Transport and Communications Science Journal

A MACHINE LEARNING APPROACH TO RISK ASSESSMENT OF EXPRESSWAY BRIDGES

Le Duc Anh[1], Dao Duy Lam[2*], Thai Thi Kim Chi[2], Bach Thi Diep Phuong[3]

[1]Ministry of Transport, No 80 Tran Hung Dao Street, Hanoi, Vietnam

[2]University of Transport and Communications, No 3 Cau Giay Street, Hanoi, Vietnam

[3]University of Transport Technology, No 54 Trieu Khuc Street, Hanoi, Vietnam

ARTICLE INFO

TYPE: Research Article
Received: 21/03/2022
Revised: 29/05/2022
Accepted: 31/05/2022
Published online: 15/09/2022
https://doi.org/10.47869/tcsj.73.7.1
Corresponding author
Email: daoduylam@utc.edu.vn; Tel: +84 912532728

Abstract. The expressway network in Vietnam is developing strongly, playing the role of the backbone of the national road system, in which bridge construction accounts for a large proportion. With many specific characteristics and complex risks always hidden in all stages of the expressway project, risk assessment to have solutions and plans to prevent and respond to risks, limiting the impacts of quality assurance and operational safety of the works is essential. However, the current risk assessment and forecasting models still have many limitations. The application of Machine Learning to all aspects of life is getting more popular. This article develops the algorithms, models and program to assess the technical risks in the period of construction and service of expressway bridges in Vietnam using Machine Learning, in order to solve the current limitations in this work. The selection of key influencing factors is especially important in the field of risk assessment. It improves the classification model's performance by focusing only on the most important factors in the data. Via the applications of artificial neural networks and the Random Forest Algorithm in data processing, the performance risks for bridge management can be analyzed, and performed in more detail and exactly. The possible multiple and non-linear relationships of the risks can be investigated. Based on the results, the proposed model helps the managers to make optimum decisions on managing the risks in advance and to obtain sustainable solutions.

Keywords: risk assessment, expressway bridge, machine learning, ANN, Random Forest.

1. INTRODUCTION

According to the Road Network Development Plan for the 2021-2030 period, vision to 2050, approved by the Prime Minister in September 2021, one of the main goals is that by 2030, Vietnam will have about 5,004 km expressway (increasing about 3,841 km in compared to 2021) [1].

The expressway project with the special characteristics of high-speed operation, the requirements for quality and reliability are very high, the components are complicated, the investment cost is high and the project duration is long, requires the cooperation of many stakeholders, requires complex high-tech technology; At the same time, the influence and domination of the socio-economic - legal - cultural environment has led to the possibility of the occurrence of risks in these projects which is unavoidable.

Among expressway infrastructures, bridges always play an important role, the excessive damage of a bridge can cause harmful consequences for the management in general. Bridge risk refers to any event or hazard caused by uncertainty that could hinder the achievement of the business goals, the delivery of the stakeholders' expectations, and the occurrence of the consequences of an event. According to different aspects, risks in bridge construction projects can be divided into different groups such as financial risk, external risk, design risk, management risk, construction risk, contractual risk, health and safety risks or materials and equipment, design, human resources, finance, management, nature and environmental conditions. In this paper we pay more attention to the technical ones.

In order to reduce the impact of risks to expressway bridges, risk management is required to be conducted. Risk management is an essential and integral part of project management on virtually all construction projects. A risk management plan is considered as an attempt of analyzing risks that can enrich the decision-making process and provide additional arguments to help choosing the optimal variant using multiple approaches [2, 3].

Risk management is defined as a systematic controlling procedure for predicted risks that are to be faced in an investment or a project (Dikemen et al., 2004). Flanagan and Norman have defined risk management as a stepwise procedure consisting of risk identification, risk classification, risk analysis, and risk response tasks [3, 4].

So, the basic procedure in expressway bridge project risk management should be:

• Risk Management Planning: deciding how to approach and plan the risk management activities for the project;

• Risk Identification: The objectives of risk identification are to identify and categorize risks that could affect the project and document these risks.

• Risk Assessment: The process of quantifying the risk events documented in the preceding identification stage. Risk assessment has two aspects. The first determines the likelihood of a risk occurring (risk frequency); risks are classified along a continuum from very unlikely to very probable. The second judges the impact of the risk should it occur (consequence severity).

