PREDICTING STEPPE FIRE USING SATELLITE DATA AND MACHINE LEARNING METHOD

Iderbayar Shiilegbat[1,3], Tsolmon Rentsen[1], Bayanjargal Darkhijav[2],
Munkh-Erdene Altangerel[3], Jargaldalai Enkhtuya[4]

[1]*Department of Physics, National University of Mongolia, Mongolia*
Email: iderbayar_sh@mas.ac.mn
[2]*Department of Applied Mathematics, National University of Mongolia, Mongolia*
[3]*Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences*
[4]*Institute of Geography and Geoecology, Mongolian Academy of Sciences*

KEYWORDS: Remote sensing (RS), Random Forest, Natural fire

ABSTRACT: Steppe fires caused by nature or humans are considered among the most dangerous and devastating disasters around the world. The purpose of this study is to develop a model for predicting steppe fire using random forest classification for natural fires. The study area is Dornod province located in the eastern part of Mongolia. Landcover is mainly a steppe area. There is a high number of fire events each year in springtime. We used satellite data such as the normalized difference vegetation index, land surface temperature, and modified soil-adjusted vegetation index for the spring of the years 2015-2022. Digital elevation model and climate data such as air temperature, precipitation, and wind speed were applied in this study. The overall accuracy of the random forest algorithm was 81%. The results showed that the random forest classification method can be used to predict steppe fires and monitor environmental issues in Mongolia.

1. INTRODUCTION

Steppe fires are hard to predict and extinguish and they cause enormous financial losses. A large number of fires are caused by human factors, although other factors like drought, wind, lightning strikes, and topography have an essential influence on fire occurrence and spread. Over the years, different authors have developed and used different methods for fire analysis. For example, Dutta et al. (2013) used neural networks on aggregated monthly meteorological data (evaporation, precipitation, incoming solar irradiance, maximum temperature, soil moisture, wind speed, pressure, and humidity) to estimate the risk of fire occurrence. The random forest algorithm has been parameterized and successfully trained to produce a satisfactory prediction of "fire and no-fire" class mapping on a daily basis (Jain et al. 2020). A recent study has shown that RS along with machine learning methods have been efficiently applied for steppe fire predictions (Jain et al. 2020).

Wildfire is one of the main factors influencing the spatial dynamics of grassland and forest ecosystems. Every year, thousands of hectares of grassland are damaged. Incidences of steppe and forest fires have been increasingly occurred in the past decades in Mongolia (Munkh-Erdene et al. 2021), and only in Dornod province a total of 29 wildfires occurred in 2020. To prevent from fires, Mongolian researchers have used different RS methods to monitor the fire disaster, determine the area affected by the fire and calculate the classes of the burnt area. For example, Byambakhuu et al. (2022) used a normalized burning ratio or "NBR" index, to determine the area of the field affected by the fire and the degree of the burned field. When spring warm temperature melts snow of the steppe of eastern Mongolia, the remains of last year's plant growth are exposed. In dry and windy conditions, these brown, dried grasslands can become tinder for wildfires.

At present, RS techniques along with thematic information stored in a geographic information system (GIS) are popular and effective means for identifying the most influencing factors affecting the fires. Over the past few years, RS platforms, techniques, and technologies have been significantly evolved and system capabilities have greatly improved, and the costs of many of these data sets have drastically decreased. Meanwhile, large archives of datasets are fully available in different Internet sites. Thus, it is possible to extract different thematic information at various scales, integrate the extracted information with other historical data sets stored in a GIS (Amarsaikhan et al. 2009). By the use of such extracted information, one can conduct sophisticated fire related analyses and use the result for decision-making processes.

The purpose of this study is to predict steppe fires using a random forest classification. At first, the fire and non-fire locations should be determined. Then, random forest classifiers should be trained with the RS dataset to predict steppe fires in the study area. Satellite-based Normalized Difference Vegetation Index (NDVI), Land Surface

Temperature (LST), and Modified Soil-Adjusted Vegetation Index (MSAVI) datasets were applied for the springs of the 2015-2022 years. Moreover, slope and aspect maps from a digital elevation model (DEM) have been derived. In addition, air temperature, precipitation, and wind speed have been used for the classification.

