Using of machines learning in extraction of urban roads from DEM of LIDAR data: Case study at Baghdad expressways, Iraq

Oday Zakariya Jasim

University of Technology, Civil Engineering Department, Baghdad, Iraq

ABSTRACT

Road extraction from remotely perceived information may be a difficult issue and has been approached in many various ways in which by photogrammetrists and digital image processors. This study was given extraction of roads from DEM of LiDAR with IKONOS image using machines learning (ML). Two set of data were used in this study. IKONOS Image and Digital Elevation Model (DEM) data were combined to produce thematic mapping. The spatial resolution of data is 1 m and was acquired on 2010. The result f thematic map based on theses images and the methods was used three models of machines learning. The problem of this study, when was used the LIDAR data to extract the road is very difficult because the LiDAR data is too noisy and employed it also so hard. Moreover, this article will describe an effective and compare between several machines learning algorithms (RF, BYO and MLP) for detection the roads from LIDAR data. The statistical indictors such as an overall accuracy (OA), kappa analysis statistic (K), (MAE) which is Mean Absolute Error and finally the (RMSE) which is Root Mean Squared Error. All these will be use to get the accuracy of classification assessment and the best model to produce the thematic map.

Keywords:	Remote sensing, LIDAR, Machines learning, DEM, Random Frost Model (RF), Bayesian Optimization (BYO), Multi layer Perceptron (MLP).

Corresponding Author:

Oday Zakariya Jasim

University of Technology, Civil Engineering Department, Baghdad, Iraq

Address. Baghdad, Iraq.

E-mail: 40004@uotechnology.edu.iq

1.1. Introduction

Light Detection and Ranging (LIDAR) Remote sensing is the technique of obtain remotely the electromagnetic radiation emitted or reflected by a surface without direct contact with it [1]. Also LIDAR is a technique for acquisition of 3D data from the surface of the terrain. With the Geographic Information System, extraction and classification roads can be made to the land or in any area [2]. Road information is highly imperative in Geographical Information System (GIS) and play significant role in transport application systems [3]. The road extraction from any satellite images with different resolutions can be cause by the different spatial resolutios of the images or spectral resoulutions, the type of sensor, and ground features [4]. So the new Machine learning is a technique that utilized to detect, recognize and extract data or land uses from the images [5]. Many approaches to road extraction have been attempted in the past. However, it was noted that different classification outcomes are achieved from the use of different classifiers. Hence, many investigations already done to determine the activities of different classifiers [6]. However, in machine learning, at least three major factors that influence accuracy of classification should be considered. Especially, selection of training samples, feature selection and settings of tuning parameters [7]. Although, the three aforementioned factors have been investigated in the past, only few investigations in relation to settings of tuning parameters have been witnessed. But, the first step and the most important among the factors use in machine learning is the setting tuning parameters. Furthermore, machine learning classifiers studies shows

that comparison between previous studies are focused mostly on grasslands, farmlands and coal mining areas which are predominantly non-urban areas.

1.2 Objective of the research

The objective of this paper is to assess to the three ways of machines learning (ML) for road and nonroad of urban landcover classification, with the approach of pixel-based by using satellite image with high spatial resolution with DEM of Lidar data. An investigation will be done to how tuning parameters and affect the the results of classification such as hidden layers, optimization algorithm analysis and the activation function analysis for hidden and output layers. As well as, comparing the proposed model (Multilayer Perceptron (MLP)) with other models of machine learning such as random frost model (RF) and bayesian Optimization (BYO). The results of models based on the statistical indictors such as (OA),Overall Accuracy, (K),Kappa statistic, (MAE), Mean Absolute Error and (RMSE), Root Mean Squared Error are used to get the accuracy assessment of the classification.

2. Methodology

2.1 Case study (area of intrest)

The case study or the AOI of this paper is in Baghdad capital of Iraq with a part of district AL-Rusafa (Figure 1) and located in Longitude 44°.45, 44°.46 decimal degrees east of Greenwich and Latitude 33°.34, 33°.34 decimal degrees north of equator It is an old quarters and in the center in Baghdad, while the expressway of the AOI is also shown in the figure below.

