracy, time complexity and space complexity. The model that performed better was chosen.

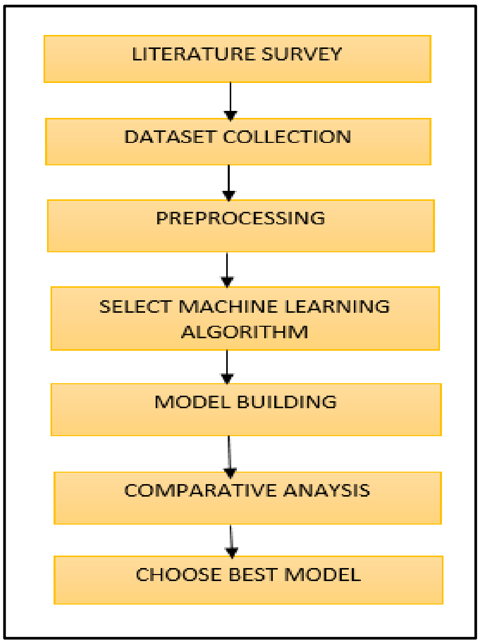

Figure 1: Process Flow Diagram

B. Data Collection and Pre - processing

The data used in this article is from kaggle. The dataset contains 5000 Fire and Unburned images. The dataset is divided into two folders called test data and training data. The test data consists of a combination of about 50 fire and unburned images. The ship's data contains 5000 combined images of fire and non - fire. Each of these folders is divided into two subfolders, Fire and Not Fire, each of these subfolders of test data contains 25 images, and each of these subfolders of train data contains 2500.

After data collection data pre - processing takes place in which dataset to be formed in standard format. The image dataset is converted to. csv file (comma - separated values) using ML python codes to extract features from the image using 1st order statistical moments (mean, variance, kurtosis, skewness), Gray Level Co - occurrence Matrix (GLCM) features (contrast, dissimilarity, homogeneity, energy, asm, correlation), Local Binary Pattern (LBP) feature (lbp_energy, lbp_entropy), Gabor (gabor_energy, gabor_entropy). The dataset is split as Train Dataset. csv and Test Dataset. csv files.

C. System Architecture

System architecture or system architecture is a computational paradigm that describes the structure, behavior and views of the system. System architecture may be composed of system modules and subsystems that will collaborate to execute the overall system.

Points regarding the system design are given below: -

- An architectural diagram is a diagram of the system used to explain the overall outline of the software system and the interactions, limits and boundaries between the elements.
- After data collection and data preprocessing a suitable algorithms are taken and model is built, based on the Train Dataset and tested with the Test Dataset

- Then the Result Analysis is done, for which each model is compared with accuracy, time complexity, space complexity respectively, and the best model is chosen

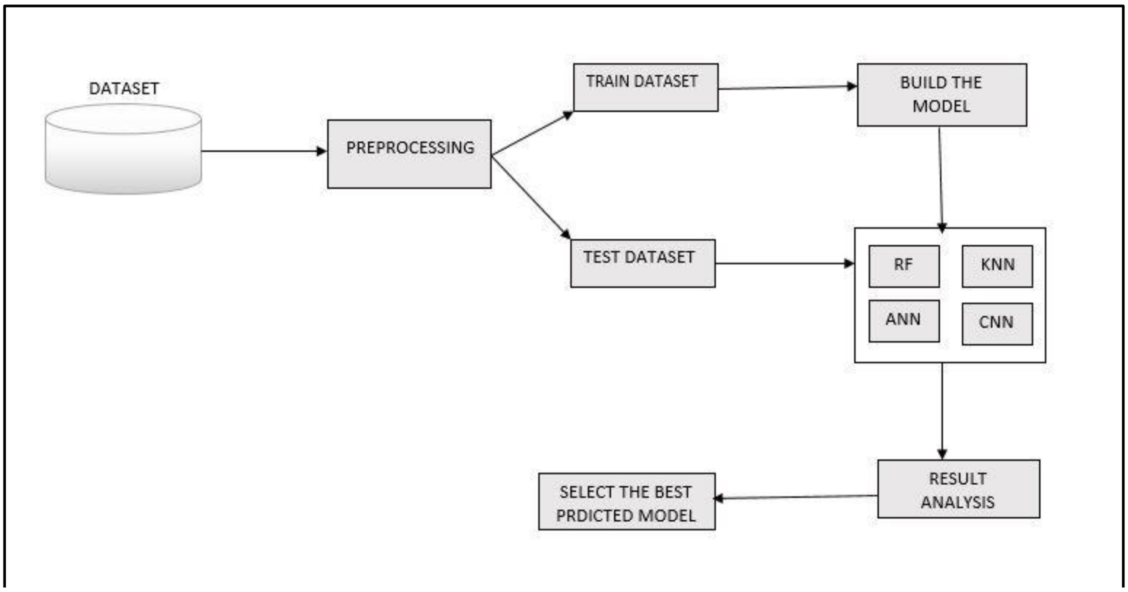

Figure 2: System Architecture

D. Machine Learning Models

Machine learning means that computers learn from data using algorithms to perform a task without being explicitly programmed. Deep learning uses a complex structure of algorithms modeled on the human brain The machine learning methods used in this article are Random Forest (RF) and K - Nearest Neighbor (KNN). Also, the deep learning used is artificial neural networks (ANN) and legitimate neural networks (CNN).