2. STUDY AREA

The study area is Dornod province, located in the eastern part of Mongolia (Figure 1). It covers the east-central Asian grassland steppe. The total area of Dornod province is 123.5 thousand square kilometers and geographically it is mainly the steppe, located at 560 – 1,300 m above sea level. The average annual rainfall is 150~300 mm, and it occurs during summertime. About 90% of the total area is steppe with hills. Bordered by Russia to the north and China to the east and southeast (Davaajargal et al. 2021). 81.3 percent of the total arable land or around 10.0 million hectares are either cultivated or used as pasture. Ten percent of the flora registered in Mongolia grows in Dornod, along with more than 40 kinds of herbs and 10 kinds of useful plants.

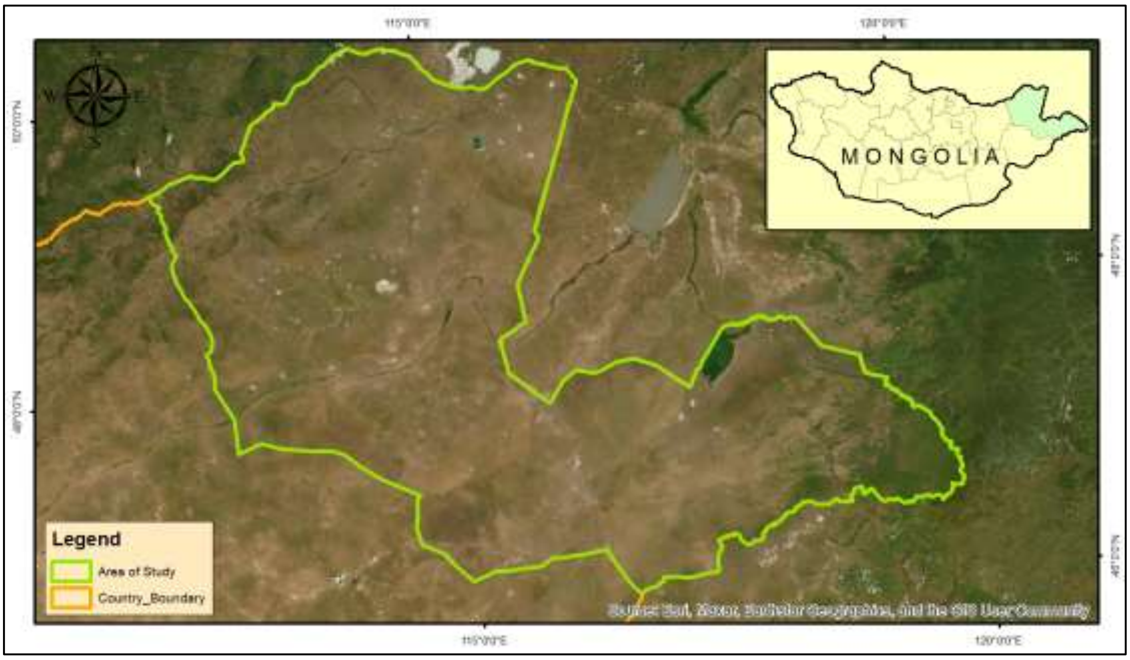

Figure 1. Study area: Dornod province.

3. DATA AND METHODOLOGY

In this study, MODIS satellite products such as the Normalized Difference Vegetation Index, Land Surface Temperature, Modified Soil-Adjusted Vegetation Index for the spring of the 2015-2022 years as well as climate data (spring precipitation, wind speed, and air temperature for 2015-2022) have been used. Table 1 shows satellite data characteristics.

3.1 Normalized Difference Vegetation Index

The NDVI is a commonly used RS technique that identifies vegetation and measures a plant's overall growth. The range of NDVI is between minus1 to plus 1. Values close to +1 show healthy and dense vegetation (Przyborski, 2000).

3.2 Land Surface Temperature

LST has been widely used for environmental studies, using satellite-derived images. Satellites only measure land surface temperature (Table 1).

3.3 Modified Soil-Adjusted Vegetation Index

MSAVI is the soil-adjusted vegetation index that seeks to address some of the limitations of NDVI when applied to areas with a high degree of exposed soil surface (Qi et al. 1994). The MSAVI was calculated by using the following formula (1).

$$\mathrm{MSAVI} = \frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - Red)}}{2} \quad (1)$$

3.4 DEM

In this study, SRTM DEM of the study area has been used. The data of raster parameters such as the slope and aspect map are derived from the DEM dataset at dimensions of 30x30m. Table 1 shows satellite data sets with spatial resolution.