Figure 1. Case study area

2.2 Datasets

A data of IKONOS satellite image and LIDAR (DEM) will be in image processing to get a final digital thematic mapping. In figure 2 the case study data of IKONOS image with the spatial resolution of one meter and image capture date in 2010 while the LIDAR (DEM) with spatial resolution of one meter and also was captured on 2010.
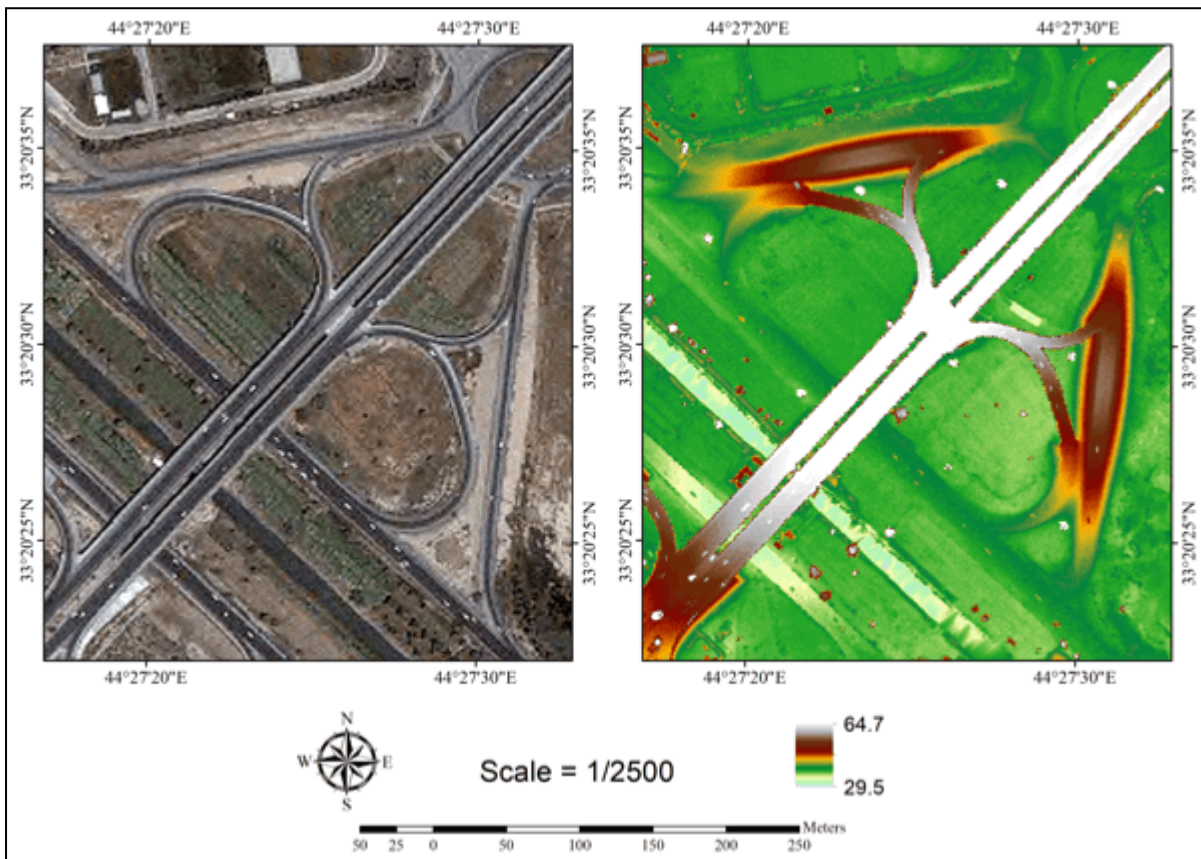

Figure 2. IKONOS Image and LIDAR (DEM) for the case study

2.3 Method

A three models will be use in this study as an extraction urban roads new techniueqs, random frost model (RF), bayesian optimization (BYO) and multilayer perceptron (MLP). After preparing the data in the above steps and will be ready to as input to these three models, we will be dived the data two sets are training 70% and 30% testing of data. in this study we will select the best model through evolution these models by calculate the Overall Accuracy (OA), Kappa analysis statistic (K) [8], (MAE) Mean Absolute Error and finally (RMSE)Root Mean Squared Error (RMSE). The figure 3 shows methodology flow chart.
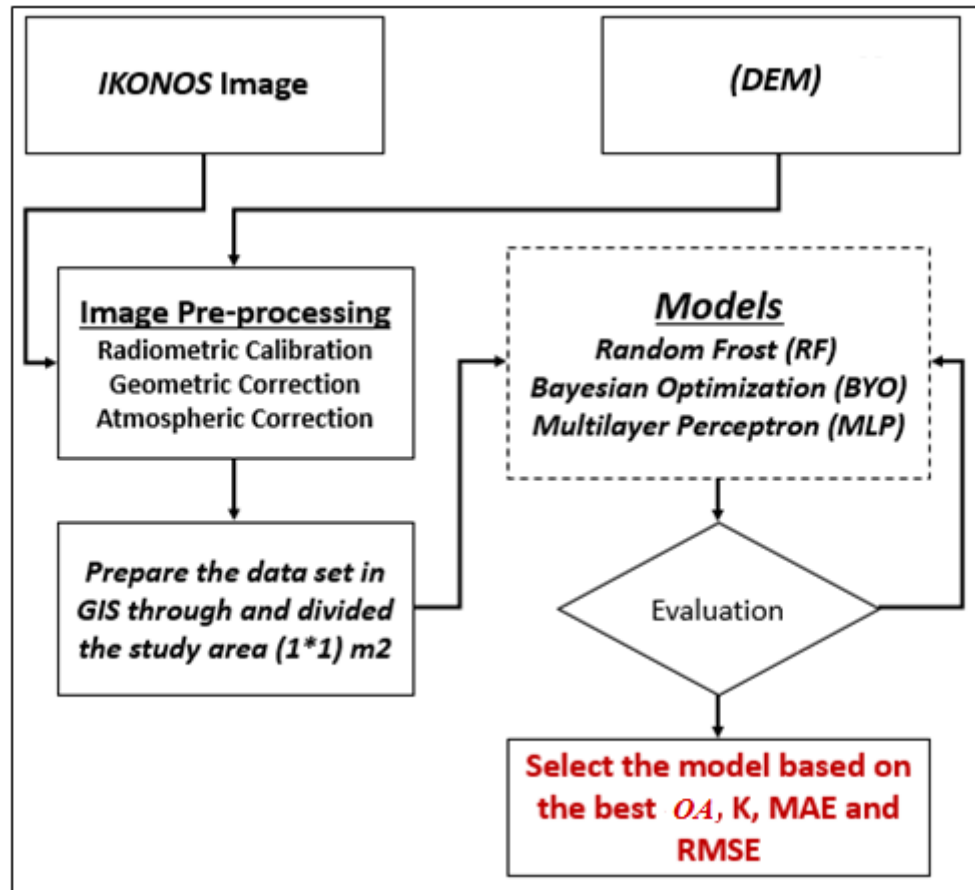
Figure 3. Methodology flow chart