• Risk Analysis: One of the core components of risk management that enables professionals to quantify and analyse risks that may pose potential threats to project performance in terms of cost, quality, safety, and time.

• Risk Response Planning: taking steps to enhance opportunities and reduce threats to meeting project objectives.

• Risk Monitoring and Control: monitoring identified and residual risks, identifying new risks, carrying out risk response plans, and evaluating the effectiveness of risk strategies throughout the life of the project.

Most of the models for multi-hazard risk assessment of the work are based on traditional probabilistic approaches such as based on questionnaires. Recently, some authors have introduced new approaches, more effective tools for risk analysis and assessment than the traditional probabilistic approach (S. Kameshwar et al., 2014), (Jelena M. Andric et al., 2015) [3, 6], (F. Soleimani et al., 2017), (E. Borgonovo et al., 2017) [3, 5], (Aaron, 2018) [5], (E. Zio et al., 2018) [7], (Z. Chen et al., 2018) [8], and some other authors [3, 5, 9].

Researchers on risk management in transport construction projects in Vietnam such as Trinh Thuy Anh, Nguyen Quoc Tuan et al, 2006 have applied Monte- Carlo for risk analysis during construction phase [3]. Nguyen Viet Trung studied risk management of bridge [10], Than Thanh Son, 2015, studied risks in the form of public-private partnership to develop road transport infrastructure projects in Vietnam has summarized 51 specific risks [3]. Nguyen and Bui, 2016 researched and managed technical risks of road construction in Vietnam [11], Dao and al. research on identification and management of technical risks of highway bridges [3, 9, 10, 12].

There are currently no universally accepted frameworks for risk identification in expressway construction. However, according to World Bank (2015), risk identification in expressway construction should begin with stakeholder mapping. The project managers should then identify the specific expectations and elements of satisfaction for each stakeholder.

Fuzzy Logic Hierarchical Modeling (FAHP) is a recent risk assessment method that has a consistent approach to risk assessment that helps to address the limitations of existing models by suggesting Based on new fuzzy logic, then propose a bridge risk assessment model as the basis for developing decision-making tools to manage global safety more effectively but still has certain limitations [3, 5, 7, 11].

The application of Machine Learning to all aspects of life is now getting more popular. In fields such as finance or services, an interpretable model is essential in addition to making predictions or classifying classes. The selection of key influencing factors is especially important in the field of risk assessment. It improves the classification model's performance by focusing only on the most important factors in the data [5, 13, 14, 15, 16].

Feature extraction is divided into two stages: developing a set of features and selecting features [15, 17]. Developing a set of features is a critical task in data processing. When developing data, we must ensure that we do not lose too much information and that we do not make it too expensive in terms of costs. The second stage aims to find the object's representative features while removing redundant and confounding features to improve the performance of data mining algorithms.

Dimensionality reduction of the sample, also known as data set compression, is accomplished through feature extraction and feature selection [15]. This is the most fundamental step in data preprocessing. Feature selection can be an inherent part of feature extraction, such as elementary component analysis, or it can be part of an algorithmic processing design, such as decision tree design. Feature selection, on the other hand, is frequently a separate, isolated step in a processing chain.

In this article, we'll propose the Random Forest Algorithm in data processing for feature selection and develop the artificial neural networks (ANN) so the performance risks for bridge management can be analyzed and performed in more detail and exactly. The possible multiple and non-linear relationships of the risks can be investigated. Based on the results, the proposed model and program can help the managers to make optimum decisions on managing the risks in advance and to obtain sustainable solutions.

2. ARTIFICIAL NEURAL NETWORK BASE

As mentioned above, the risk assessment of the expressway bridge is a complicated evaluation system. Using Artificial Neural Networks (ANN) the evaluation system can comprehensively and effectively assess the risk probability of the performance such as bridge risk score and risk categories. ANN can learn from past examples, generalizing the information for future solutions, and self-updating. The subjectivity of ascertaining the weight of risky factors and the burdensome task of computing can be avoided effectively.

Artificial Neural Networks are flexible mathematical structures that are capable of identifying complex non-linear relationships between input and output data sets. A neural net consists of a large number of simple processing elements that are variously called neurons, units, cells, or nodes. Each neuron is connected to other neurons by means of direct communication links, each with an associated weight that represents information being used by the net to solve a problem. The net usually has two or more layers of processing units where each processing unit in each layer is connected to all processing units in the adjacent layers.