Table 1. Satellite data (Earthdata, 2021).

Terra/MODIS Product code	Product name	Spatial resolution (m)	Temporal resolution (days)
MOD13Q1.006	RED	250	16
MOD13Q1.006	NIR	250	16
MOD11A2.006	LST	1000	8
MOD13Q1.006	NDVI	250	16
SRTM	DEM	30	-

3.5 Climate data

Climate data, such as spring precipitation, wind speed, and air temperature for 2015-2022 were used from the Mongolian statistical information (Mongolian Statistical Information Service, 2022).

3.6 Random forest algorithm

The random forest method developed by Breiman and Cutler, is an ensemble learning technique based on classification and regression trees. It can evaluate the relationship between covariates and dependent variables, and calculate the relative importance of covariates (Cutler et al. 2007; Kubosova et al. 2010). It has been applied in various fields in the past, including medicine, genetics, ecology, and RS. In recent years, it has been used in forest fire forecasting, and it has a good predictive capacity (Oliveira et al. 2012; Rodrigues et al. 2014; Kane et al. 2015). The method consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest is split into a class prediction and the class with the most votes becomes the model’s prediction. A decision tree, in a mathematical sense, is an acyclic graph with a fixed root. Each tree is trained using a bootstrap sample of the training data, and at each node, the best split is selected from random subsets of the predictor variables. It ensures that each tree utilizes the training data and predictor variables in a different way, reducing its statistical dependence on the other trees (Breiman, 2001).

4. RESULTS AND DISCUSSION

In this paper, we have used the MSAVI, NDVI, DEM, and LST indices over Dornod province in Mongolia, using satellite data for the period 2015-2022. Satellite image products have been downloaded from the database of the Information and Research Institute of Meteorology, Hydrology, and Environment of Mongolia. In order to apply the proposed methodology and to build our dataset using RS data, we selected some fire zones "no-fire" and "fire" that occurred in the Dornod province steppe between 2015 and 2022. Figure 2 shows data on the locations of 152 fires and 137 non-fires collected using satellite images as well as the slope of the study area. As required in the random forest model, the data were converted into binary format. The character variables like, “fire and no fire” were converted into dummy variables. For instance, if it provides input as “fire” for the variable, the binary value 1 is assigned, and "no-fire" responses are assigned with 0.

Since there is no high and multi−correlation among the data, we considered all variables for further analysis (Figure 3). We split the data set into a training dataset (70% of data) and a validation dataset (the remaining 30 %) for the RF algorithm. The datasets were created with nine dependent variables to run the model. The model is built on training datasets and tested on validation datasets. Figure 4 shows the feature importance score and rank that contribute to fires. LST and Aspect have the greatest impact on classification (Feature important scores: 24%, 19%).