2.3.1 Random frost model (RF)

RF is a machine-learning model based on randomized decision trees [9]. It provides a direct measurement of feature importance. In addition, RF can be used for both classification and regression tasks, which makes it a good model for a variety of tasks. RF has predominant execution on high dimensional information since it performs an implicit feature selection using a small subset of the training data. An indicator called "Gini importance" is often used to visualize the result of this certain component choice process. This component significance score gives a relative positioning of the otherworldly highlights, and is – in fact – a result in the preparation of RF classifier: At every hub τ inside the parallel trees T of RF, the ideal split is looked for utilizing the Gini polluting influence i (τ) – a computationally effective estimation to the entropy – estimating how well a potential split is isolating the examples of the two classes in this specific hub.

2.3.2 Bayesian optimization (BO)

Many issues have made it challenging to obtain an acceptable spectral index for enhancement for any feature and extraction. Firstly, the required coefficients needed to establish the spectral bands together with the index are mostly continuous variables. Hence, many rounds of iterations are necessary to achieve the best variables. The optimization processes based on trial-and-error are very slow. Second, the coupling effects arise in the determination of the coefficients which would lead to delay in identifying the optimal range of values. Therefore, a vigorous optimization approach is necessary to enable effective attainment of the best values through mathematical index.

Bayesian optimization is a smart way used to maximize or minimize a cost function compared to the conventional method refers to as grid and random search approach [10]. This approach has been used widely in machine learning algorithms to optimize the hyperparameters. Therefore, in this work, it is to be improve a spectral index for extraction of any features. This method differs from grid and random search that employs all possible patterns to select the best possible combination of values for a particular assignment. In Bayesian optimization, decision is made based on next parameter configuration to be evaluated. This brings in the issue of objective function and historical data to learn the way former configurations accomplish results. Interestingly, Bayesian optimization determines the best possible parameter configuration swiftly even faster than other approaches.

2.3.3 Multilayer perceptron (MLP)

MLP is one of the machines learning family that is based on biological and statistical learning approach. It a simple system that consists interconnected nodes or neurons [11] as shown in Figure 4. The model represents a non-linear mapping between inputs and outputs that are systematically structured in layers with random or full interconnections between layers. Theoretically, there exist three layers in the system that receive input and process results: input, hidden and output layers. The nodes are generally assigning numeric weights which are process as output signals. The node inputs are modified using an activation function.

The possibility of learning in ANN model has attracted more attention to it use in many researches works. It minimizes error function in an iterative manner through a designed gradient algorithm shown in (Equation 1). ANN has recorded a tremendous success in remote sensing applications, although with limitations in terms of computational complexity and drawback of overlearning [12].

$$E = \frac{1}{2} \sum_{i=1}^{L} \left(d_j - o_j^M \right)^2 \tag{1}$$

In which, d_j and o_j^M denotes the required output and response at the node j in the output layer, respectively, and L represents the number of nodes in the output layer. Generally, corrections are made to weight parameters through computation which are added to the initial outcomes as shown in (Equation 2):

$$\begin{cases} \Delta w_{i,j} = -\mu \dfrac{\partial E}{\partial w_{i,j}} \\ \Delta w_{i,j}(t+1) = \Delta w_{i,j} + \alpha \Delta w_{i,j}(t) \end{cases} \tag{2}$$