Multilayer perceptrons (MLPs) are the most common type of feed-forward networks with one or more hidden layers. The most common learning rule for multi-layer perceptrons is the back propagation algorithm [13].

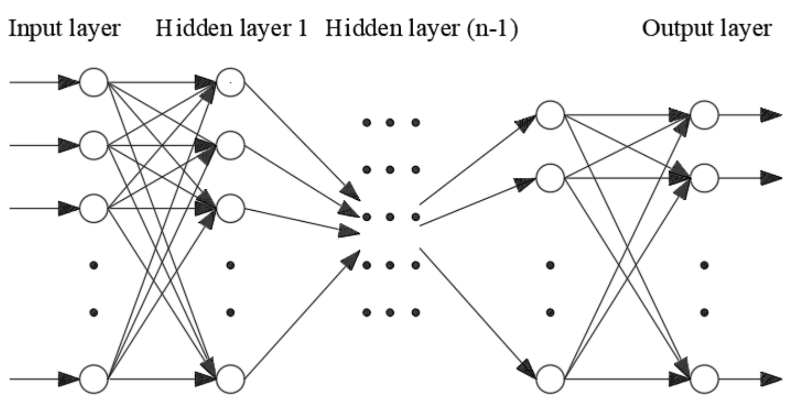

Figure 1. General Multi-Layer Perceptron (MLP) structure [13].

In the MLP structure, the neurons are grouped into layers. The first and last layers are called input and output layers respectively, because they represent inputs and outputs of the overall network. The remaining layers are called hidden layers. Typically, an MLP neural network consists of an input layer, one or more hidden layers, and an output layer, as shown in Fig.1 [13, 18].

Depend on the specific projects, input and output variables are carefully selected from the database, which includes information about: structure key code, structure name, risk event description, risk score, risk categories, outline works cost, maintenance classification, etc. Optimization of the neuron number in the hidden layer is performed using the trial-and-error method.

In order to assess the accuracy and performance of the neural network models, different criteria are utilized such as sum of squared error (SSE), mean-square error (MSE), mean absolute error (MAE), rootmean-square error (RMSE), mean absolute percentage error (MAPE) (Elhag and Boussabaine, 2002), root-mean-square percentage error(RMSPE) (Law 2000; Martin and Witt 1989), correlation coefficient (R), and Theil's inequality coefficient (U) (Martin and Witt 1989; Theil 1966) [5]. The MSE, and R are the most widely used criteria that are defined as follows:

$$R=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} \tag{1}$$

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2 \tag{2}$$

In which: X_i, Y_i actual value and calculated value; $\bar{X}$, $\bar{Y}$: the average value.

Along with the advantages of neural networks, there will be disadvantages, the most common of which are as follows: High hardware configuration is required, training a neural network necessitates the use of expensive hardware such as a GPU or a TPU; Inability to interpret network output, it's like a black box that, while training on the data, has taught itself inference methods that we have no idea why it does; Inability to debug, in neural network programming and training, these problems can be extremely difficult. As a result, finding a perfect architecture for any problem is impossible. And selecting a good neural network architecture is heavily influenced by the network builders' experience.

In the world, ANN is applied quite broadly in risk management. Kim used the neural network for identifying risks. Wenxi used a back-propagation neural network for assessing risks in highway projects in China. In 2009, Wenxi and Danyang developed an approach that used a combination of fuzzy logic and neural network to evaluate risk of highway projects [5, 19]. Elhag and Wang used ANN to model bridge risk score and risk categories [5]. Pedro used ANN to assess project risk through the prediction of the contractor's profit considering risk factors [5, 17]. A representative neural network with four inputs including safety, functionality, sustainability and environment risk is shown in Fig. 1 [3, 9]. However, these researches did not state clearly the values of risk factors and the method to achieve them as well as a lot of difficulty for risk prediction.

3. MACHINE LEARNING FOR RISK ASSESSMENT AND FORECAST

3.1. Model introduction

Authors have built a machine learning model for expressway bridge risk assessment and forecast after analyzing and collecting a data set of highway bridges [3, 20]. The first thing that must be clarified is to determine which features among 2313 data samples corresponding to 30 features, will have the greatest influence on the risk level of highway bridges. The selection of important features will assist the model in making more accurate predictions and will limit the phenomenon of over-fitting [21]. The artificial neural network architecture to make a prediction model and estimate the risk after selecting the important set of features using the Random Forest algorithm is presented below. An overview of the processing is shown in Fig. 2 including following main steps: Collecting data sets of risks on expressway bridges; Building a feature extractor using Random Forest; Building a risk prediction model using an artificial neural network; Risk prediction of real bridges for verification.

