PREDICTING FOREST FIRE USING MODIS SATELLITE DATA AND MACHINE LEARNING METHOD

Tsolmonbayar Shagdar[1], Tungalag Amar[1], Batchuluun Tseveen[1]
Oyunsanaa Byambasuren[1], Iderbayar Shiilegbat[2], Bayanmunkh Norovsuren*[3]

[1]School of Engineering and Applied Science, National University of Mongolia, Ulaanbaatar, Mongolia
Email: tsoko.tsomi@gmail.com, tungalag0504@gmail.com, batchuluun@num.edu.mn, oyunsanaa@seas.num.edu.mn

[2]Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences,
Ulaanbaatar 13330, Mongolia
Email: iderbayar_sh@mas.ac.mn

[3]Senior Analyst, Laboratory for Geo mineralization, Mongolian National University
Email : n.bayanmunkh@mnun.edu.mn

KEY WORDS: Machine learning, random forest classification, MODIS

ABSTRACT: In Mongolia, the physical features are the climate that ranges from semi-arid in the North to desert in the South; it is continental with seasonal temperature ranges, with large temperature differences between years. In a country with an extremely dry climate and low precipitation, the input of forest resources on the ecosystem functioning is relatively higher than in humid areas. The impacts of climate change and human wrong activities are critically affected by desertification, biodiversity losses, poverty of nomadic herders' livelihood, losses of livestock numbers, a decrease in water sources, and increased forest and steppe fires. There is a limited research methodology for forest fire risk in Mongolia. This paper focuses on predicting a machine-learning method and analysing forest fires. We are analysing for random forest classification. The selected area is located in Dornod province in the northern part of Mongolia bordering the Siberian Forest area. We used products of MODIS (250m) satellite data and image calculation to indexes such as MSAVI, LST, Normalised Burn Ratio (NBR), and NDVI for the years 2015-2022. We used statistical data detailed by climate (precipitation and temperature), and DEM (slope and aspects) were used for modelling. Outputs of the model for the forest fire map were compared to ground truth data and with other government database resources it was 75% agreement. Random forest methods can be used for forest fire risk in different areas or steppe fires.

1. INTRODUCTION

In Mongolia, due to the effects of environmental impacts such as low precipitation, windstorms, lightning, and extreme heat, as well as human-caused forest fires, the number of forest fires increases and reaches catastrophic levels in spring and autumn. Globally, more than 50 million forest and field fires occur annually. The study area borders Russia's Siberian forest-steppe region, where wildfires in January 2019 burned 6.7 million acres and released 49 megatons of carbon dioxide into the atmosphere, equivalent to the emissions from 36 million cars. As of 2022, more than 700 forest and field fires have occurred in Mongolia in the last 5 years, burning 44.2 million hectares and causing 160 billion MNT in damage.
In 43 countries of the world, in the last 5 years, about 3 million forest and field fires broke out, and more than 23 thousand people lost their lives. Long-term data can be used from satellite data and is widely used in environmental research. In recent years, artificial intelligence has been developing rapidly, and most researchers are using machine learning and deep learning methods in their research. Environmental factors were selected based on the results of other researchers to predict wildfires. This research aims to predict forest fires by combining MODIS satellite data, meteorological statistics, and machine learning Random Forest (RF) classification method for the spring period of March-May 2015-2022. In the study "The impact of forest fire on forest cover types in Mongolia" by (Nandin-Erdene et al., 2021) they mapped the burned area, and as a result, 25,239 hectares of forest were burned, and 52,603 hectares of forest were transformed into steppe areas. (Elbegjargal N., et al.,) "A spatial distribution map of the wildfire risk in Mongolia using a decision support system", studied the distribution of wildfires throughout Mongolia in 2018.
The study identified and mapped areas at risk of fire with the help of 17 factors such as socio-economic, environmental, climate and wildfire risk. Also, the risk map is classified into five categories, and based on spatial statistics, fire risk is divided into high risk, high risk, medium risk, low, and very low risk (Elbegjargal N., et al.,). The results show that the percentage of the study area is at risk of fire in nature reserves in each category. Wildfires are one of the main factors affecting the spatial dynamics of grassland and forest ecosystems. Every year, thousands of hectares of grasslands are burned and destroyed. In Mongolia, the number of fires in fields and forests has been increasing in recent decades (Munkh-Erdene A., et al. 2021), and a total of 29 fires occurred in 2020 in Dornod Province alone. Mongolian researchers are still using different methods and methods to prevent fires. Remote sensing methods can monitor fire disasters, identify areas