In which, $w_{i,j}$ denotes weight parameter obtained between i and j nodes. The Δ is a constant generally positive and regulates the extent of adjustment required called learning rate. The sign α refers to a momentum factor that accommodates only values between 0 and 1, while ''t'' represents the number of iterations made. The parameter α is considered as stabilizing factor or smoothing as it streamlines the rapid changes in the weights.
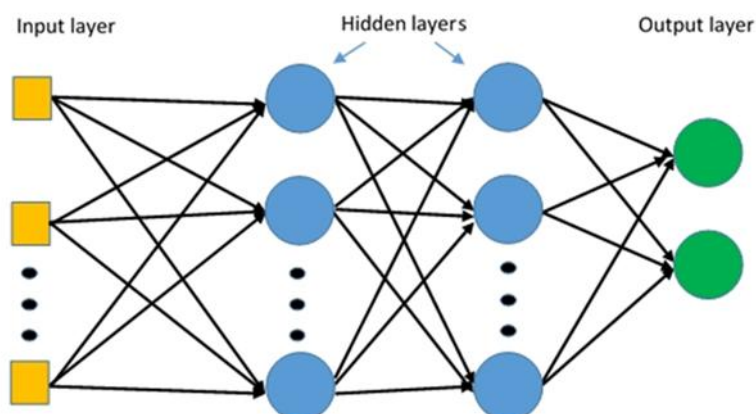
Figure 4. Architecture of typical MLP

2.4 Methods of evaluation

The Error matrix analysis is used to determine the Kappa statistic (K) and Overall Accuracy (OA) in other to achieve the best classification accuracy. The computation of OA is achieved by dividing the total sum of the major diagonal by the total number of pixels available in the error analysis matrix. While, the Kappa analysis is a technique of discrete multivariate deployed in many accuracy assessments. Kappa analysis is a measure of accuracy yields or K statistic is calculated based on Equation 1 [13]:

$$K = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_i + x_{+i})}{N^2 - \sum_{i=1}^{r}(x_i + x_{+i})} \tag{1}$$

Where the number of rows in the matrix is represented by r. The symbol x_{ii} stand for the number of observations in the row i. While column i, x_{i+} and x_{+i} represent the marginal totals for both row i and column i and N denotes the total number of pixels.

(MAE) and (RMSE) are used in statistical analysis. The MAE estimates the average error magnitude in a forecast set regardless of their direction. It also assesses accuracy level for continuous variables with the equation as provided in the library references.A linear score, of MAE will represent in which indicates that all individual differences are weighted equally in the average. However, the RMSE is based on quadratic scoring rule that assesses the average magnitude of an error. The equation for the RMSE is provided in the references. The RSME can also be define as the difference between the corresponding forecast and observed values each squared over the average number of samples. The square root of the average is finally taken since the errors are squared before averaged. The RMSE provides a high weight to large errors. This indicates that RMSE is more important when large errors need to be eliminated.

3. Results

Through the table 1 the best training model is MLP model than other models. Through the MLP has the highest accuracy (97.7) and kappa (0.95) with the lest for mean absolute error (0.04) and rout mean square error (0.13) for training model. While, the second model RF is better than OYB model through RF model has high Acc (94.81) and K (0.89) with less MAE (0.8) and RMSE (0.2). Whereas, the worst model in this study is the OBY through has the least ACC with K and highest MAE with RMSE.

As well as, the table 2 shows the testing models for this study. And also, we found the MLP is the best model for testing the other models through achieved the highest Acc with K and lest MAE and RMSE which are 96.24, 0.93, 0.06 and 0.16 receptively. Hence, through two tables for training and testing models we found the MLP is the proposed model which overcome other models (OBY and RF).

Table 1. Evaluation of training models

Models	OA	K	MAE	RMSE
RF	94.81	0.89	0.08	0.20
OYB	88.44	0.76	0.22	0.31
MLP	97.70	0.95	0.04	0.13

Table 2. Evaluation of models for testing

Models	OA	K	MAE	RMSE
RF	92.73	0.88	0.10	0.37
OYB	85.67	0.74	0.25	0.45
MLP	96.24	0.93	0.06	0.16

Table 3 shows the datils accuracy for the proposed model (MLP), the TP rate is true positives which are the number of the predicted positive examples that are actually positive. False positives FP rate is the which are the number of examples predicted positive that are actually negative. As well as, the TN is the true negative which represents the number of negative predictions that are truly negative, and the FN is the false negative which indicates the number of negative predictions that are truly positive.