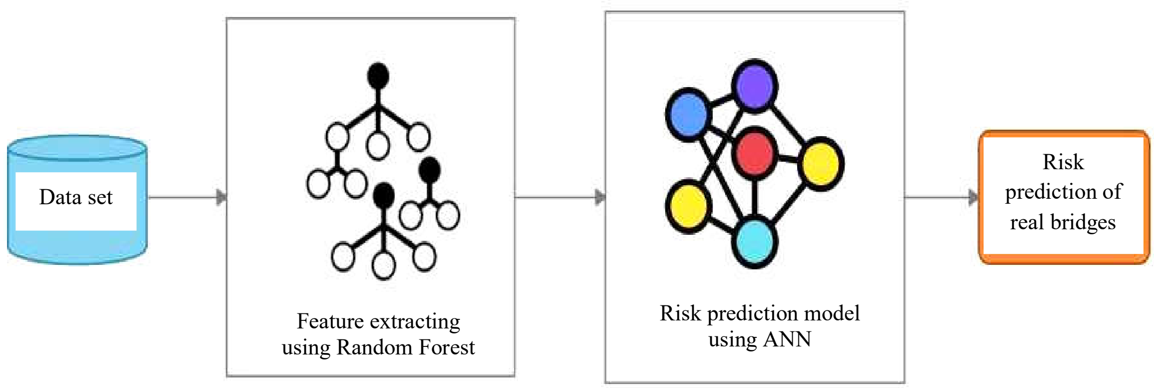

Figure 2. Processing steps of machine learning approach.

Dimensionality reduction of the sample, also known as data set compression, is accomplished through feature extraction and feature selection [16]. This is the most fundamental step in data preprocessing. Feature selection can be an inherent part of feature extraction, such as elementary component analysis, or it can be part of an algorithmic processing design, such as decision tree design. Feature selection, on the other hand, is frequently a separate, isolated step in a processing chain.

Feature selection can be defined as the process of finding a subset of features from M feature sets of the original N data set, thereby defining the feature selection criteria. A quantitative criterion is used to minimize the size of the feature space in this manner. As the size of a field increases, so does the number of elements in the N set, making finding the best representative set difficult, and there are numerous problems associated with the chosen set. An extraction algorithm, in general, consists of four basic steps: generation of subsets, evaluation of subsets, stopping conditions, and validation of results [15].

3.2. Data collection

In the research, we use the dataset taken from the bridge management system under the Directorate for Roads of Vietnam (http://www.vbms.vn) [20] and from site survey [3]. This data includes a lot of information such as bridge structure type, girder structure type, span structure type, year of operation, average traffic volume ... and especially risk index to assess damage level of a bridge at the survey time. There are 59 attributes collected for each bridge. This information is very valuable for us to analyze and build risk assessment models for different types of bridges. We collected a total of 11823 data samples of 11823 bridges in the country which include many different types of bridges. The quantity of each bridge corresponding to each management unit is presented in Table 1.

Table 1. Number of bridges on the VBMS system by management unit.

Management unit	Quantity
Road Administration Department I	901
Road Administration Department II	988
Road Administration Department III	890
Road Administration Department IV	1393
Expressway Administration Department	643
Departments of Transport of provinces and cities	7008
Total	11823

3.3. Enrichment of expressway bridge data

As shown in Table 1, the number of bridges on the highways is managed by the Expressway Administration Department. The total number of bridges on the highways on this dataset is 643 units [3]. This is a rather modest number compared to a dataset that can be analyzed and trained by machine learning methods. For machine learning problems in general and application of algorithms on neural networks in particular, it always requires a large enough dataset, enough quality and enough diversity [16]. Therefore, data enrichment not only helps us to obtain a dataset with a larger number of samples, but also enhances the diversity of the data, making machine learning models more generalizable [21]. The main goal behind data enrichment methods is to add additional data samples to the original dataset that have similar distributions to the old data samples. For each data type, there are different data enrichment methods. For example, with image data we can use geometric transformations such as cropping, zooming in, zooming out, rotating the image, adding noise to the image... [22]. With text data, we can use techniques such as replacing synonyms, replacing synonymic phrases, replacing synonymic sentences [14, 23] to create new data samples with similar characteristics to the original data. We propose a data enrichment method using the KNN algorithm [17] to determine the K bridges that have the most similarity with the 643 highways mentioned above. In this algorithm implementation, we use hyperparameter K = 5 and similarity threshold = 0.6. To calculate the similarity, we use the cosine distance (3) to calculate the similarity between two feature vectors of the bridges.

$$\cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{3}$$

After calculating and removing duplicate bridges, we obtain a dataset consisting of 2313 bridges. This dataset is used by us in the following analyses.

