FLOOD RISK MAPPING USING RANDOM FOREST AND SUPPORT VECTOR MACHINE

M.Ganjirad[1] *, M.R.Delavar[2]

[1]MSc. Student, GIS Dept., School of Surveying and Geospatial Eng., College of Engineering, University of Tehran, Tehran, Iran – mohammad.ganjirad@ut.ac.ir
[2]Centre of Excellence in Geomatic Eng. in Disaster Management, School of Surveying and Geospatial Eng., College of Engineering, University of Tehran, Tehran, Iran – mdelavar@ut.ac.ir

Commission IV, WG IV/3

KEY WORDS: Random Forest, Support Vector Machine, Google Earth Engine, Flood Risk Mapping, Machine Learning

ABSTRACT:

Floods are among the natural disasters that cause financial and human losses all over the world every year. By production of a flood risk map and determination of potential flood risk areas, the possible damages of this phenomenon can be reduced. To map the flood extend in Calcasieu Parish, Louisiana, US, conditioning factors affecting the flood occurrence including elevation, slope, plan curvature, land use, distance from rivers, density of rivers, rainfall, normalized difference vegetation index (NDVI), modified normalized difference water index (MNDWI), and normalized difference built-up index (NDBI) were identified and their information layers produced using the Google Earth Engine (GEE) cloud platform. Then, for flood risk mapping, Random Forest (RF) and support vector machine (SVM) as two machine learning models have been implemented and their results compared. RF and SVM models have been validated based on the maximum absolute error (MAE) index with an accuracy of 0.043 and 0.097, respectively. Visualization of the predicted values in QGIS software confirms that the RF model has provided better outputs than that of the SVM model. By analysing the features importance of the layers in the RF model, it was verified that the elevation, slope, and plan curvature layers have the highest degree of influence on the flood risk with degrees of importance of 0.197, 0.135, and 0.123.

1. INTRODUCTION

Flood is one of the important natural hazards and a special threat to many communities and businesses. Floods can affect all human activities such as economic and social activities, disrupt people's lives, and cause many human and financial losses. It is necessary to provide a flood risk map as one of the most basic information to assist mitigation of the damages caused by the disaster. The flood risk map has a number of usages, among which the following can be mentioned (Marco et al., 1994):

- It provides basic initial information for land use planning;
- It assists informed development of new urban areas;
- The cost of flood compensation and insurance plans can be adequately evaluated by using the flood risk map;
- The feasibility of unstructured flood control measures such as flood proofing can be correctly assessed;
- Flood risk map increases the public awareness of potential risk.

In many countries, e.g. Germany, France, etc., areas where floods have been common in a 100-year period have been identified as flood risk maps (Marco, 1994, Watt, 2000). Due to the dangerous effects caused by floods, the National Flood Insurance Program (NFIP) was approved by the United States Congress in 1968 through the National Flood Insurance Act. This program pursues two main goals: 1) Reducing flood risk through flood damage insurance and 2) Reduction of floodplains in order to reduce flood hazard. In order to identify flood plains, it is necessary to produce flood risk maps. Flood risk can be defined as a probability function of flooding and its

impact (Fernández et al., 2010). The modeling and spatial analyses capabilities provided by geospatial information system (GIS) along with remote sensing (RS) greatly enhance the ability to determine flood risk susceptible zones in the desired domain.

Various techniques with different advantages and disadvantages have been developed in order to mapping a flood risk. The choice of the desired technique for flood risk mapping depends on the data, available capabilities and project requirements (Dung, 2021). An important challenge in the process of flood risk mapping is the modeling of multivariate and non-linear relationships with different degrees of risk levels (Wang et al., 2015). Researchers have used different techniques for flood modeling such as analytic hierarchy process (AHP) (Ouma and Tateishi, 2014; Ghosh and Kar, 2018; Sinha et al., 2008), fuzzy logic (Parsian et al., 2021; Kumar et al., 2020), as well as set pair analysis (SPA) (Zou et al., 2013; Guo et al., 2014; Zeng et al., 2018), and even their combinations (Ekmekcioglu et al., 2021). Weaknesses of various methods should be considered, e.g. There are the high computational for the AHP method, As the number of problem features increases, there will be more pairwise comparisons (Karthikeyan et al., 2016; Oguzitimur 2011). In fuzzy logic, to achieve more accuracy, more fuzzy grades are needed which results to increase exponentially the number of rules. In addition, there is less speed and longer run time of the system (Behrooz et al., 2018). The weight and importance of indicators have an effect on set pair analysis

---

* Corresponding author

201

(SPA) evaluation (Feng and Luo, 2009; Zou et al., 2013). With the development of artificial intelligence (AI), it is possible to use machine learning (ML) algorithms for flood risk mapping. Smart spatial data fusion using machine learning algorithms can provide high level of quality for flood risk modeling and mapping. In recent years, the use of techniques such as Random Forest (RF) (Farhadi and Najafzadeh, 2021; Esfandiari et al., 2020; Wang, 2015), support vector machine (SVM) (Opella and Hernandez, 2019; Mojaddadi et al., 2017) and artificial neural network (ANN) (Avand et al., 2020; Andaryani et al., 2021) has increased in identifying flood risk areas. Each of the above methods has different challenges. In RF, variables with different number of levels, are biased in favor of those attributes with more levels (Prajwala, 2015). SVMs use complex mathematical functions that are difficult for humans to understand (Martens et al., 2007). In ANN, the slow convergence speed is one of the main challenges (Li and Yeh, 2002).