On the other hand, F-measure is measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$F = 2.\frac{precision.recall}{precision + recall} \qquad (2)$$

$$Recall = \frac{tp}{tp + fn} \qquad (3)$$

$$Precision = \frac{tp}{tp + fp} \qquad (4)$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \qquad (5)$$

Table 3. Detailed accuracy by class for the proposed models (MLP)

	Rate of TP	Rate of FP	Precision	Recall	F-Measure	Class
	0.336	0.133	0.632	0.336	0.439	Road
	0.867	0.664	0.658	0.867	0.748	NonRoad
Weighted Avg.	0.652	0.449	0.647	0.652	0.623	

Table 4 shows a contingency table which is also generally regarded as a confusion matrix. In this regard, there exist two classes. Hence, it can be represented by a 2 * 2 confusion matrix which can be arbitrarily large. The number of properly classified instances can be computed by summing the diagonals found in the matrix. This implies that other parameters in the matrix are wrongly classified (class "a" represents misclassified; "b" exactly twice; class "b" gets misclassified as "a" three times).

Table 4. Confusion matrix for the proposed models (MLP)

a	b	< -- classified as
8850	17505	a = Road
5143	33629	b = NonRoad

True Positive (TP) rate refers to the ratio of classified examples as x, out of the total number of examples that truly belong to class x. This identifies the actual part of the class that has been captured. The False Positive (FP) rate is the ratio of classified examples in class x, but does not belong the class from among examples that are from different class other than class x. The matrix is computed by summing the column of class x and subtract the diagonal members, then divide by the sum of rows of all other classes. Lastly, Precision is defined as the ratio of the examples that truly belong to class x over all others classified as class x, but, may not necessarily belong to the class. So, in this case, the diagonal elements were divided into the sum of the relevant column.
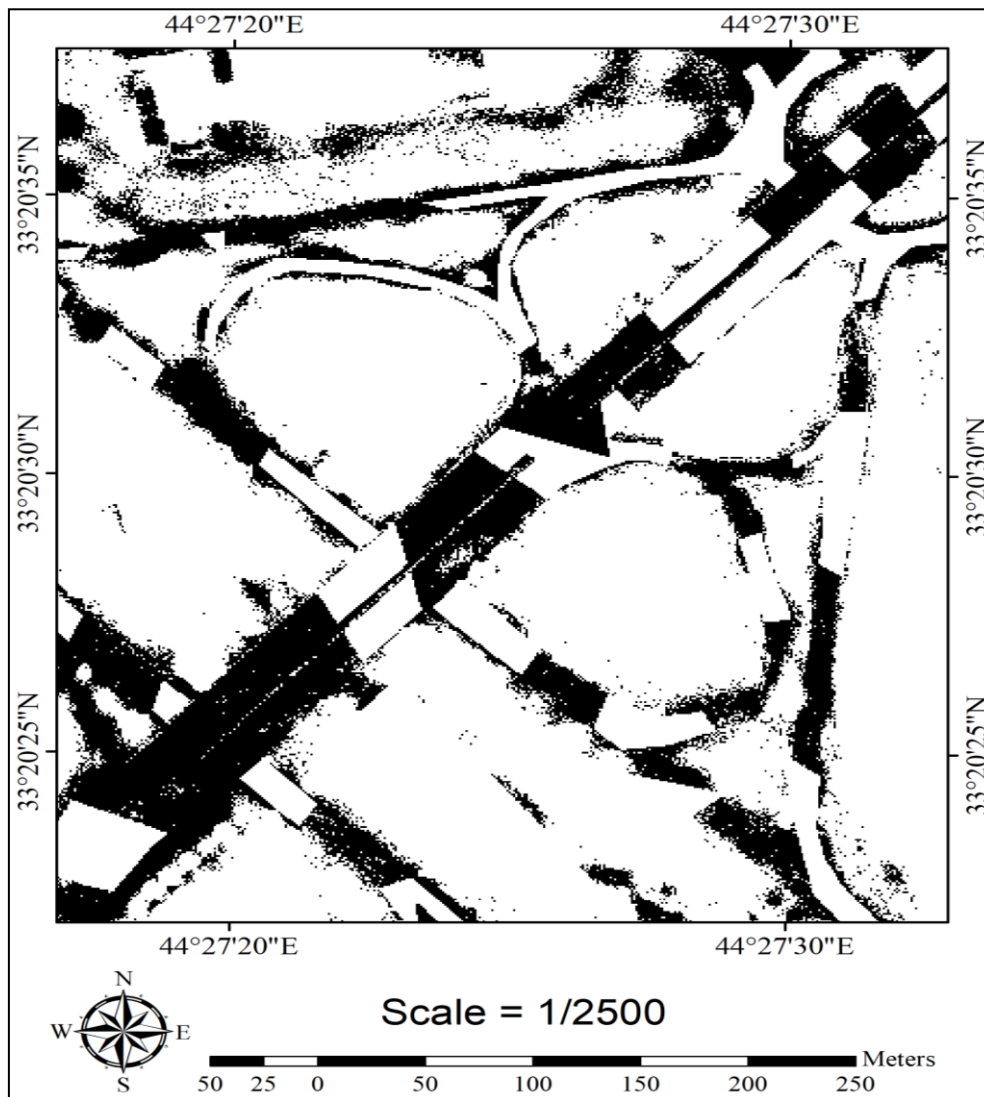