3.4. Missing value processing and dataset division

With this dataset of 2313 data, we proceed with basic data processing such as missing value processing. For each type of adequacy, we use different processing.

After conducting the data preprocessing steps as described above, we proceed to divide the dataset into training dataset and test dataset. The test ratio used is 0.2, which means that out of a total of 2313 data samples, 1850 samples are used for model training and 463 samples are used for model testing. This division is shown in Fig. 3.

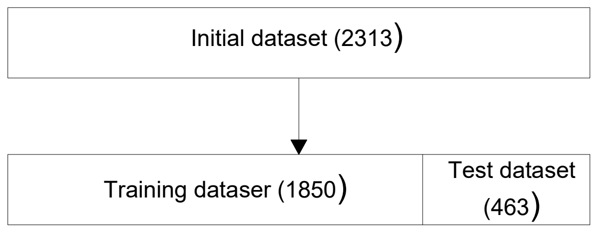

Figure 3. Training dataset and test dataset division procedure.

3.5. Machine learning experimental setup

Experiments was conducted according to the following two scenarios:

Using ANN without feature extraction: To test the effectiveness of selecting important attributes by Machine learning approach, a test run of ANN on the dataset of 15 attributes at first without feature extraction step was conducted.

Using ANN with important attributes: This is the combination of ANN with important feature extraction by Random Forest algorithm. The ANN algorithm will take as input what are the 6 most important features (out of 15 features) to calculate the risk level on a scale of 1 - 5, corresponding to very low risk, low risk, medium risk, high risk, very high-risk levels.

To build an important set of attributes with the Random Forest algorithm, there are several hyperparameters that need to be selected such as *n_estimator, criterion, max_depth, max_samples, max_features*... The choice of values of these hyperparameters will affect the efficiency of the Random Forest algorithm. However, finding some optimal parameters is a difficult and time-consuming task. It usually requires the knowledge and experience of the modeler. For each scenario, the neural network using Levenberg–Marquardt algorithm was launched [8, 13] and its effectiveness was measured with mean squared error (MSE).

4. RESULTS AND DISCUSSION

For each experimental scenario as described above, we measure the parameters with the accuracy of the model according to the MSE index. These parameters are measured per epoch of the training process with both the training dataset and the test dataset. The best evaluation results are obtained at the time when the model has the MSE value on the smallest test set. Here are the results for each case.

4.1. Using ANN without feature extraction

For the case only using the ANN on all 15 attributes without feature extraction by Random Forest, it is easy to see that the model falls into the overfitting phenomenon [21, 23, 24] and being overfitted into the training dataset without the ability to generalize to the test dataset. This problem is clearly shown in Fig. 4. Obviously, this method is not effective because the neural network architecture only learns well on the training dataset without generalization on the test set. The best MSE loss value of the model on the training set is 0.0024 and on the test set is 0.034.