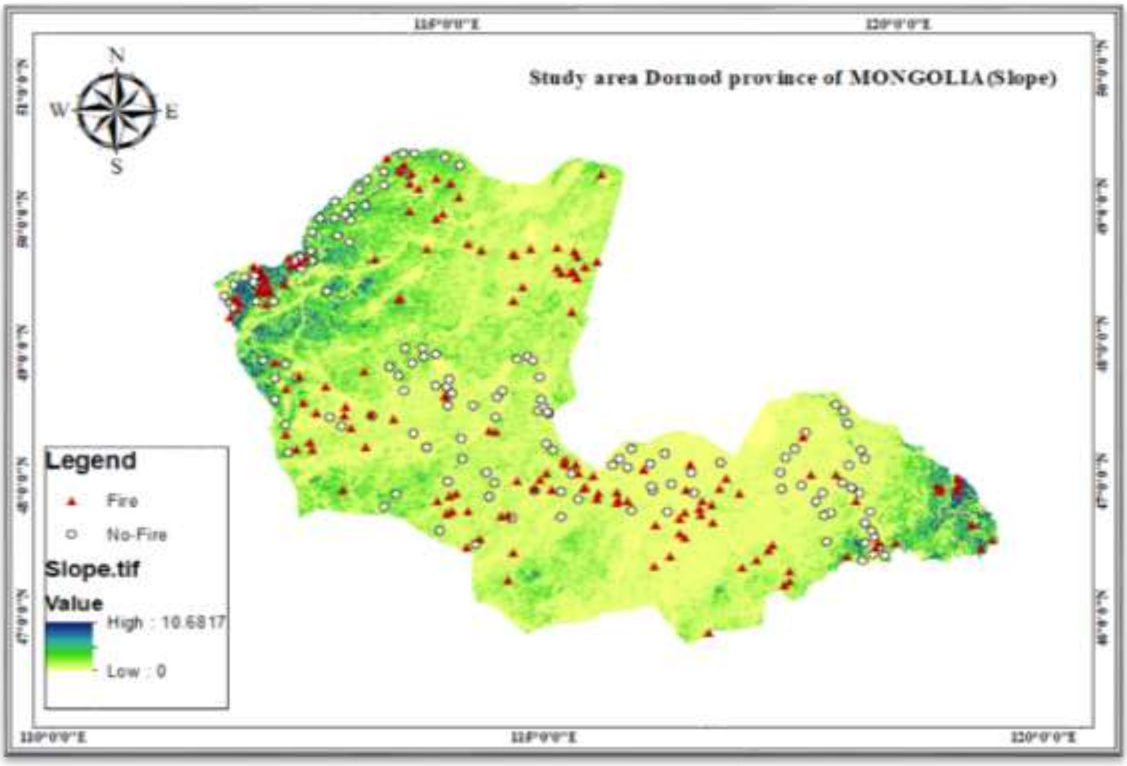

Figure 2. Location of fires in the study area.

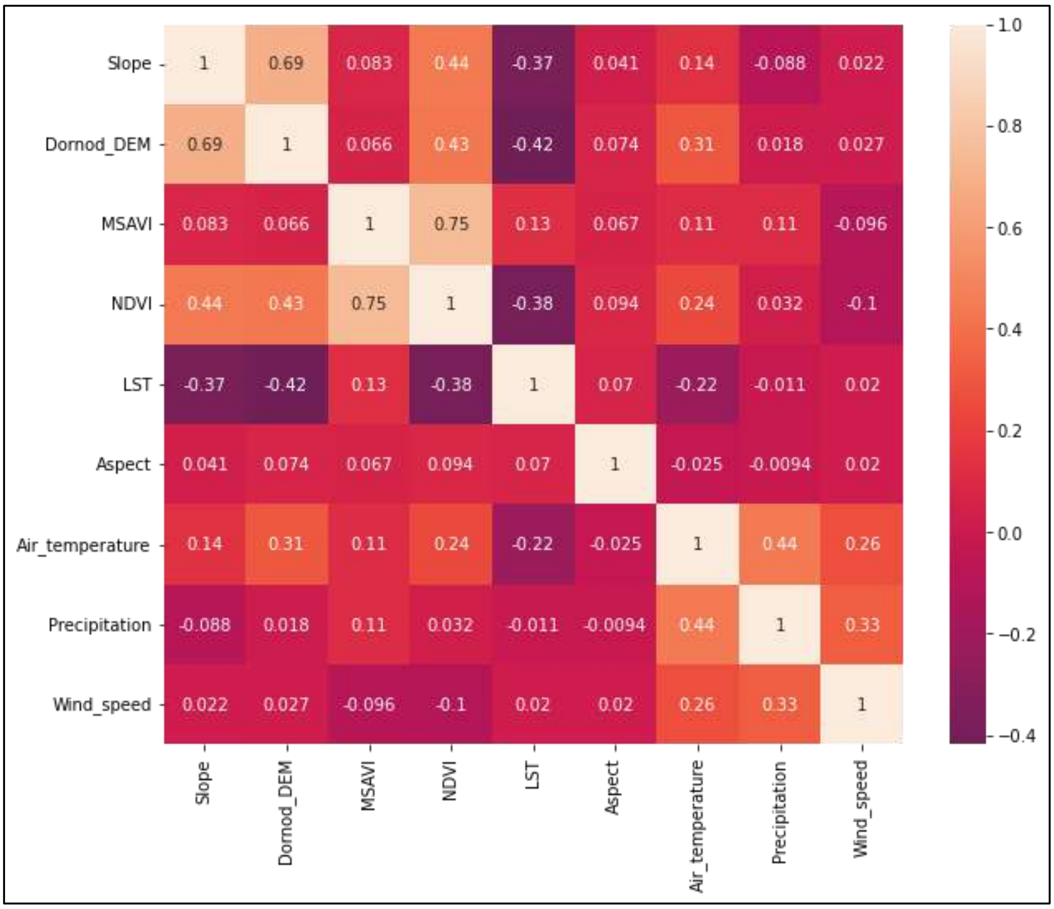

Figure 3. A correlation matrix.