On the other hand, with the development of cloud-based computational systems, a number of benefits were provided to scientists. One of the spatial systems based on the cloud space is the Google Earth Engine (GEE) platform. GEE makes it easier to retrieve and access data and their processes. GEE has the ability to perform a set of spatial analyses at different scales in order to investigate diverse phenomena and issues including deforestation, drought, disease, food insecurity, water scarcity, climate change and global warming. Such processes require high computing power, which makes it possible to use Google servers (Gorelick et al., 2017).

The main objective of this paper is to use and compare RF and SVM models to produce flood risk maps of Calcasieu Parish located in Louisiana, US. In addition, GEE cloud system was used to prepare the required data.

The structure of the remaining parts of this paper is as follows. Section 2 elaborates the research methodology where the RF and SVM models will be explained. Section 3 explains the implementation of the research and the result evaluation. Section 4 concludes the paper and suggests some future areas of the research.

2. MATERIALS AND METHODS

In this section, the study area, RF and SVM models as the employed machine learning models and factors affecting flood risk are discussed.

2.1 Study Area

Louisiana is located in the South-East of United States near the Gulf of Mexico. Calcasieu is situated on the border of the south-western part of the state of Louisiana. According to the last census[1] (US Census, 2020), Calcasieu has a population of 205,282 and an area of 2,833 km². Figure (1) illustrates the study area. Figure (2) shows the main steps in mapping and evaluating the flood risk areas using the employed machine learning algorithms. Between May 17 and 20, 2021, rain in the areas around Lake Charles, Louisiana has caused severe flooding which left 4 death. Lake Charles mayor "Nic Hunter" verified that this was the third heaviest rain event in the history of the city (Louisiana Radio Network, 2021).

Figure (3) shows the damage and water rise in the flood that has been occurred in the Calcasieu Parish of Louisiana on May 17, 2021.

2.2 Machine Learning Models

The use of machine learning algorithms and the implementation of smart spatial data fusion methods improve the quality of the final product of spatial data analyses. In addition, in machine learning algorithms, the learning process is based on data. For this reason, data quality control is important in the machine learning process. In this section, the general structure of the RF and SVM models are reviewed. Then the specification of the collected dataset is discussed.
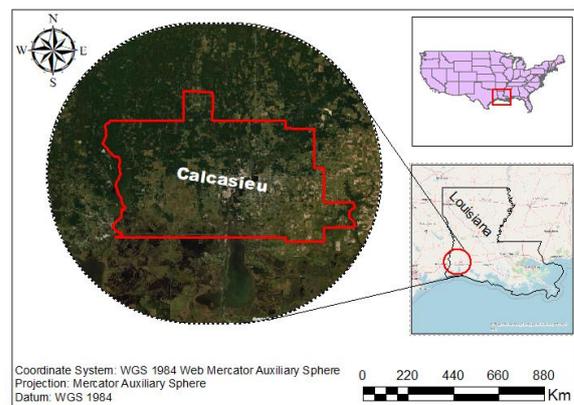
Figure 1. Study area.
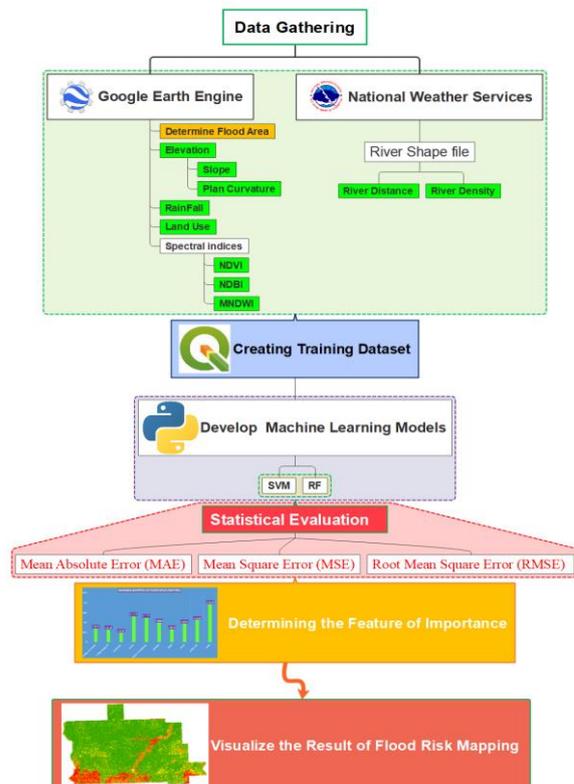
Figure 2. Research methodology.

---

[1] www.census.gov

Figure 3. Flooding and rising water level in residential areas due to heavy rain on May 17, 2021 (Calcasieu Parish Sheriff's Office, 2021).