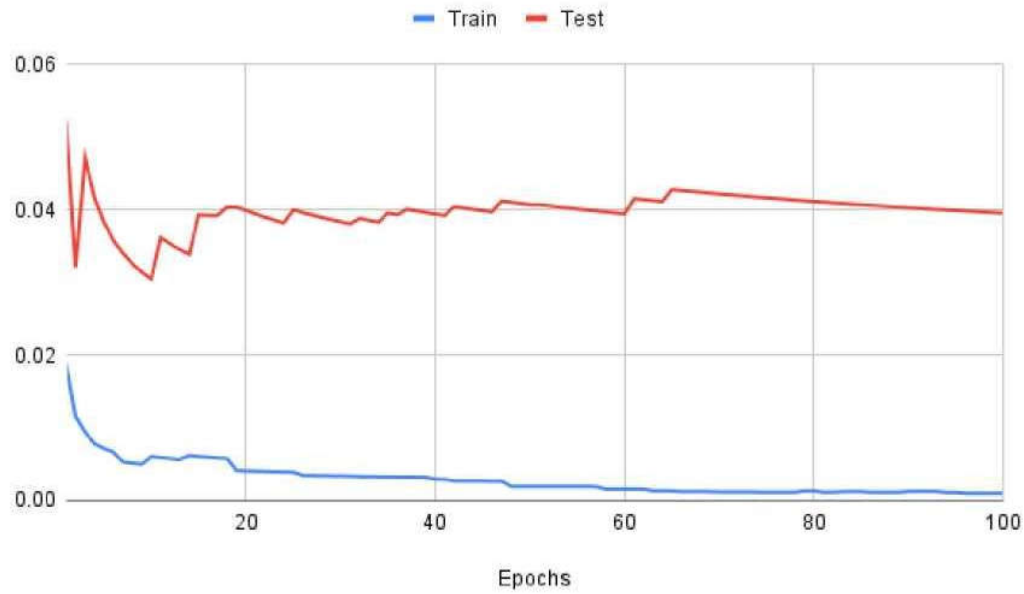

Figure 4. MSE index for the case of using ANN without feature extraction.

4.2. Using ANN with important feature extraction by Random Forest algorithm

For the case where the Random Forest algorithm is used for feature extraction, the ANN works more stably, as the loss on the training set and the test set both decreases quite smoothly as shown in the Fig. 5.

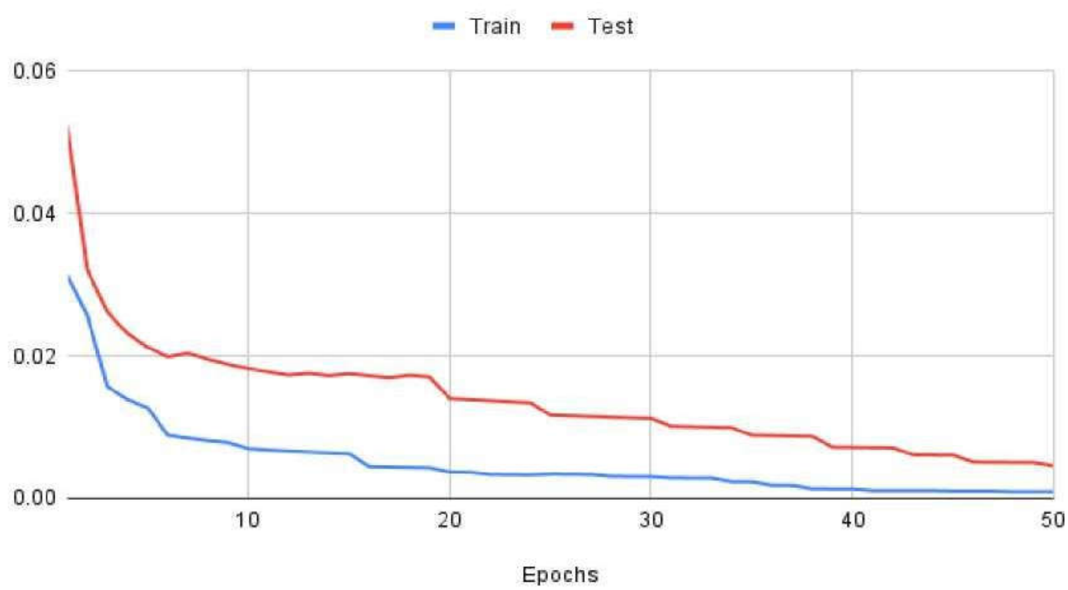

Figure 5. MSE index for the case of using ANN training with Random Forest.

This shows that the ANN model has the ability to generalize to the data on the test set. This proves that feature extraction has helped the model remove information that is not related to the level of risk. This makes the model simpler and avoids being fitted with outliers in the dataset. The best MSE loss value of the model on the training set is 0.0008 and on the test set is 0.0019, which is much better than the first method as in Table 2.

Table 2. Result comparisons.

Experiment	Error on training set	Error on test set
ANN without feature extraction	0.0024	0.034
ANN with important feature extraction	**0.0008**	**0.0019**

4.3. Risk prediction on real bridges

We have selected 12 real data samples to present in this section. These data are samples drawn from the test dataset, including actual expressway bridges collected from the VBMS dataset [20] and site survey [3, 25]. Details of these expressway bridges are described in Table 3. In which we have representative and the most influencing factors such as Service Year - Number of service years of the bridge, Average speed - Average service speed, Average Load - Average Service Load of Vehicle, Average Traffic Volume - Average Service Traffic Volume, Score - Risk Score assessed through the survey. Predicted Score - the risk score assessed by the proposed model.