- Random Forest: Random forests are generated from a subset of the data, and the final result is based on an average vote or majority vote to solve the problem of overfitting. Random forests combine multiple trees to predict the class of a dataset, so some decision trees may or may not be able to predict the output correctly. But together all the trees predicted the correct yield. It is based on the concept of group learning, which combines multiple classifiers to solve complex problems and improve model performance.
- K - Nearest Neighbour: This is a supervised machine learning algorithm. This algorithm can be used to solve both classification and regression problems. The number of nearest neighbors of a new unknown variable to be predicted or classified is indicated by the symbol "K". The ANN algorithm uses the same principle. Its purpose is to find all the nearest neighbors around the new unknown data point and determine which class it belongs to. This is a distance - based approach. ANN calculates the distance of all points near unknown data and excludes the point with the shortest distance. Therefore, it is often referred to as a distance - based algorithm. To properly classify the results, you must first determine the value of K (nearest neighbor).
- ANN: Artificial Neural Networks (ANN) consists of layers with input and output dimensions. The latter is determined using a large number of neurons (also known as "nodes"). This is an arithmetic unit that connects weighted inputs via an activation function (which allows neurons to be switched on and off). The weights are randomly initialized, like the maximum of the system research algorithm, and optimized at some point in the training to reduce the loss function.
- CNN: A convolutional neural network, also known as a CNN or ConvNet, is a layer of neural networks that specializes in processing mesh topological data, such as images. The human brain processes a large amount of information the moment we see an image. Each neuron operates in its own receptive field and is connected to other neurons in such a way that they cover the entire visual field. Just as each neuron only responds to stimuli in a limited region of the visual field known as the receptive field in biological vision systems, each neuron in a CNN processes data only.

5. Experimental Results

The results obtained from the analysis performed are given below:

The dataset contains two csv files, Train Dataset. csv and Test Dataset. csv. The coding is done on a Jupyter notebook with python code. Model was built using Train Dataset. csv for each algorithm (RF, KNN, ANN, CNN) and prediction was made with Test Dataset, the confusion matrix of each algorithm was calculated with a built - in function to find the accuracy of each model. Time complexity, i. e. the time taken by the algorithm and the space complexity it uses, is also calculated.

A. Result Table

Algorithms	Accuracy	Confusion Matrix	Time Complexity (in seconds)	Space Complexity (in bytes)
Random Forest	0.892	{[224 27][27 222]}	2	9826724
KNN	0.742	{[218 96][33 153]}	2	992051
ANN	0.858	{[202 49][22 227]}	29	1046723
CNN	0.794	{[229 22][81 168]}	24	7988142

Figure 3: Train set is 50% and test set is 50%

Algorithms	Accuracy	Confusion Matrix	Time Complexity (in seconds)	Space Complexity (in bytes)
Random Forest	0.903	{[134 14][15 137]}	1	790023
KNN	0.733	{[129 60][20 91]}	1	997520
ANN	0.893	{[134 15][17 134]}	57	1181218
CNN	0.803	{[112 37][22 129]}	19	4017311

Figure 4: Train set is 70% and test set is 30%

Algorithms	Accuracy	Confusion Matrix	Time Complexity (in seconds)	Space Complexity (in bytes)
Random Forest	0.935	{[93 7] [6 94]}	1	802331
KNN	0.785	{[91 35] [8 66]}	1	1006246
ANN	0.86	{[90 9] [19 82]}	50	1195458
CNN	0.835	{[79 20] [13 88]}	19	3817310

Figure 5: Train set is 80% and test set is 20%

From the tables understand that Random Forest has highest accuracy and took less time to run the code. And while comparing between ML algorithm (RF, KNN) Random Forest performed better and with DL algorithms (ANN, CNN) Artificial Neural Network performed better.