2.2.1   Random Forest: RF is one of the supervised machine learning algorithms in the field of regression and classification which was introduced by Breiman (2001).

The technique used to reduce the estimated variance is called bagging. Bagging seems to work especially well for high variance, and low bias procedures such as trees in decision tree models. RF is a basic modification of bagging, which is a large collection of trees (Hastie, 2009). In other words, a RF model is a collection of decision trees, as the building block of a RF model is a decision tree (Caigny et al., 2018).

Each tree is trained on a sample of training data. Then, if the goal is classification; prediction is undertaken by majority vote of trees. If the goal is regression, the average output of the trees will be obtained (Figure 4) (Svetnik, 2004).
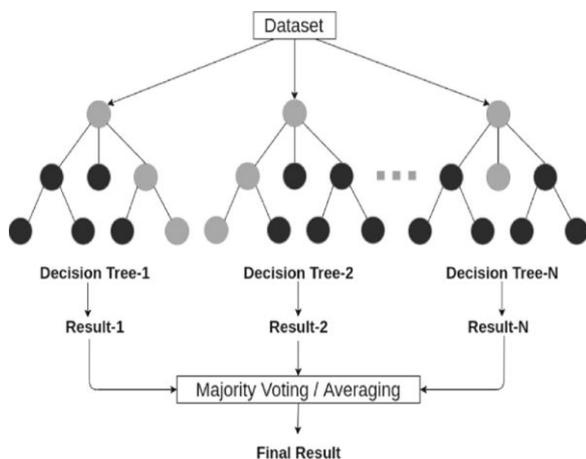
Figure 4.The general structure of RF model (Sarker, 2021)

2.2.2   Support Vector Machine: Support vector machine (SVM) is a supervised machine learning algorithms (Wan and Lei, 2009). SVM is used in classification and regression modelling (Choubin et al., 2019). The model is based on the optimization principle and tries to fit a hyper plane on the training data set for separating the different classes. The hyper plane, is orientated in such a way that is as far as possible from the closest data points from each of the classes. These closest points are called support vectors (Koggalage et al., 2004). A hyper plane is also known as decision boundary. According to Figure (5) the desired kernel is implemented on the features and finally classification is undertaken by applying sigma and bias on the output.
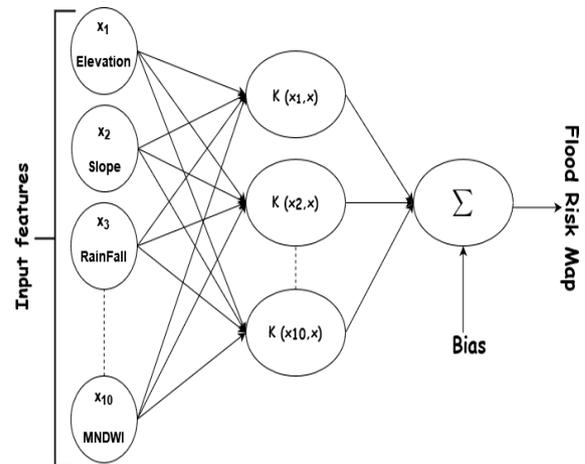
Figure 5. The general structure of the SVM model (Adapted from Xu et al., 2021)

As illustrated in Figure (5), the x_i is the inputs feature, k is the kernel and final result is the flood risk map.

2.3 Dataset

RF and SVM are supervised machine learning algorithms, so preparing a training dataset is essential. The effective features were determined from the previous research for flood risk mapping (Dung et al., 2021).

The initial layers were collected by the Google Earth Engine cloud system and the United States Weather Service[2]. Then, using QGIS software, a training dataset of layers affecting floods with resolution of 10 meters was prepared (QGIS Development Team, 2020). Then, using the collected training data, the RF and SVM models were implemented and the obtained outputs were visualized in the QGIS software. The conditioning factors considered are explained below.

2.3.1   Elevation: Areas with a lower height are more affected by the risk of flooding (Jati et al., 2019). USGS 3DEP National Map data with a resolution of 10 meters has been used to produce a digital elevation model (DEM). This layer is obtained from the Google Earth Engine. The southern and southwestern regions of Calcasieu have lower elevations and the probability of flooding is higher in these regions.

2.3.2   Slope: Generally, Areas with low altitude have a lower slope, the risk of flooding is higher, and can dispose the runoff faster (Tehrany and Kumar, 2018). A slope layer with the degree unit is created from the elevation layer by Raster terrain analysis in QGIS open source software (QGIS Development Team, 2020).

2.3.3   Plan Curvature: Curvature describes the shape of the Earth surface and reflects the water holding capacity. Therefore, the probability of flooding is inversely proportional to the curvature of that area (Costache, 2019). This information layer has been obtained by the elevation data and Curvature tool available in Saga GIS software with a 10-meter resolution (SAGA GIS Development Team, 2020).