Table 3. Risk prediction of bridges.

No	Service Years	Average speed (Km/h)	Average Traffic Volume	Score	Predicted
1	15	75	100,000	22	23.4
2	15	120	100,000	34	35.1
3	15	50	2,000	92	91.2
4	10	65	100,000	30	31.2
5	10	50	100,000	46	45.3
6	10	50	2,000	90	89.3
7	10	55	1,000	95	95.3
8	10	50	1,000	97	97.6
9	5	50	100,000	36	35.6
10	5	75	25,000	65	66.4
11	5	55	1,000	95	96.3
12	5	50	1,000	99	98.2

A lower risk score represents a higher level of risk, calculated on a 100-point scale. It means that if the score is small the risk is higher. In Table 3, we can find that the old bridge with big traffic volume has more risk with the low score, but a new bridge with high traffic volume also has a low score. In this case, it shows the important impact of traffic volume in risk assessment. The speed will be more important in service safety assessment. Risk prediction by machine learning is quite consistent with the Risk Score survey. For application, we have started to construct a risk assessment and forecast program for expressway bridge at the link: https://bit.ly/BridgeRiskAssDA.

5. CONCLUSIONS

With the rapid development of the expressway systems in which the bridge works play a significant role, potential risk factors that can be synthesized from similar researches in Vietnam and abroad combined with surveys to assess actual conditions to ensure simplicity and effectiveness to help prevent possible serious incidents, reduce risks, improve construction quality, and effectively control progress and cost of highway projects are of great interest. However, the research results for risk classification assessment, scenario analysis and response solutions, as well as providing procedures and regulations on risk management for expressway bridges in Vietnam is currently incomplete, the risk prediction model should be developed soon as a tool for all stage of the project.

In this paper, we have developed the algorithms, models and program using Machine Learning approach to assess the technical risks in the period of construction and service of expressway bridges. The applied results for real projects in Vietnam are presented to verify and show the advantages and accurate results which can help to overcome current limitations in risk assessment and prediction.

Through the study of actual data, it shows that there exists an influence between the design, construction and operation conditions on the occurrence of the incident [3, 9]. Therefore, the research to build predictive models will be based on statistics that have happened in the past. On the basis of the analysis of the factors that most affect the construction incidents, the research has been conducted to select the variables on the types of bridge works and traffic conditions to serve as a basis for preparation of input data to build the model. In general, the selected variables are suitable and sufficient to build predictive models according to Machine Learning.

The accuracy of the forecasting model depends heavily on the quality of the data, which is also a limitation in predictive analysis. In order for the model to have higher accuracy, it is necessary to continue researching and perfecting, especially adding input data to retrain the ANN network to be more accurate. The application of Machine Learning in current research is very useful, especially in the Fourth Industrial Revolution (4.0), it needs to be fully promoted. It is recommended to further promote the research and application of Machine Learning for many other activities in the field of transport.

ACKNOWLEDGMENT

This research is funded by University of Transport and Communications (UTC) under grant number T2021-CT-025

REFERENCES

[1]. Prime Minister, Road Network Development Plan for the 2021-2030 period, with a vision to 2050, approved by Decision No 1454/QĐ-TTg, 2021. (in Vietnamese)

[2]. Alberto Deco, Dan M. Frangopol, Risk assessment of highway bridges under multiple hazards, Journal of Risk Research, 14 (2011) 1057-1089. https://doi.org/10.1080/13669877.2011.571789

[3]. D. D. Lam, L. D. Anh, N. T. Dong, Report of scientific and technological research project of Ministry of Transport "Research on identification and management of technical risks of expressway bridges", 2019. (in Vietnamese)

[4]. A. Clark-Ginsberg, L. Abolhassani, E. Azam Rahmati, Comparing networked and linear risk assessments: From theory to evidence, International Journal of Disaster Risk Reduction, 30B (2018) 216-224. https://doi.org/10.1016/j.ijdrr.2018.04.031

[5]. L. D. Anh, D. D. Lam, N. T. Dong, Identification of technical risks for bridges on the highway in Vietnam, The Transport Journal, 12 (2019) 40-44. (in Vietnamese)