B. Graphs

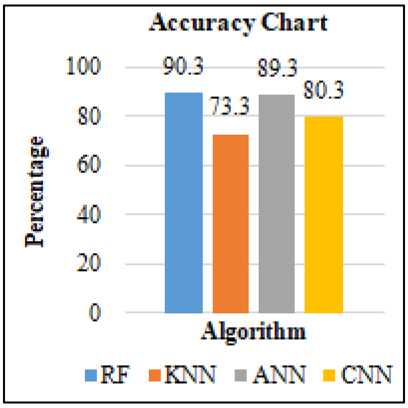

Figure 6: Accuracy Chart

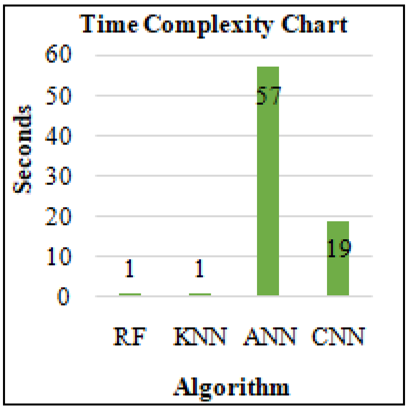

Figure 7: Time Complexity Chart

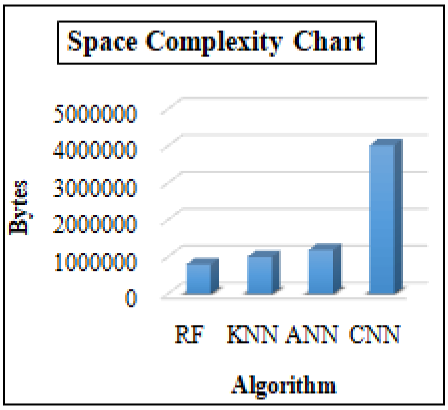

Figure 8: Space Complexity Chart

6. Conclusion

The project includes two ML algorithms: Random Forest and K - nearest neighbors, and two DL algorithms, artificial neural networks and evolutionary neural networks. The model is built to determine the accuracy of each algorithm and to predict the forest fire when input is inserted to the model. Also, to understand which algorithm is the most efficient in terms of space - time and accuracy. The results show that Random Forest performs better than other algorithms.

7. Scope and Future Enhancements

This project is currently using a. csv file for model building, training, and testing. Future scope could be working with image dataset and also working with longitude and latitude value, to predict burnt area.

References

[1] Richa Sharma, Shalli Rani, and Imran Memon “A smart approach for fire prediction under uncertain conditions using machine learning”August 2020, doi. org/10.1007/s11042 - 020 - 09347 – x

[2] Lakshmisri Surya “Risk Analysis Model That Uses Machine Learning to Predict the Likelihood of a Fire Occurring at A Given Property”, Volume 5, Issue 1 March 2017, ISSN: 2320 - 2882

[3] Zouiten Mohammed, Chaaouan Hanae, and Setti Larbi “Comparative study on machine learning algorithms for early fire forest detection system using geodata”, Vol.10, No.5, October 2020, pp.5507~5513, ISSN: 2088 - 8708, DOI: 10.11591/ijece. v10i5. pp5507 - 5513

[4] Pragati, Sejal Shambhuwani, Piyusha Umbrajkar “Forest fire detection using machine learning”, Volume 4, Special Issue 12, ICCEME 2019 - 2020, ISSN (Online) 2456 - 0774

[5] Faroudja Abid "A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems", doi. org/10.1007/s10694 - 020 - 01056 - z
[6] Jintao Qu, Xiaohui Cui "Automatic machine learning Framework for Forest fire forecasting", doi: 10.1088/1742 - 6596/1651/1/012116
[7] Aditi Kansal, Yashwant Singh, Nagesh Kumar, Vandana Mohindru "Detection of Forest Fires using Machine Learning Technique: A Perspective" 2015 Third International Conference on Image Infonnation Processing
[8] Guoli Zhang, Ming Wang, Kai Liu "Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China" doi. org/10.1007/s13753 - 019 - 00233 - 1
[9] Preeti T, Dr. Suvarna Kanakaraddi, Aishwarya Beelagi, Sumalata Malagi, Aishwarya Sudi "Forest Fire Prediction Using Machine Learning Techniques", DOI: 10.1109/CONIT51480.2021.9498448
[10] B S Negara, R Kurniawan, M Z A Nazri, S N H S Abdullah, R W Saputra, A Ismanto "Riau Forest Fire Prediction using Supervised Machine Learning", doi: 10.1088/1742 - 6596/1566/1/012002
[11] Ahmed M. Elshewey, Amira. A. Elsonbaty "Forest Fires Detection Using Machine Learning Techniques", ISSN No: 1006 - 7930
[12] Jonathan L. Hodges, and Brian Y. Lattimer, Jensen Hughes "Wildland Fire Spread Modeling Using Convolutional Neural Networks", doi. org/10.1007/s10694 - 019 - 00846 - 4