---

[2] www.weather.gov/gis

2.3.4 Land Use: One of the effective parameters in flood is the type of land use. Apollonio et al. (2016) have proved that there is a high correlation between flood-prone areas and land use changes, especially in areas where moisture insulation due to urbanization has increased. The European Space Agency (ESA) has provided a land use map with a resolution of 10 meters based on Sentinel-1 and Sentinel-2 data in 11 different classes in 2020-2021. In this study, by the Google Earth Engine cloud system, this product has also been prepared for Calcasieu Parish**.**

2.3.5 Rainfall: The amount of rainfall is directly related to the probability of flooding. It should be noted that the total amount of rainfall in a period of time has a greater effect than the intensity of rainfall (Bracken et al., 2008). In order to prepare the total amount of rainfall on May 17, 2021, the GSMap product, which provides the hourly rate of rainfall with a resolution of 0.1 x 0.1 degrees, has been used in the Google Earth Engine system. In order to be compatible with other data, the pixel size of this layer was changed to 10 meters in the pre-processing operation.

2.3.6 River Distance and Density: The density of rivers and waterways significantly affects the time of concentration and magnitude of the flow. In other words, increasing the density of waterways and rivers will increase flood peaks (Pallard et al., 2009). In addition, The risk of flooding is higher in areas close to rivers.

To prepare the data layers, first, the shape file of the US Rivers from the National Weather Service was downloaded[3], and then based on the rivers of the Calcasieu Parish, the density and distance layers from the rivers were computed.

2.3.7 Spectral Indices: A number of spectral indices for completing the training dataset can be computed. In this study, 3 indices including Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973), Modified Normalized Difference Water Index (MNDWI) (Xu, 2007), and Normalized Difference Built-up Index (NDBI) (Zha et al., 2003), along with other features based on Sentinel-2 images with a resolution of 10 meter, were produced and extracted by Google Earth Engine system on May 17, 2021. In the following, these spectral indices will be introduced.

• *Normalized Difference Vegetation Index (NDVI)*

In NDVI, negative values indicate water and positive values indicate vegetation. NDVI has an inverse relationship with floods where higher NDVI values indicate a lower probability of flooding. NDVI is calculated using Eq. (1) (Ullah et al., 2020).

$$NDVI = \frac{NIR - RED}{NIR + RED} = \frac{B8 - B4}{B8 + B4} \qquad (1)$$

where, B4/B8 is the 4th /8th band of Sentinel-2 in this study.

• *Normalized Difference Built-up Index (NDBI)*

NDBI index is used to analyze built-up areas. Negative values of NDBI represent waters bodies and positive values represent built-up areas. NDBI is calculated based on Eq. (2) (Zha et al., 2003).

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} = \frac{B11 - B8}{B11 + B8} \qquad (2)$$

where, B11/B8 is the 11th /8th band of Sentinel-2.

• *Modified Normalized Difference Water Index (MNDWI)*

The modified NDWI (MNDWI) can detect water bodies and eliminate some noises (Eq. 3). In this index water bodies have positive values (Xu, 2007).

$$MNDWI = \frac{Green - SWIR}{Green + SWIR} = \frac{B3 - B11}{B3 + B11} \qquad (3)$$

where, B3/B11 is the 3th /11th band of Sentinel-2 in this study.

2.3.8 Target Classes: In order to complete the training dataset, it is necessary to determine the flooded footprints. For this purpose, the algorithm proposed by Notti (2018) which was developed under the United Nations programs in the Google Earth Engine cloud system, was employed in this paper. This algorithm uses Sentinel-1 SAR Data. SAR sensors work independently of weather conditions. They do not need sunlight and produce images at all hours of the day and night. First, the study area is determined by uploading the shape file of Calcasieu Parish. Then the desired time frame is determined. Using the studies conducted and the United States Weather Service reports, the time frame is considered during May 17-20, 2021. Then the algorithm is executed, and a binary image with a resolution of 10 meters is prepared from the flooded areas. Areas with values of 1 are flooded, and areas with values of 0 are not flooded. 4500 points were extracted from this image, 235 of which were flooded and the rest were non-flooded points (Figures 6).

Figures (7) illustrate the layers affecting the flood, and Table (1) provides metadata information including coordinate system, projection, datum, unit and pixel size of the employed data layers.

Coordinate System	*USA Contiguous Albers Equal Area*
Projection System	*Albers*
Datum	*North American Datum 1983*
Unit	*Meter*
Pixel Size	*10 meters*

Table 1. Metadata of training dataset.
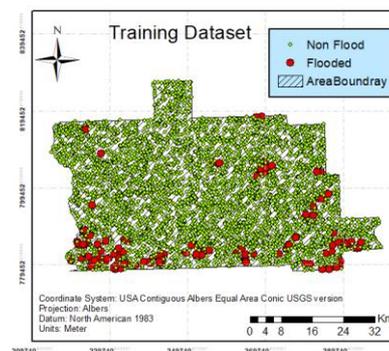
Figure 6. The obtained training data flooded and non-flooded points.