affected by fires, and classify burned areas. For example, Byambakhuu et al. used the normalized burn ratio or "NBR" index to determine the burned area and extent of burned area in 2022. When the warm weather of spring comes and the snow melts in the eastern steppes of Mongolia, the remnants of last year's plant growth are exposed (Munkh-Erdene, A., et.al., 2021)., In dry and windy conditions, these brown, dried-up grasslands can become hotbeds for wildfires. Currently, remote sensing, together with data stored in geographic information systems (GIS), is a popular and effective tool for determining the most influential factors in wildfires (Byambakhuu G., et.al. 2022). Over the past few years, remote sensing platforms, techniques, and technologies have evolved significantly, system capabilities have dramatically improved, and the cost of many of these data sets has decreased dramatically. At the same time, large archives of datasets have been made available on different sites on the Internet.

Therefore, it is possible to extract different data at different scales and integrate the extracted data with other historical data sets stored in GIS (Amarsaikhan et al. 2009). By using such mined data, detailed fire-related analysis can be performed, and the results can be used in the decision-making process. For decades, forest fires have been a major environmental problem in Southeast Asia, with significant impacts on the atmosphere, carbon cycle, and various ecosystems (Cortez p., et.al. 2007). Forest fires can cause economic hardship, business disruption, and poor health (Davies D., et.al.2009). Parameters affecting wildfires include NDVI, elevation, slope, direction, soil moisture, fuel, and climate data, which show that they are factors that significantly affect fire risk and fire occurrence (Jaafar Z., et.al. 2014) and (Chisholm R., et.al. 2016) and (Davies D., et.al. 2009). In 2013, data networks were used on monthly meteorological data (evaporation, precipitation, incident solar radiation, maximum temperature, soil moisture, wind speed, and humidity) to estimate fire risk (Huijnen V., et.al.2015). Jain et al., “A review of machine learning applications in wildfire science and management”, 2020 parameterized the Random Forest algorithm and successfully trained their model to make effective predictions about daily “fire and non-fire” class maps (Stockwell C., et.al. 2016) and (Dutta R et.al. 2013). Recent studies have shown that remote sensing combined with machine learning methods can be used effectively to predict wildfires (Piyush Jain et al 2020). Nowadays, interdisciplinary research and analysis is gaining momentum, not just research in that field. For this reason, it is desirable to follow the standards of other countries and improve the methods and methods of machine learning and artificial intelligence in research, and in Mongolia, the results of research using this method to predict forest and field fires have not been done so far. Therefore, we aimed to predict wildfires using machine learning and machine learning methods.

In our country, land degradation and dryness are expected to increase from year to year, as well as the occurrence of forest and field fires have also increased compared to previous years (Magsar E.,et.al. 2012). Forest fires should be predicted in order to prevent fires from occurring rather than reacting to them after they occur. Although there are methods to calculate fire risk classification, frequency, burned area, etc., but they are not very effective in predicting, one example may be related to the increasing number of fires from year to year (Roger D., et.al. 1990). However, by using a combination of artificial intelligence and remote sensing methods, it is possible to train models on big data and prevent risks. Also, by using satellite data, we can save time, budget, equipment, etc. by measuring the land. Dividing into 2 categories: with fire and without fire, and aiming to predict using machine learning methods, is innovative compared to the previous research work of Mongolian researchers.

1.1 Study area

The study area located at the Dornod province is situated in the northern part of Mongolia borders with the Russian Federation. Dornod province is located in the taiga zone, forest steppe zone and steppe zone. The north of the province is characterised by alpine forests, gradually blending in the arid steppe plains of the central Mongolian highland (Figure 1). 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 kilometres and geographically it is mainly the steppe, located 560 – 1,300 m above sea level. The average annual rainfall is 150~300 mm, 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 J., et al 2020). 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.