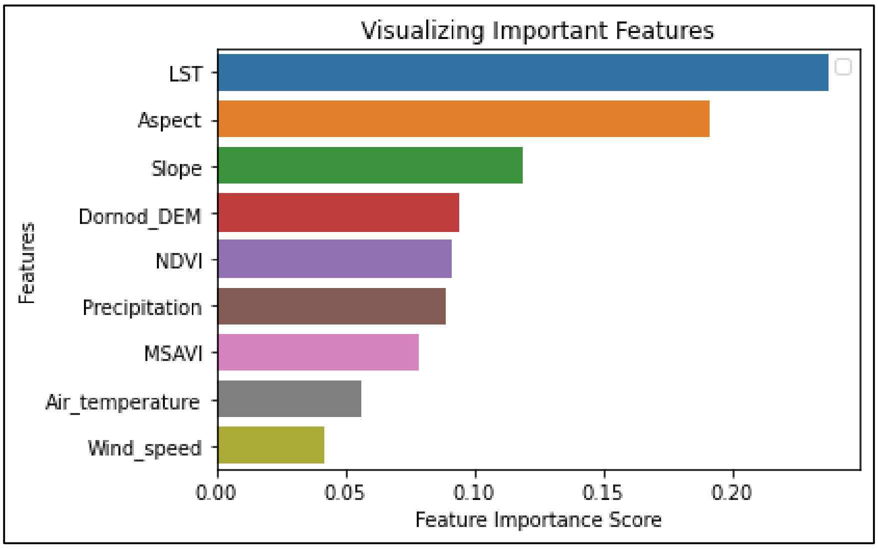

Figure 4. Factor importance score.