---

3 www.weather.gov/gis/Rivers
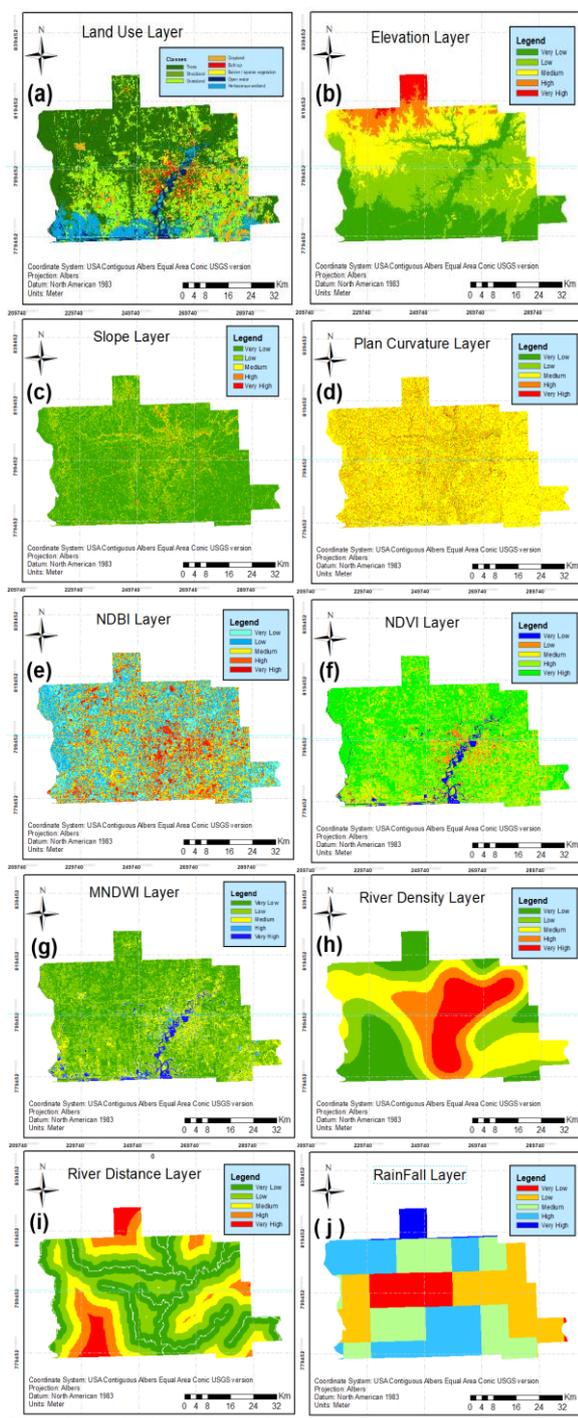
Figure 7. Set of effecive spatial layers in flooding (a) Land Use, (b) Elevation, (c) Slope, (d) Plan Curvature, (e) NDBI, (f) NDVI, (g) MNDWI, (h) River Density, (i) River Distance and (j) RainFall.

3. RESULT

By completing the training dataset, SVM and RF models have been implemented by using the Python programming language. This study is considered as a regression problem. The final goal is to determine the probability of flooding in each of the pixels of the domain. 70% of the data is considered as training and 30% as the test data. It should be noted that in

RF model, the number of trees was 128. Also, the radial basis function (RBF) kernel is implemented in the SVM model.

Statistical indices including Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) were calculated to evaluate the model in training and validation processes according to Eq. (4, 5, and 6) (Chai and Draxler, 2014).

$$MAE = \frac{1}{n}\sum(y - \hat{y}) \tag{4}$$

$$MSE = \frac{1}{n}\sum(y - \hat{y})^2 \tag{5}$$

$$RMSE = \sqrt{MSE} \tag{6}$$

where, n = number of observations.
      y = actual value.
      ŷ = predicted value.

According to Table (2), RF has performed better than SVM in the test and validation steps. The error rate in the learning step based on the MAE was 1.73% for RF and 9.53% for SVM. When the model is trained, the validation process is performed on 1350 reference points (30% of the training data). According to Table (2), MAE, MSE, and RMSE indices were calculated on the results of both of the models. For example, the validation result for the RF model based on the index of MAE has a 4.32% error. While the same error rate for the SVM model has reached to 9.76%.

Models	MAE Training	MAE Validation	MSE Validation	RMSE Validation
RF	0.0173	0.0432	0.0219	0.1481
SVM	0.0953	0.0976	0.0292	0.1711

Table 2. Error rates of RF and SVM models in the process of training and validation.

The output maps were classified into 5 classes including very low, low, medium, high and very high flood risk areas (Figure 8 and 9). The results illustrate that the RF model produces more accurate output than that of the SVM. The risk map on the RF model has been able to distinguish different areas based on their features in the flood disaster. For example, the southwestern areas of Calcasieu, which had low elevation and frequent floods, are in the very high-risk class.

Another evaluation that can be performed in the RF model is determining the relative importance of the employed conditioning factors (features). The importance of the features is determined using a sensitivity analysis that calculates a score for all the input features.