figure 5. road map by using the maximum likelihood method

Finally, a classification map was produced and based on the proposed MLP model compared with the traditional method maximum Likelihood which is available in the GIS to show its performance. The thematic maps, which produced by the maximum Likelihood and MLP models, are shown in Figure 5 and 6. The MLP model was selected over other statistical and machine learning methods because this model has the ability to learn complex features by itself. The results suggest that the MLP model outperformed the other models which are RF, OBY and maximum Likelihood considering all the used accuracy metrics. The overall realization and performanceof of the MLP model was about 98% due to the MLP model also showed a better result when RMSE and MAE with K metrics considered. Overall, this study suggests that the MLP model with simple sensitivity analysis can significantly outperform the RF, OBY and maximum Likelihood models. However, the latter model can be improved by optimizing its hyperparameters and architecture. This may be worked out in the future studies, as it requires lots of data, computing power, and extensive experiments.
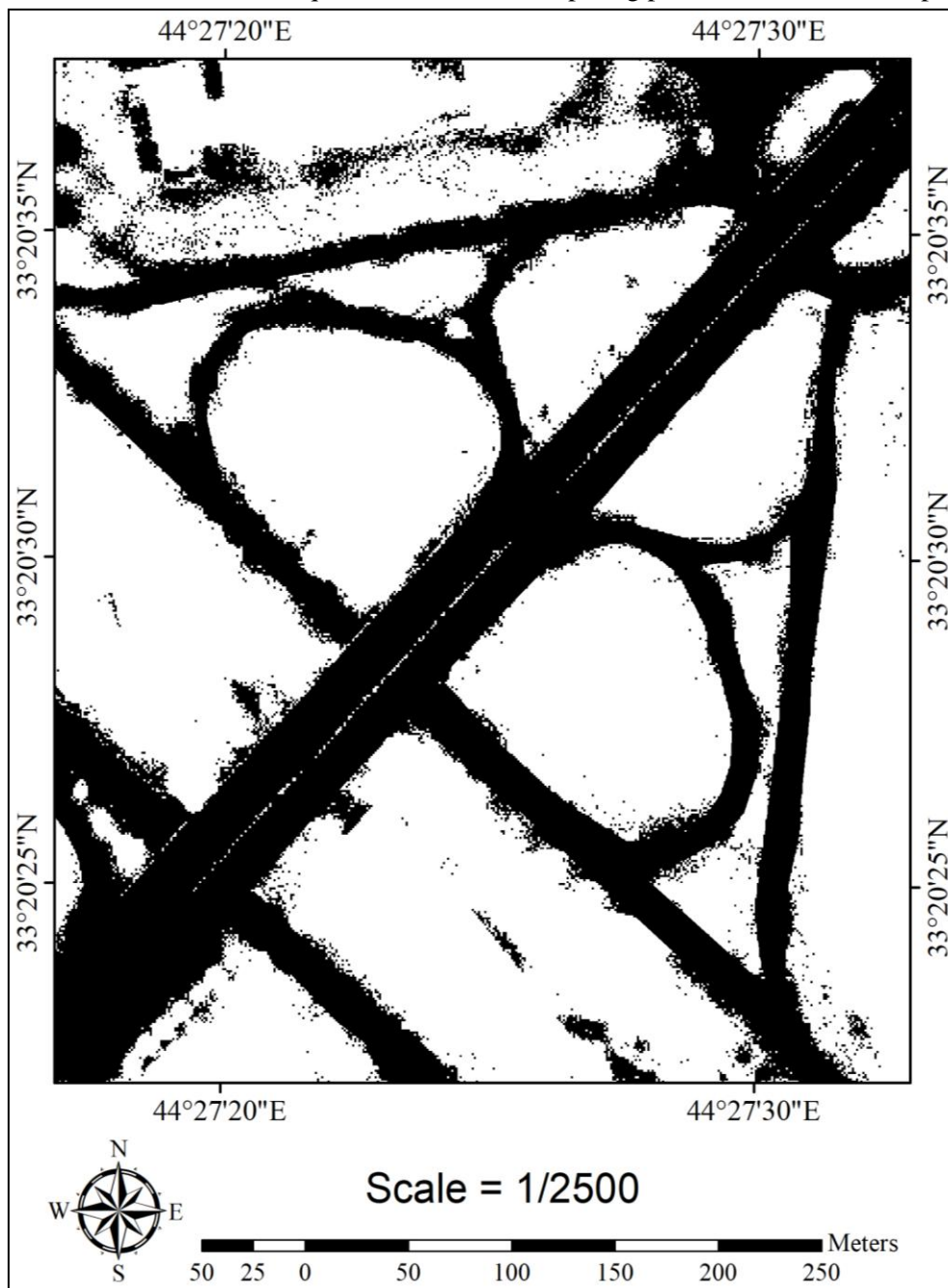

Figure 6. Road map by using the MLP model

4. Discussion

In this section a description and discussion will done for the proposal model which is MLP model through the sensitivity analysis, optimization algorithm analysis and activation function analysis for hidden and output layers.