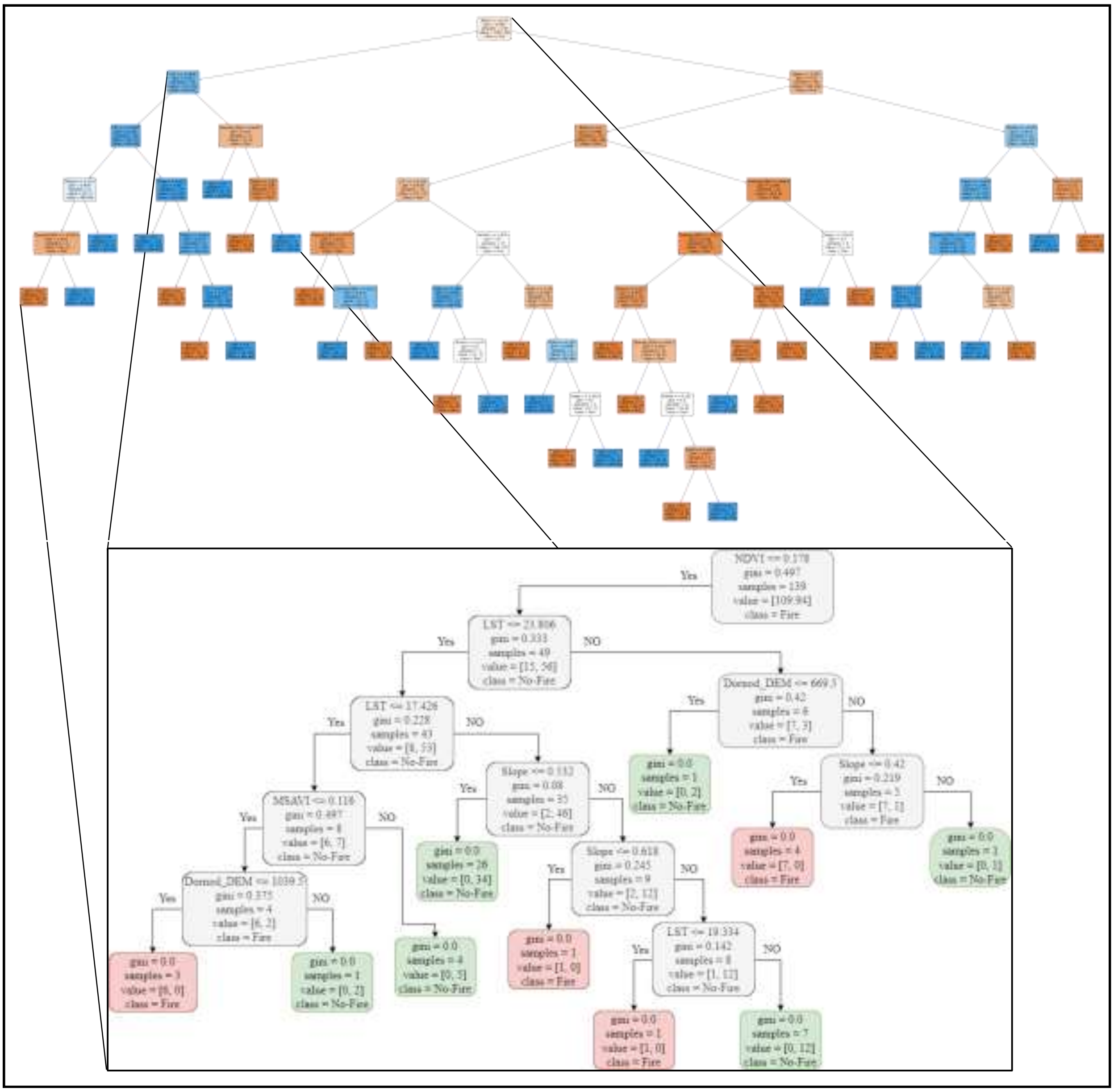

Figure 5. Random forest classification tree and one branch parts were magnified.

We applied the random forest classification for fire zone and non-fire zone. Random forest classifiers were performed by applying the values of satellite data and climate data at these points and the results are shown in Figure 5. The number of decision trees of the model parameter is 15 which is most suitable in our case. Figure 5 shows first a decision tree selected from the random forest. Depending on the features, it shall predict fire risk and occurrence. To construct a decision tree, we used the Gini impurity method for splitting the tree. Gini impurity has a maximum value of 0.5, which is the worst we can get, and a minimum value of 0 means the best we can get. As a result, we get the leaf node having GINI = 0; which means it's the purest node that can't be further divided.

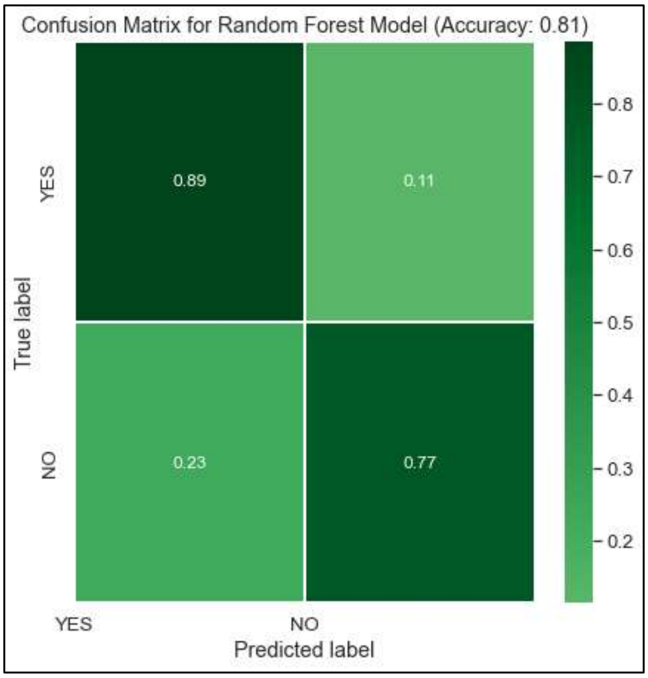

Figure 6. Confusion matrix.

We made predictions using the remained 30% which includes 87 points. As seen from Figure 6, points with fire occurrence were predicted with 89% accuracy as fire points, and 11% of them were incorrectly predicted. Moreover, it is seen that 23% of no–fire points were predicted as fire points, and 77% of them were predicted correctly.