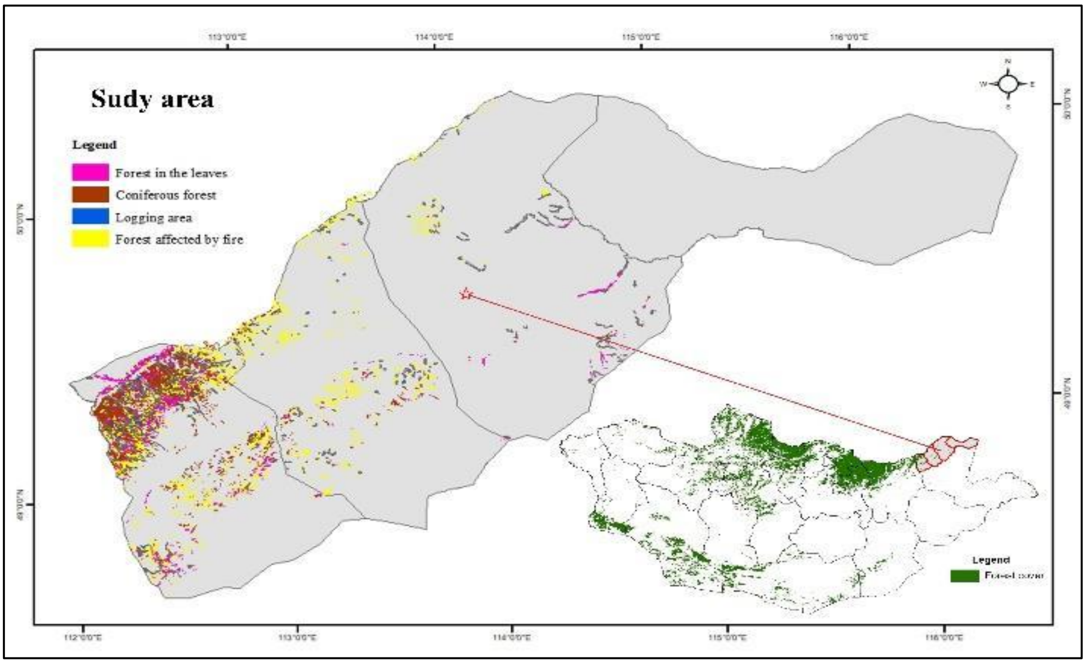

Figure 1. Location map (b) The distribution of natural zones in Mongolia and (b) The selected area located in Dashbalbar, Chuluunkhoroot, Bayandun, Bachi-Uul in Dornod province.

Most of the moisture coming from the air currents from the west and north-west is stopped in the Khentii mountain range, so the Eastern Plain has a humid-cool, dry-cooler, dry-warm climate. The region was chosen as the research region because most of the forest and field fires occurring in Mongolia are in the Dornod province.

2. MATERIALS AND DATASETS

2.1 Satellite Remote Sensing Data

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

Table 1. Satellite data bands, and time period covered.

	Thermal band (s)	Spatial resolution (m)	Time Period
Terra/MODIS	RED	250	2015-2022
	NIR	250	
	LST	1000	
	NDVI	250	
SRTM	DEM	30	

The data of raster parameters such as the slope and aspect map derived from the DEM dataset at dimensions of 30x30m. figure 2 shows satellite data aspect, slope, dem.