The importance of input parameters is determined by scores. Any feature with a higher score means that feature has a greater impact in predicting the flood risk. The sum of scores for all features is equal to one. According to Figure 10, the features of elevation, slope, and plan curvature have the greatest effects on the occurrence of floods. The importance of these features is 0.1931, 0.1312, and 0.1230, respectively.
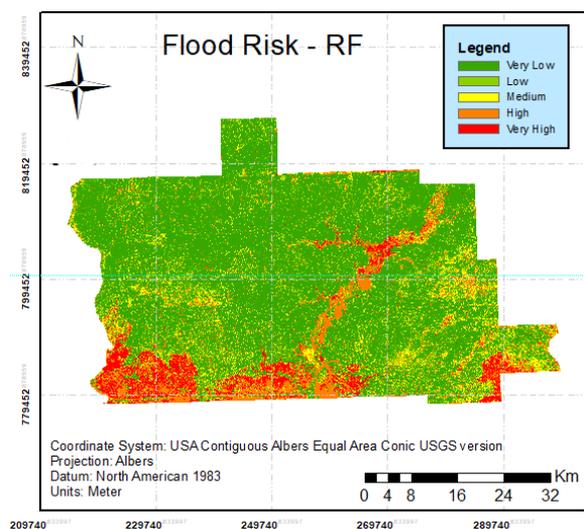
Figure 8. Flood susceptibility mapping by RF model.
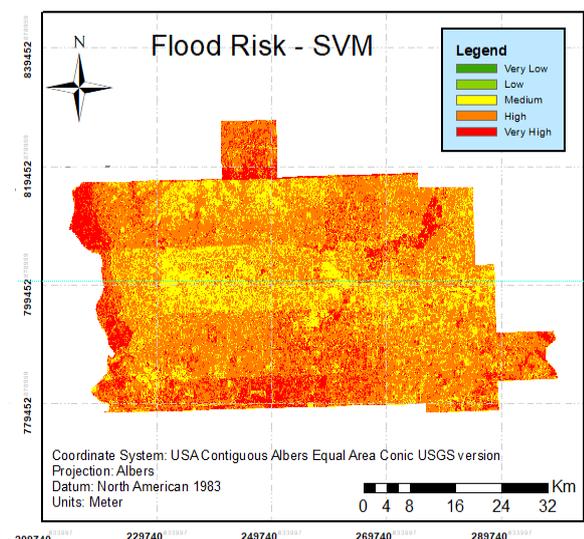
Figure 9. Flood susceptibility mapping by SVM model.
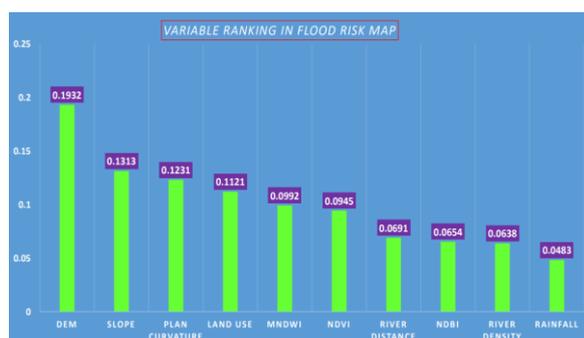
Figure 10. Bar chart for Relative importance of the features.

4. CONCLUSION AND FUTURE DIRECTIONS

Identifying factors affecting floods and producing a risk map is one of the basic solutions to control this natural disaster. To estimate the severity of flood risk, several methods have been presented by different researchers. In this study, by implementing machine learning models including RF and SVM, the flood risk maps in Calcasieu section of Louisiana State have been produced.

According to the archived results of this study, 10 effective factors including elevation, slope, plan curvature, land use, NDVI, NDBI, MNDWI indices, distance from rivers, river density, and rainfall intensity were obtained using the Google Earth Engine cloud system and USA National Weather Services.

After performing the necessary pre-processing and creating the training dataset in QGIS software, RF and SVM machine learning models were implemented by Python programming language and the probability of flood occurrence in each pixel of the domain was estimated. The results of the machine learning models are highly dependent on the employed training dataset. As a result, the quality and size of the training data is very important. The size of the training data should be enough to complete the learning process and there should be no duplicate data (Brownlee, 2019). As a result, in order to develop a machine learning model to produce a flood risk map, it is necessary to pay attention to the amount and characteristics of the training data available in the target area. In this study, a training dataset with a resolution of 10 meter was prepared. It is clear that as the data quality decreases, the quality of the results of the analyses also decreases.

The outputs of the models based on the employed statistical indices (MAE, MSE and RMSE) proved that the RF provides more accurate results and has a higher efficiency than those of the SVM model. In addition, by examining the features importance considered in the RF model, it was found that the elevation, slope, plan curvature, and land use have greater effects on the occurrence of floods than those of other inputs. These four effective factors compared to the other ten factors account for 55% of the flood risk. It should be noted that the order of feature importance depends on the geographical area and in different researches that have been undertaken in different places, the order of feature importance has changed (Wang et al., 2015; Farhadi and Najafzadeh, 2021).