The hidden layers were determined In this paper that affected on the performance of network analysis depth. And also, was checked for the   several network depths ranging from 3 to 20 hidden layers. The MLP evaluation model was based on (OA), (K) accuracies that were measured on training model. The figure 7 was presented the summary of the results. The map of the classification map was depended on the testing ,and also training samples. The experiment number 4 with 7 hidden layers were achieved the best accuracy. On the training model the overall accuracy and kappa accuracy were 97.7 and 0.95 respectively. The results were indicated there are no specific patterns or number of hidden layers to improve MLP model.
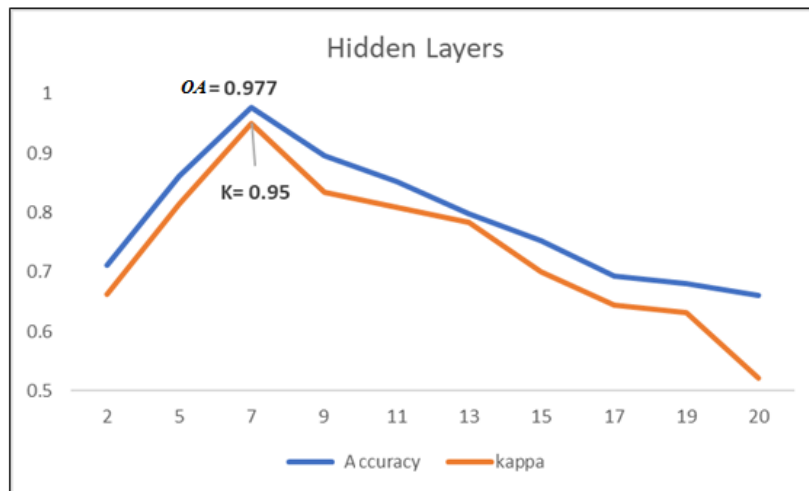

Figure 7. Hidden layers with accuracy and kappa.

The optimizations of MLP is very important because lead to improve the performance the network model. There are many algorithms and several ways help to develop this network such as LBFGS, SGD and ADAM. The figure 8 shows the best accuracy and the suitable algorithm for optimizing MLP classifier to classify the image. The experiment number 3 with solver LBFGS was the suitable algorithm for thematic map, as well as the same classifier achieved the Acc and K were 97.7 and 0.95 respectively.
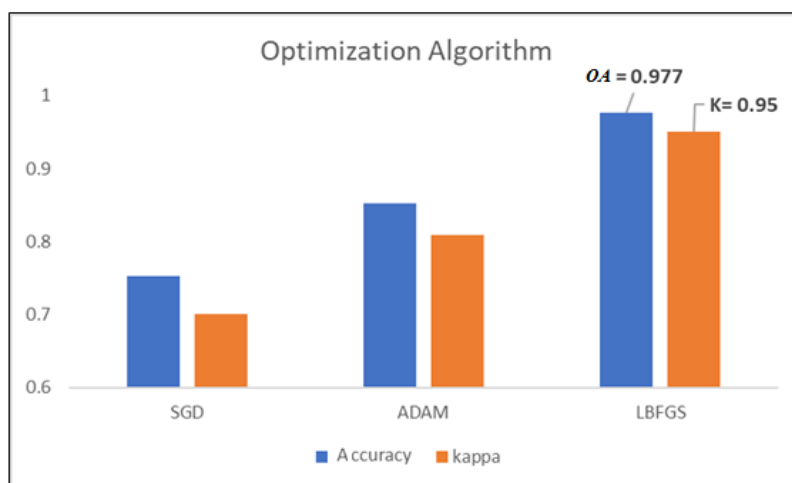

Figure 8. Optimization algorithm analysis with accuracy and kappa

Finally, the activation function was considered important for both the hidden and output layers because this function influence on the overall performance of ANN modelling. The figure 9 shows the activation function analysis for hidden layers and output layers respectively. This model analyzed several activation functions such as Logistic, relu, identity and tanh. So, the analysis of MLP shows the best accuracy was achieved with relu for hidden layers and output layers, while the worst result of MLP model was logistic activation function used.
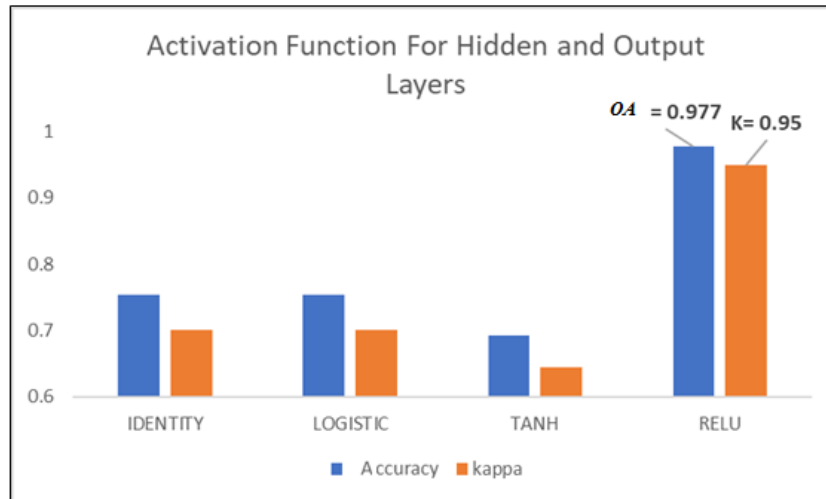