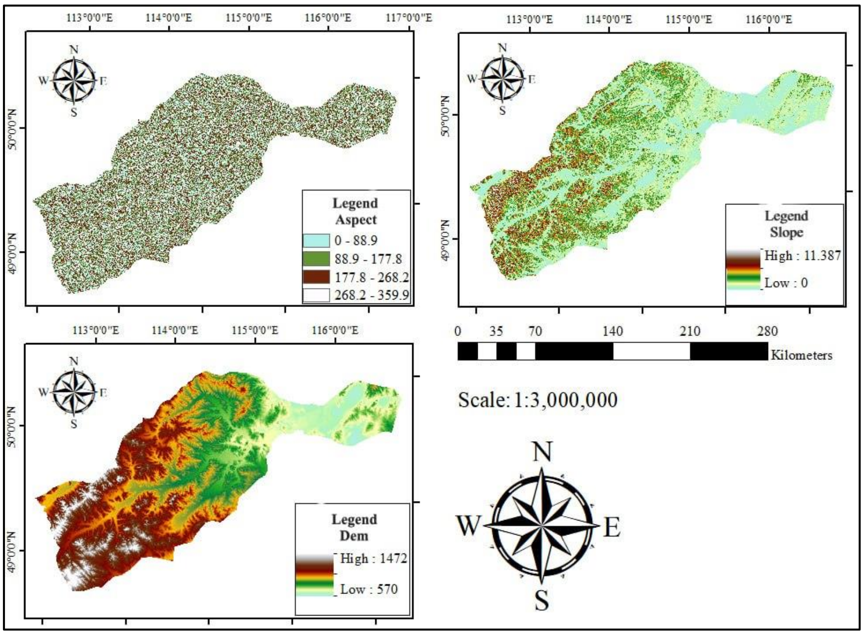

Figure 2. Map of the Aspect, Slope, Dem

2.2 Forest Inventory Data

The systematic national forest inventory (NFI) has been conducted nearly every ten years since late 1956 and the last national forest inventory was conducted for the period 2015. The statistics of the national forest inventory in Mongolia are based on large numbers of field plots and are the most important data sources in research on forest (Norovsuren B, 2019).

2.3 Climate data

Climate data, such as spring precipitation, wind speed, and air temperature for 2015-2022, were used from the Mongolian statistical information. Meteorological station and MODIS data were collected and used by the Institute of Water, Meteorology, Environmental Research and Information for the spring season of 2015-2022, average precipitation, wind speed, and air temperature for each month of 3, 4, and 5 months.

3. METHOD

We used Remote sensing methodology for the large-resolution satellite data. Assessment processed with the layer of the data such as forest taxation data of the NFI, climate factors, burned area and GIS software's. We used NBR index, Machine learning method and we defined forest cover area of study area. Following schema shows methodology step by step (Figure 3).

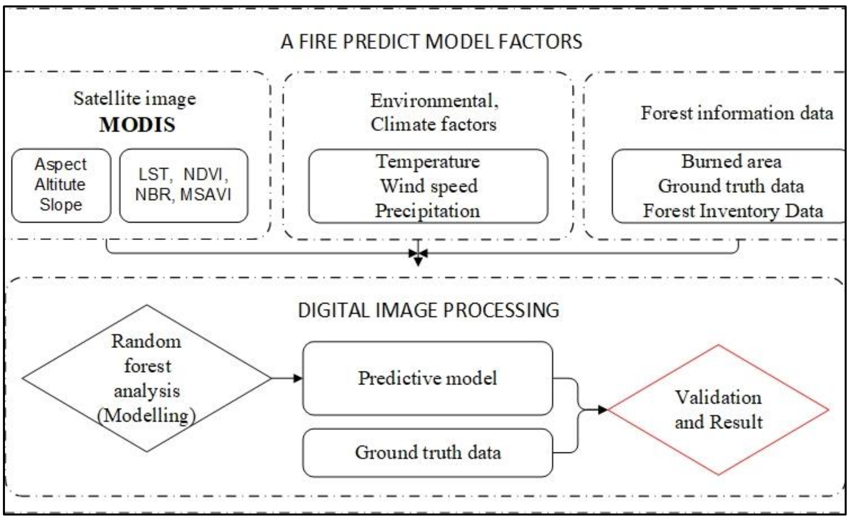

Figure 3. Methodology schema

3.1 Random Forest algorithm

Random Forest (RF) developed by Breiman and Cutler, is an ensemble learning method 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). RF has been applied in various fields in the past, including medicine, genetics, ecology, and remote sensing. In recent years, it has been used in forest fire forecasting, it has a good predictive capacity (Oliveira et al. 2012; Rodrigues and Riva 2014; Kane et al. 2015). A random forest consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest is split in a class prediction and the class with the most votes becomes the model's prediction. It ensures that each tree utilises the training data and predictor variables in a different way, reducing its statistical dependence on the other trees (Breiman L., 2001).

3.2 Estimation Algorithm

Normalised Difference Vegetation Index
The NDVI is a commonly used remote sensing technique that identifies vegetation and measures plant’s overall growth. The range of NDVI is between minus1 to plus 1. Values close to +1 show healthy and dense vegetation (Paul Przyborski, 2000).

$$NDVI = \frac{NIR-Red}{NIR+Red} \qquad (1)$$

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, J. et al 1994). The MSAVI was calculated by using the formula (2).