In future research, it is suggested that by removing the less important factors in the flood risk mapping and adding some other influencing layers of information such as soil type, river water discharge rate, land subsidence map and population distribution map, a more accurate estimate of the flood occurrence may be obtained. In addition, it is useful to use synthetic aperture radar (SAR) measurements that operate independently of weather conditions and can provide valuable information related to flood assessment. The reason is that water surfaces do not reflect in the microwave range, and water surfaces appear black in SAR images. Furthermore, the use of deep learning algorithms has become common in environmental fields (Lamba et al., 2019; Hassan et al., 2020). Deep learning networks perform the feature extraction process automatically. In addition, machine learning models such RF and SVM are pixel-based approaches, while deep learning methods are patch-based models (Guo et al., 2022; Helber et al., 2019). For this reason, the learning process is carried out in deep learning networks based on various patches that have been extracted. In addition to the methods proposed in the field of data science, physical models such as weather research and forecasting model (WRF) (UCAR and NCAR, 2020) are used in the preparation of flood hazard maps. These physical models are implemented based on physical schemas at different regions, and modeling is undertaken based on the physical conditions of the region. Using smart data fusion and integrating data from machine/deep learning with physical models is also recommended for flood risk mapping.

ACKNOWLEDGEMENTS

REFERENCES

Andaryani, S., Nourani, V., Haghighi, A.T., Keesstra, S., 2021: Integration of hard and soft supervised machine learning for flood susceptibility mapping. *Journal of Environmental Management*. 291, 112731.

Apollonio, C., Balacco, G., Novelli, A., Tarantino, E., Piccinni, A.F., 2016: Land Use Change Impact on Flooding Areas: The Case Study of Cervaro Basin (Italy). *Sustainability*. 8(10). doi.org/10.3390/su8100996.

Avand, M., Moradi, H.R., Ramazanzadeh Lasboyee, M., 2021: Spatial Prediction of Future Flood Risk: An Approach to the Effects of Climate Change. *Geosciences*. 11(1).

Behrooz, F., Mariun, N., Marhaban, M., Mohd Radzi, M.A., Ramli, A., 2018: Review of Control Techniques for HVAC Systems—Nonlinearity Approaches Based on Fuzzy Cognitive Maps. *Energies*. 11, 495. doi.org/10.3390/en11030495.

Bracken, L.J., Cox, N.J., Shannon, J., 2008: The relationship between rainfall inputs and flood generation in south-east Spain. *Hydrological processes*. 22(5), 567-723.

Brownlee, J., 2019: How to choose a feature selection method for machine learning. *Machine Learning Mastery*. 10.

Chai, T., Draxler, R.R., 2014: Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific model development*. 7(3), 1247-1250.

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A., 2019: An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of The Total Environment*. 651, 2087-2096. doi.org/https://doi.org/10.1016/j.scitotenv.2018.10.064.

Costache, R.-D., 2019: Flood Susceptibility Assessment by Using Bivariate Statistics and Machine Learning Models - A Useful Tool for Flood Risk Management. *Water Resources Management*. 33, 3239–3256. doi.org/10.1007/s11269-019-02301-z.

De Caigny, A., Coussement, K., De Bock, K.W., 2018: A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*. 269(2), 760-772.

Dung, N.B., Long, N.Q., Goyal, R., An, D.T., Minh, D.T., 2021: The Role of Factors Affecting Flood Hazard Zoning Using Analytical Hierarchy Process: A Review. *Earth Systems and Environment*. doi.org/10.1007/s41748-021-00235-4.

Ekmekcioğlu, Ö, Koc, K., Özger, M., 2021: District based flood risk assessment in Istanbul using fuzzy analytical hierarchy process. *Stochastic Environmental Research and Risk Assessment*. 35(3), 617-637..

Esfandiari, M., Abdi, G., Jabari, S., McGrath, H., Coleman, D., 2020: Flood Hazard Risk Mapping Using a Pseudo Supervised Random Forest. *Remote Sensing*. 12(19).

Farhadi, H., Najafzadeh, M., 2021: Flood Risk Mapping by Remote Sensing Data and Random Forest Technique. *Water*. 13(21). doi.org/10.3390/w13213115.

Feng, L., Luo, G., 2009: Practical Study on the Fuzzy Risk of Flood Disasters. *Acta Applicandae Mathematicae*. 106(3), 421-432. doi.org/10.1007/s10440-008-9305-4.

Fernández, D.S., Lutz, M.A., 2010: Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. *Engineering Geology*. 111(1), 90-98. doi.org/10.1016/j.enggeo.2009.12.006.

Ghosh, A., Kar, S.K., 2018: Application of analytical hierarchy process (AHP) for flood risk assessment: a case study in Malda district of West Bengal, India. *Natural Hazards*. 94(1), 349-368. doi.org/10.1007/s11069-018-3392-y.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017: Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. 202, 18-27.

Guo, E., Zhang, J., Ren, X., Zhang, Q., Sun, Z., 2014: Integrated risk assessment of flood disaster based on improved set pair analysis and the variable fuzzy set theory in central Liaoning Province, China. *Natural Hazards*. 74(2), 947-965. doi.org/10.1007/s11069-014-1238-9.