[6]. J. M. Andric, D. Lu, Risk assessment of bridges under multiple hazards in operation period, Safety Science, 83 (2016) 80-92. https://doi.org/10.1016/j.ssci.2015.11.001

[7]. E. Zio, The future of risk assessment, Reliability Engineering & System Safety, 177 (2018) 176-190. https://doi.org/10.1016/j.ress.2018.04.020

[8]. Z. Chen, Z. Li, C. Huang, G. Zhang, H. Yu, Safety Assessment Method of Bridge Crane Based on Cluster Analysis and Neural Network, Procedia Computer Science, 131 (2018) 477-484. http://dx.doi.org/10.1016/j.procs.2018.04.235

[9]. D. D. Lam, L. D. Anh, N. T. Dong, Risk management of expressway bridge projects in Vietnam: current status and future researches, Proceedings of the Tenth International Conference on Bridge Maintenance, Safety and Management (IABMAS 2020), 1st edition (2021) 1962-1965. https://doi.org/10.1201/9780429279119

[10]. N. V. Trung, D.C. Tam, Risk analysis and management in bridge construction, Civil Construction Public house, 2011. (in Vietnamese)

[11]. N. V. Chau, Researching and managing technical risks in road construction in Vietnam, PhD thesis, University of Transport and Communications, 2016. (in Vietnamese)

[12]. D. D. Lam, N. V. Trung, N. Q. Huy, Solution pour améliorer la maintenance des ponts. Academic Journal of Civil Engineering, 36 (2019) 370-373. https://doi.org/10.26168/ajce.36.1.89

[13]. M. W. Gardner, S. R. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmospheric Environment, 32 (1998) 2627-2636. https://doi.org/10.1016/S1352-2310(97)00447-0

[14]. H. Jia, J. Lin, J. Liu, An earthquake fatalities assessment method based on feature importance with deep learning and random forest models, Sustainability, 11 (2019) 2727-2748. https://doi.org/10.3390/su11102727

[15]. K. Kira, L. A. Rendell, A practical approach to feature selection. Machine learning proceedings, 1992 (1992) 249-256. https://doi.org/10.1016/B978-1-55860-247-2.50037-1

[16]. T. Shaikhina, D. Lowe, D. Daga, D. Briggs, R. Higgins, N. Khovanova, Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation, Biomedical Signal Processing and Control, 52 (2019) 456-462. https://doi.org/10.1016/j.bspc.2017.01.012

[17]. C. V. Gonzalez Zelaya, Towards explaining the effects of data preprocessing on machine learning, 2019 IEEE 35th international conference on data engineering (ICDE), 2019, pp. 2086-2090. https://doi.org/10.1109/ICDE.2019.00245

[18]. UK ORR, Annual assessment of highways England's performance, 2018.

[19]. S. Mangalathu, G. Heo, J. S. Jeon, Artificial neural network based multi-dimensional fragility development of skewed concrete bridge classes, Engineering Structures, 162 (2018) 166-176. https://doi.org/10.1016/j.engstruct.2018.01.053

[20]. Directorate for Roads of Vietnam, Data from bridge management system, http://www.vbms.vn.

[21]. W. M. P. Van der Aalst, V. Rubin, H. M. W. Verbeek, B. F. Van Dongen, E. Kindler, C. W. Gunther, Process mining: a two-step approach to balance between underfitting and overfitting, Software & Systems Modeling, 9 (2010) 87-111. https://doi.org/10.1007/s10270-008-0106-z

[22]. W. Chen, A. Shirzadi, H. Shahabi, B. B. Ahmad, S. Zhang, H. Hong, N. Zhang, A novel hybrid artificial intelligence approach based on the rotation forest ensemble and native Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China, Geomatics, Natural Hazards and Risk, 8 (2017) 1955-1977. https://doi.org/10.1080/19475705.2017.1401560

[23]. W. Chen, Z. Sun, J. Han, Landslide susceptibility modeling using integrated ensemble weights of

evidence with logistic regression and random forest models, Applied Sciences, 9 (2019) 171-197. https://doi.org/10.3390/app9010171
[24].D. M. Hawkins, The problem of overfitting, Journal of Chemical Information and Computer Sciences, 44 (2004) 1-12. https://doi.org/10.1021/ci0342472
[25]. International Organization for Standardization, ISO 31000:2018, Risk management – Guidelines.