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

Land Surface Temperature
LST has been widely used for environmental studies, using satellite-derived images. Satellites only measure land surface temperature (Table 1). The LST was calculated using Equation (3) by (Norovsuren et al., 2019).

$$LST = \left(BT + w * \frac{BT}{p}\right) * ln\ (e) \qquad (3)$$

Normalised Burn Ratio
To benefit from the magnitude of spectral difference, NBR uses the ratio between NIR and SWIR bands, according to the formula shown below. A high NBR value indicates healthy vegetation while a low value indicates bare ground and recently burnt areas.

$$NBR = \frac{NIR-SWIR}{NIR+SWIR} \qquad (4)$$

Data from 2015 and 2019 were used to train the Random Forest classification model. However, the data of 2020 was used for model testing. Out of a total of 644 fire and non-fire point data, 493(76.5%) were used as training data and 151(23.5%) were used for testing. The breakdown of total data by year is shown in the following table (Table 2).

Table 2. Selected years.

Data	Years	Fire	No-fire
Train	2015	217	218
	2019	27	31
Test	2020	73	78
Total		317	327

4. ANALYSIS AND RESULTS

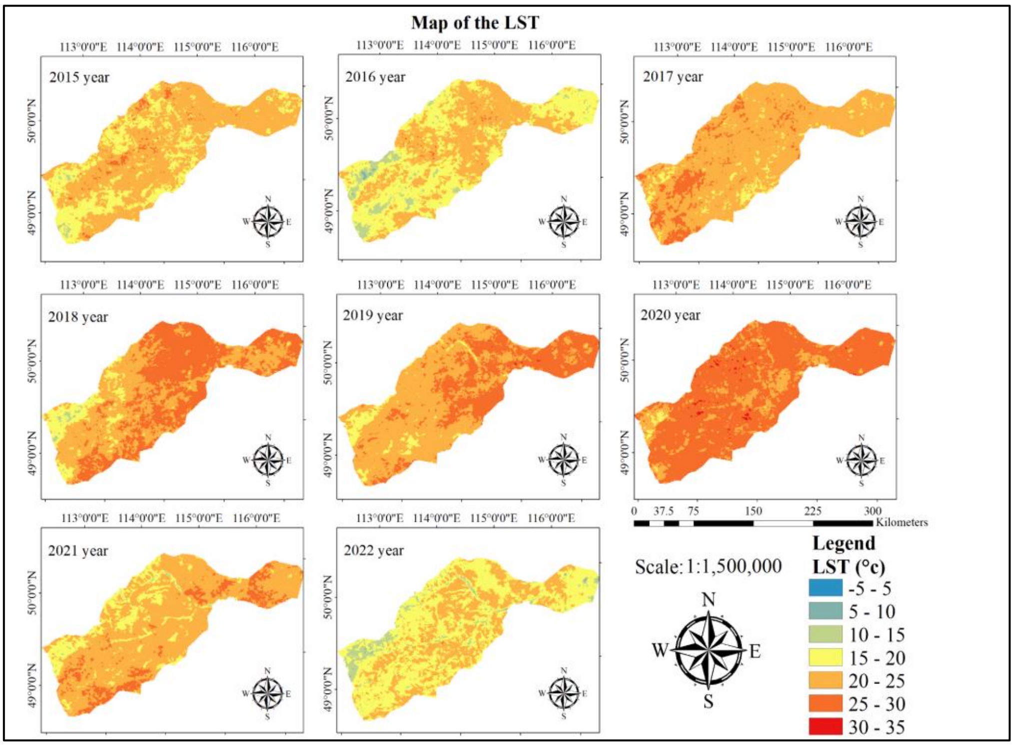

Figure 4. Example mapping: Land Surface Temperature in 2015-2022 years

In This research, we have used the remote sensing indexes such as MSAVI, NDVI, DEM and LST (Figure 4) over the northern part of Dornod province in Mongolia using MODIS data from the period 2015-2022. In order to apply the proposed methodology and. In order to apply the proposed methodology to build our dataset using Remote Sensing data, we selected some fire zones "no-fire" and "fire" that occurred in the selected area in the Forest region between 2015 and 2022. The following figure 5 shows the correlation between the factors. It is worth noting that NDVI and MSAVI have a positive correlation, because these 2 indices are calculated using information from the visible light red ray (Red) channel of the spectral region, so there is a positive or 100% correlation.