5. CONCLUSIONS

This study used satellite and climate data and applied the machine learning method for fire prediction analysis. Our proposed model used various factors with different spatial and temporal resolutions. In the study, a total of 289 locations were identified, and 15 decision trees were built for the RF algorithm. With the help of satellite data, it was possible to determine the locations of fire and non-fire areas. The RF classification performed well, forecasting fire occurrence in the study area with a high prediction accuracy of 81% by using all factors. The accuracy of the prediction of fire points was 89%. As seen, the outcomes of the study could be used for different fire prevention activities and for supporting fire-related decision-making processes. Further, the number of fire and no-fire points should be increased to improve the outcome of the machine learning method.

REFERENCES

Amarsaikhan, D., Blotevogel, H.H, Ganzorig, M., and Moon, T.H., 2009. Applications of remote sensing and geographic information systems for urban land-cover change studies in Mongolia, Geocarto International, 24:4, 257-271, https://doi.org/10.1080/10106040802556173.

Breiman L., 2001. Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Byambakhuu, G., Battsengel, V., Narantsetseg, Ch., et al, 2022. A wildfire monitoring study for burn severity and recovery process using remote sensing techniques: A case study near Shiliin Bogd mountain, Eastern Mongolia. Journal of Geographical Issues, Volume 22 (1), ISSN: 2312-8534, 20-31, https://doi.org/10.22353/.v22i1.

Cutler DR, Edwards TJ, Beard KH, Cutler A, Hess KT, Gibson J, and Lawler JJ., 2007. Random forests for classification in ecology. Ecology, 88 (11), 2783–2792.

Davaajargal, J., Bayanjargal, D., and Tsolmon, R., 2021. Estimation of crop suitability using NDVI in The Kherlen Basin Dornod province Mongolia. International Journal of Science, Environment, and Technology 10, 19-21.

Dutta, R., Aryal, J., Das, A., and Kirkpatrick, J.B., 2013. "Deep cognitive imaging systems enable estimation of continental-scale fire incidence from climate data", Scientific reports, vol. 3, no. 1, pp. 1–4.

Earthdata, 2021. AppEEARS, from https://lpdaacsvc.cr.usgs.gov/appeears/.

Jain P., Coogan S.C.P., Subramanian S.G., Crowley M., Taylor S., Flannigan M.D., 2020. A review of machine learning applications in wildfire science and management (2020).

Kane, VR., Lutz, JA., Alina Cansler, C., Povak, NA., Churchill, DJ., Smith, DF., Kane, JT., and North, MP., 2015. Water balance and topography predict fire and forest structure patterns. Forest Ecol Manag, 338:1–13.

Kubosova K, Brabec K, Jarkovsky J, Syrovatka V., 2010. Selection of indicative taxa for river habitats: a case study on benthic macroinvertebrates using indicator species analysis and the random forest methods. Hydrobiologia, 651:101–114.

Mongolian Statistical Information Service, 2022. from https://www.1212.mn/

Munkh-Erdene, A., Amarsaikhan, D., Jargaldalai, E., Nyamjargal, E., 2021. Forest Fire Risk Assessment Model Using Remote Sensing and GIS Techniques in Tujiin Nars National Park, Mongolia, Proceedings of the Mongolian Academy of Sciences, Ulaanbaatar, Mongolia. https://doi.org/10.5564/pmas.v61i01.1557

Oliveira S, Oehler F, San-Miguel-Ayanz J, Camia A, Pereira J., 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. Forest Ecol Manage, 275:117–129.

Paul Przyborski., Normalized Difference Vegetation Index (NDVI), 2000. Earth observatory, from https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php.

Piyush Jain, Sean C.P. Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D. Flannigan, 2020. A review of machine learning applications in wildfire science and management, Environmental Reviews, Volume 28 (4), https://doi.org/10.1139/er-2020-0019.

Qi J., Chehbouni A., Huete A.R., Kerr Y.H., Sorooshian S., 1994. A modified soil adjusted vegetation index. Remote Sensing of Environment, Volume 48, Issue 2, Pages 119-126.

Rodrigues M., Riva JDL., 2014. An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ Model Softw, 57:192–201.