Figure 9. Activation function analysis for hidden and output layers

5. Conclusion

Machines learning with Geographical Information Systems (GIS) approach will save and protect time and labor to a good degree in updating Geodatabase of the road. In this study was used three different models of ML such as MLP, RF and BO. The evaluation of those three models based on statistical indictors such as (OA), (K), (MAE) and (RMSE) are used to get the best model to use as the proposed model to produce the thematic map in GIS. After preparing the data in the above steps and will be ready to as input to these three models, we will be dived the data two sets are training 70% and 30% testing of data. The results showed the MLP is the best model and overcome of other models through it achieved the highest accuracy (97.7) and kappa (0.95) with the lest for mean absolute error (0.04) and rout mean square error (0.13) for training model. While, the testing of MLP model was achieved 96.24, 0.93, 0.06 and 0.16 for OA with K and lest MAE and RMSE receptively. On the other hand, the result of the proposed model showed, seven of hidden layers was achieved the best accuracy, while the LBFGS of optimization algorithm shows the accuracy of the accepted model. Finally, the best activation function of the MLP model was the relu algorithm for hidden layers and output layers.

Reference

[1]   Y. Oubrahim, S. Lbazri, S. Ounacer, A. Rachik, R. Moulouki, and M. Azzouazi, "A New Architecture for monitoring Land use and land cover change based on remote sensing and GIS: A Data mining Approach," Periodicals of Engineering and Natural Sciences, vol. 6, no. 2, pp. 406–414, 2018.

[2]   H. Tokgöz and H. Bayraktar, "GIS risk analysis of a first-degree seismic zone: Kaynaşlı (Turkey) Keyword: Geographic information system Settlement suitability Disaster Street screening Corresponding Author," Periodicals of Engineering and Natural Sciences Journal, vol. 6, no. 2, pp. 10–23, 2018.

[3]   J. C.Thill, "Geographic information systems for transportation in perspective," *Elsevier* Transportation Research Part C: Emerging Technologies, vol. 8 pp.1-6, 2000

[4]   W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, P. Eklund "A review of road extraction from remote sensing images," *Elsevier* , Journal of Traffic and Transportation Engineering (English Edition) , vol 3,issue 3,pp(271-282, 2016

[5]   M. Adel Al-Shaher, "Detection and classification of vehicle types using machine learning technology," Periodicals of Engineering and Natural Sciences Journal ,vol. 7, no. 3, pp. 985–815, 2019.

[6]   D. Geman and B. Jedynak, "An Active Testing Model for Tracking Roads in Satellite Images." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.18 issue 1, pp.(1-14), (1996).

[7]   Z. Zhong, W. Cui, H. Jiang, "Fully convolutional networks for building and road extraction: Preliminary results". IEEE International Geoscience and Remote Sensing Symposium (IGARSS) pp.1591-1594, 2016.

[8]   O. Jasim, K. Hassoon, N.I. Sadiqe, "Mapping LCLU Using Python Scripting," *engtechjournal.org*. vol.37 ,(A) ,pp.140-147, 2019

[9]   C. Leistner, A. Saffari, and H. Bischof, "MIForests: Multiple-instance learning with randomized trees," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* , vol. 6316 LNCS, no. PART 6, pp. 29–42, 2010.

[10]   B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1. Institute of Electrical and Electronics Engineers Inc., pp. 148–175, 01-Jan-2016.

[11]   F. Gasparetti, C. De Medio, … C. L.-T. and, and undefined, "Prerequisites between learning objects: Automatic extraction based on a machine learning approach," *Elsevier*, Telematics and Informatics, 35(3), 595–610. 2018.

[12]   A. A. Ahmed and B. Pradhan, "Vehicular traffic noise prediction and propagation modelling using neural networks and geospatial information system," *Environ. Monit. Assess.*, vol. 191, no. 3, Mar. 2019.

[13]   A. A. Ahmed, B. Kalantar, B. Pradhan, S. Mansor, and M. I. Sameen, "Land use and land cover mapping using rule-based classification in Karbala City, Iraq," *Lect. Notes Civ. Eng.*, vol. 9, pp. 1019–1027, 2019.