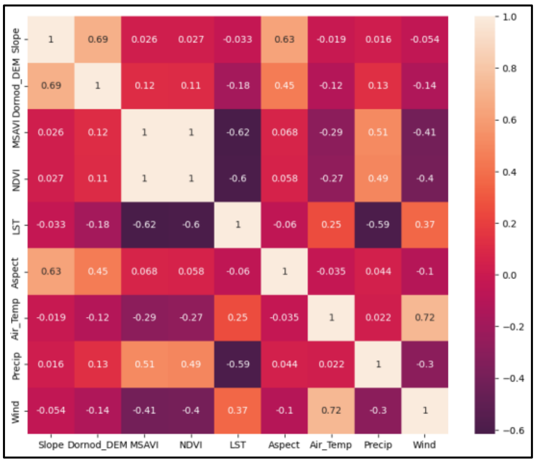

Figure 5. Correlation matrix between selected 9 factors

Therefore, NDVI is used to calculate the land cover and MSAVI is excluded because it can be seen from the multi-year change map. The following figure shows the relationship of the factors after removing MSAVI (figure 6).

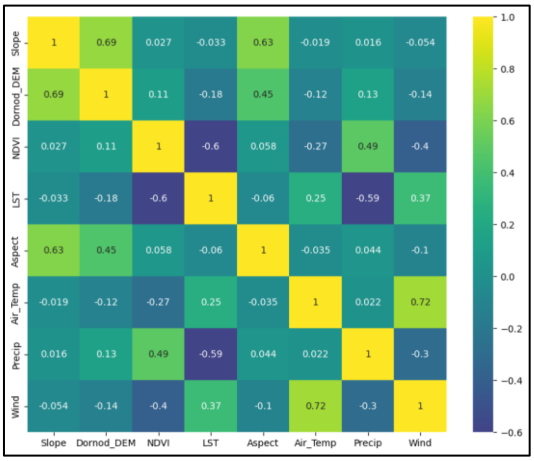

Figure 6. Correlation matrix between 8 factors excluded MSAVI

The result of the model, or one decision tree of the Random Forest model, is shown in the following figure 7 as an example. Standardisation of a Factor's dataset is a common requirement for many machine learning estimators: they might behave badly if the individual features do not more or less look like standard normally distributed data. Standardise features by removing the mean and scaling to unit variance.

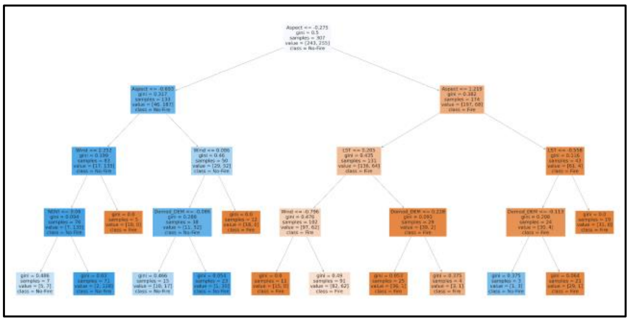

Figure 7. Results of the Random Forest Model

The following figure shows the percentage of effects that the factors contribute to the model. Here, aspect has 48% or the most influence. On the other side, air temperature, precipitation and slope have less than 6% influence. This climate data is believed to be due to the use of statistics. Climate data should be used as CRU data to improve model results (figure 8).