Guo, Z., Moosavi, V., Leitão, J.P., 2022: Data-driven rapid flood prediction mapping with catchment generalizability. *Journal of Hydrology*. 609, 127726.

Hassan, M.M., Gumaei, A., Alsanad, A., Alrubaian, M., Fortino, G., 2022: A hybrid deep learning model for efficient intrusion detection in big data environment. *Information Sciences*. 513, 386-396.

Hastie, T., Tibshirani, R., Friedman, J., 2009: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer New York.

Helber, P., Bischke, B., Dengel, A., Borth, D., 2019: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 12(7), 2217-2226.

Jati, M., Suroso, S., Santoso, P., 2019: Prediction of flood areas using the logistic regression method (case study of the provinces Banten, DKI Jakarta, and West Java). *Journal of Physics: Conference Series*. 1367, 012087.

Koggalage, R., Halgamuge, S., 2004: Reducing the number of training samples for fast support vector machine classification. *Neural Information Processing-Letters*. 2(3), 57-65.

Kumar, N., Goyal, P., Kapil, G., Agrawal, A., Ahmad Khan, R., 2020: Flood risk finder for IoT based mechanism using fuzzy logic. *Materials Today: Proceedings*.

Lamba, A., Cassey, P., Segaran, R.R., Koh, L.P., 2019: Deep learning for environmental conservation. *Biology*. 29(19), R977-R982.

Li, X., Yeh, A.G.-O., 2002: Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*. 16(4), 323-343.

Marco, J.B., 1994: Flood risk mapping, in: Rossi, G., Harmancioğlu, N., Yevjevich, V. (Eds.), Coping with Floods. Springer Netherlands, Dordrecht, pp. 353-373.

Martens, D., De Backer, M., Haesen, R., Vanthienen, J., Snoeck, M., Baesens, B., 2007: Classification with ant colony optimization. *IEEE Transactions on Evolutionary Computation.* 11(5), 651-665.

Notti, D., Giordan, D., Caló, F., Pepe, A., Zucca, F., Galve, J.P., 2018: Potential and Limitations of Open Satellite Data for Flood Mapping. *Remote Sensing*. 10(11).

Opella, J.M.A., Hernandez, A.A., 2019: Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework, *2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA)*. pp. 260-265.

Ouma, Y.O., Tateishi, R., 2014: Urban Flood Vulnerability and Risk Mapping Using Integrated Multi-Parametric AHP and GIS: Methodological Overview and Case Study Assessment. *Water*. 6(6), 1515-1545.

Pallard, B., Castellarin, A., Montanari, A., 2009: A look at the links between drainage density and flood statistics. *Hydrol. Earth Syst. Sci*. 13(7), 1019-1029.

Parsian, S., Amani, M., Moghimi, A., Ghorbanian, A., Mahdavi, S., 2021: Flood Hazard Mapping Using Fuzzy Logic, Analytical Hierarchy Process, and Multi-Source Geospatial Datasets. *Remote Sensing*. 13(23).

Sarker, I., 2021: Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN Computer Science*. 2. doi.org/10.1007/s42979-021-00765-8.

Sinha, R., Bapalu, G.V., Singh, L.K., Rath, B., 2008: Flood risk analysis in the Kosi river basin, north Bihar using multi-parametric approach of Analytical Hierarchy Process (AHP). *Journal of the Indian Society of Remote Sensing*. 36(4), 335-349. doi.org/10.1007/s12524-008-0034-y.

Svetnik, V., Liaw, A., Tong, C., Wang, T., 2004: Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules, in: Roli, F.,

Kittler, J., Windeatt, T. (Eds.), Multiple Classifier Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 334-343.

T R, P., 2015: A Comparative Study on Decision Tree and Random Forest Using R Tool. IJARCCE. 196-199. doi.org/10.17148/IJARCCE.2015.4142.

Ullah, K., Zhang, J., 2020: GIS-based flood hazard mapping using relative frequency ratio method: A case study of Panjkora River Basin, eastern Hindu Kush, Pakistan. *PLOS ONE*. 15, e0229153. doi.org/10.1371/journal.pone.0229153.

Wan, S., Lei, T.C., 2009: A knowledge-based decision support system to analyze the debris-flow problems at Chen-Yu-Lan River, Taiwan. *Knowledge-Based Systems*. 22(8), 580-588. doi.org/10.1016/j.knosys.2009.07.008.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X., 2015: Flood hazard risk assessment model based on random forest. *Journal of Hydrology*. 527, 1130-1141.

Xu, D., Qi, X., Li, C., Sheng, Z., Huang, H., 2021: Wise Information Technology of Med: Human Pose Recognition in Elderly Care. *Sensors*. 21,7130.

Xu, H., 2006: Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*. 27(14), 3025-30.doi.org/10.1080/01431160600589179.

Zha, Y., Gao, J., Ni, S., 2003: Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing* - INT J REMOTE SENS. 24, 583-594.

Zou, Q., Zhou, J., Zhou, C., Song, L., Guo, J., 2013: Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. *Stochastic Environmental Research and Risk Assessment*. 27(2), 525-546. doi.org/10.1007/s00477-012-0598-5.