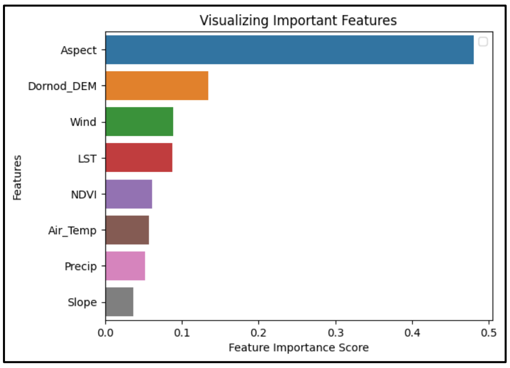

Figure 8. Factor importance score

We applied the Random Forest classification for fire zone and non-fire zone. Number of decision trees of the model parameter is 15 which is most suitable in our case. 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 which has GINI = 0; which means it's the purest node that can't be further divided. The results were shown in figure 5, this is a decision tree selected randomly from forest. Therefore, branch parts were magnified in the figure. Depending on features, it shall predict fire risk and occurrence. For this, we need data of such predicting days. The following figure shows the confusion matrix between the predicted values of the model and the actual data 2. In this, 75% predicted the spot where there was a fire, and 94% predicted the place without fire (figure 9).

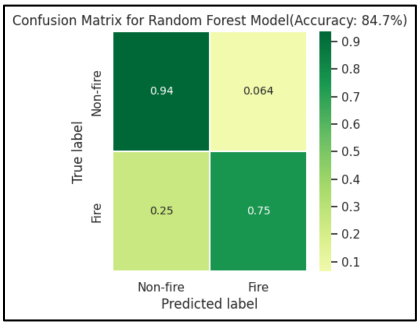

Figure 9. Confusion matrix between area of non-fire and fire

The numerical value of the confusion matrix is shown in the following figure. The random forest classification model correctly predicted 55 (75%) of the 73 fire points (Figure 10).

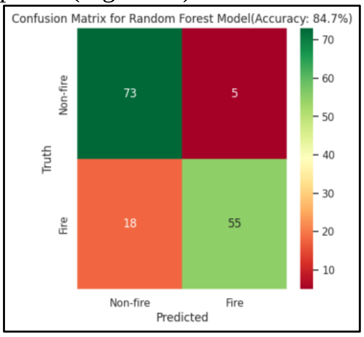

Figure 10. Confusion matrix between

Figure 11 shows the validation between research results with NFI data. Accuracy is 91.5 percent. Also, the following figure shows the number of points with and without forest and field fires in each year. In the experimental data, 86% of the forest fire areas contain forest fire areas, and 97% of the non-forest fire areas contain forest fire-free points.

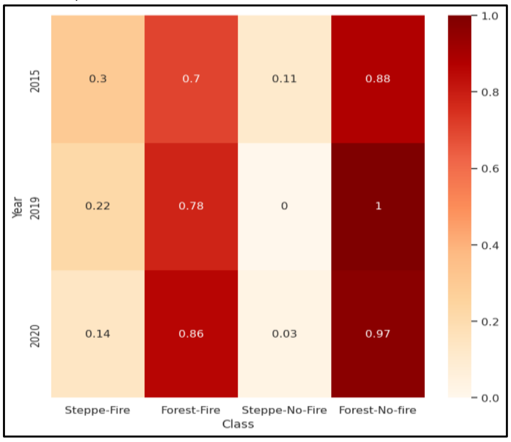

Figure 11. Validation matrix.

5. CONCLUSIONS

This study used satellite and climate data and applied machine learning methods for fire cases. Our proposed model used the various factors with different spatial and temporal resolutions. With the help of satellite data, it was possible to determine the locations of fire and non-fire hotspots. We have a total of 644 locations identified. This including 76.5 % is training data and 23.5 % test data. The number of the decision trees is 5 for the Random Forest algorithm. The Random Forest Classification is performed well, forecasting fire occurrence in selected areas with a high prediction accuracy of 84.7% by using all factors. The accuracy of the prediction of fire points is 75%. 86 percent of the fire point for test data is forest area and 14 percent is steppe fire. Random forest methods can be used for forest fire risk in different areas or forest fires. Further methodological recommendations will be used related to other experts and experts and would be connected to the international researcher's transit zone of the Siberian part.

6. ACKNOWLEDGEMENTS (OPTIONAL)

This research was partially supported by “NUM-ITC-UNESCO" Space Science and remote sensing international laboratory, Physics Department, National University of Mongolia. Authors are thankful for MODIS data from the National Remote Sensing Centre and climate data supporting from Information and research Institute of Meteorology, Hydrology and Environment, Ulaanbaatar, Mongolia.

7. REFERENCES

Nandin-Erdene Geserbaatar, Elbegjargal Nasanbat, Ochirkhuyag Lkhamjav, 2020. “The impact of forest fire on forest cover types in mongolia”, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vo. XLIII-B3-2020,

Elbegjargal Nasanbat, Ochirkhuyag Lkhamjav, Amonjol Balkhai,Chuluunbaatar Tsevee-Oirov, Amarzaya Purev and Munkhzul Dorjsuren, 2018. “A spatial distribution map of the wildfire risk in mongolia using decision support system”, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Vo. XLII-3/W4,

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.

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.

Cortez P. and Morais A., 2007. A Data Mining Approach to Predict Forest Fires using Meteorological Data. *In New Trends in Artificial Intelligence, 13th EPIA 2007* - Portuguese Conference on Artificial Intelligence, 512–523.

Davies, D. K.; Ilavajhala, S.; Wong, M. M.; and Justice, C. O., 2009. Fire Information for Resource Management System: Archiving and Distributing MODIS Active Fire Data. *IEEE Transactions on Geoscience and Remote Sensing*, 72–79.

Jaafar Z. and Loh T. L., 2014. Linking land, air and sea: potential impacts of biomass burning and the resultant haze on marine ecosystems of Southeast Asia. Global Change Biology, 2701–2707.

Chisholm R. A., Wijedasa L. S., and Swinfield T., 2016. The need for long-term remedies for Indonesia’s forest fires. Conservation Biology, 5–6.

Huijnen V., Wooster M., Kaiser J., Gaveau D., Flemming J., Parrington M., Inness A., Murdiyarso D., Main B., and van Weele M., 2016. Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997. *Scientific Reports* 26886.

Stockwell C. E., Jayarathne T., Cochrane M. A., Ryan K. C., Putra E. I., Saharjo B. H., Nurhayati A. D., Albar I., Blake D. R., Simpson I. J., Stone E. A., and Yokelson R. J., 2016. Field measurements of trace gases and aerosols emitted by peat fires in Central Kalimantan, Indonesia, during the 2015 *El Nino. Atmospheric Chemistry and Physics*, 11711–11732.

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

Kontoes C., Keramitsoglou I., Papoutsis I., Sifakis N., Xofis P., 2013. National scale operational mapping of burnt areas as a tool for the better understanding of contemporary wildfire patterns and regimes. *Sensors* 13(8), 11146–11166,

Pourtaghi Z.S., Pourghasemi H.R., Aretano R., and Semeraro T., 2016. *Investigation of general indicators influencing forest fire and its susceptibility modelling using different data mining techniques.* Ecol. Indic, 64, 72–84,

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.*

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)*

Magsar Erdenetuya, *2012* “Fire occurrence and burning biomass statistics in Mongolia”, *Ambassador City Jomtien Hotel Pattaya,*

Roger D. Hungerford, Michael G. Harrington, William H. Frandsen, Kevin C. Ryan, Gerald J. Niehoff, *1990* “Influence of fire on factors that affect site productivity”, *Symposium on Management and Productivity of Western-Montane Forest Soils,*

Sergi Costafreda-Aumedes, Carles Comas, and Cristina Vega-Garcia, *2017* “Human-caused fire occurrence modelling in perspective: a review”, *International Association of Wildland Fire,* Davaajargal Jargalsaikhan, Bayanjargal Darkhijav, Tsolmon Rentsen, 2020 “Estimation of crop suitability using NDVI in the Kherlen Basin Dornod province, Mongolia”, International *Journal of Science, Environment and Technology,* Vol. 10, No 1, 2020, 19-28;

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.

ZHENGMING WAN, 1999. “Modis land-surface Temperature algorithm theoretical basis document”, Institute for Computational Earth System Science University of California, *Santa Barbara,*

Gareth Lames, Daniela Witten and others, “An Introduction to Statistical learning with Applications in R”, *Springer, 2017*

B. Norovsuren., T.Renchin., B.Tseveen., B. Batsaikhan., A. Yangiv., A. Tolmon (2019). Estimation of forest coverage using sentinel and landsat in northern region of mongolia. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, DOI: 10.5194/isprs-archives-XLII-5-W3-71-2019