International Journal of Applied Power Engineering (IJAPE)
Vol. 12, No. 4, December 2023, pp. 349~358
ISSN: 2252-8792, DOI: 10.11591/ijape.v12.i4.pp349-358

An ensemble based data mining model for contingency analysis of power system under STLO

Ravi V. Angadi[1], J. Alamelu Mangai[2], V. Joshi Manohar[1], Suresh Babu Daram[3], Paritala Venkateswara Rao[2]
[1]Department of Electrical and Electronics Engineering, School of Engineering, Presidency University, Bengaluru, India
[2]Department of Computer Science and Engineering, School of Engineering, Presidency University, Bengaluru, India
[3]Department of Electrical and Electronics Engineering, School of Engineering, Mohan Babu University, Tirupati, India

Article Info

Article history:

Received Feb 1, 2023
Revised May 19, 2023
Accepted May 25, 2023

Keywords:

AdaBoost classifier
Bagging classifier
Contingency analysis
Data mining
Decision tree
Performance indices
Severity prediction

ABSTRACT

In a large, interconnected power system, contingency analysis is a useful tool for pinpointing the potential consequences of post-event scenarios on the system's safety. In this work, the Newton-Raphson technique is applied to every single outage of a transmission line to compute the load flows. For the static security classification of the power system, the line voltage stability performance index (LVSI) is used. There are three levels of static security of power system namely: non-critical (the least severe), semi-critically insecure (the next lowest severe), and critical (the next highest severe). The various data mining techniques such as decision trees, bagging-based ensemble methods, and boosting-based ensemble methods were applied to assess the severity of the line under various loading and contingency conditions. Test systems based on the IEEE 30 bus system were used with the proposed machine learning classifiers. The experimental results proved that bagging based ensemble method provided better accuracy compared to the decision tree and the AdaBoost ensemble method for predicting the power system security assessment. The bagging-based ensemble method has a predictive accuracy of 85% and an AUC of 0.94.

Corresponding Author:

Ravi V. Angadi
Department of Electrical and Electronics Engineering, School of Engineering, Presidency University
Bengaluru, Karnataka 560064, India
Email: raviangadi4045@gmail.com

1. INTRODUCTION

A key part of power system security is keeping an eye on and evaluating the different possible problems that could happen in the system and then choosing the worst-case scenarios from those evaluations. For reliability in a power grid, it's essential that there be no interruptions in the flow of electricity and no drops in load. In order to accomplish this, security analysis is carried out in order to establish various control mechanisms that ensure the avoidance and survival of emergency situations while also operating the system at the lowest feasible cost. In an emergency situation, the power system is said to be in a state of emergency when a predetermined limit of the system is violated. The occurrence of these limits being exceeded is due to activities occurring in the power system. In today's sophisticated energy management systems, contingency analysis plays a crucial role. The study of contingency analysis entails doing efficient calculations of system performance from a set of simplified system settings in order to estimate system stability immediately after outages are experienced. The calculation of the performance index determines the severity of the contingencies mentioned in [1]. Contingencies are commonly described as potentially dangerous disruptions that occur while a power system is operating in its steady-state functioning described in [2]. In order to do a contingency analysis, it is necessary to compute entire load flow estimates following each and every probable outage event, including outages occurring on multiple transmission lines and generators as described in [3]. Consequently, the list of possible contingency scenarios becomes extremely long, and the process becomes extremely time-consuming. In order to mitigate this problem, automatic contingency screening is being adopted. This method locates and ranks the power system's worst-case possibilities, as presented in [4]. In order to screen out the contingencies, they are ranked according to their performance indexes, with higher values indicating greater seriousness, as presented in [5]. With increasing uncertainty, it's hard to plan transmission systems in this setting. Obviously, the most significant causes of uncertainty in transmission system planning are load demand growth and unscheduled exchanges with neighbouring systems, which are compensated by incorporating the suitable flexible AC transmission system (FACTS) devices discussed in [6]–[8]. However, in the present day, due to the unbundling of electrical firms, there is a lot of uncertainty over the functioning of existing generation plants, the decommissioning of generation units, and the location of future power plants in [9]. In addition, numerous approaches to adjusting transmission planning functions should be explored due to the diversity in the energy markets as a result of the varying economic, political, social, and regulatory settings. The evolution of transmission power flows, the volume of power imported and exported, and the size and location of new power plants are all factors that must be taken into account by transmission planning functions in order to be successful, as described in [10]. In order to evaluate the power flow analysis, the contingency study needs to be carried out for the various scenarios. In this study, the huge amount of data collected through the rigorous simulation needs to be processed and pre-processed to convert it into the useful information mentioned in [11]. Power systems contingencies can utilise big data analytics to make the most of the massive volumes of data they generate. This data can then be used to leverage the optimisation processes that are already taking place in power grids. The application of big data techniques will result in an increase in the overall efficiency of the electric power network, as mentioned in [12].

Electric utilities are undergoing a technological revolution that includes the implementation of two-way communications networks, information technologies, and distributed intelligent devices to improve distribution system monitoring and control [13]. A utility's information systems will have to store and manage more information as a result of these developments. There can be a substantial amount of information produced by AMI/AMR, SCADA, simulation results, and other intelligent devices. One frequent approach is to simply amass as much information as possible and figure out what to do with it later. There is a direct correlation between the growth in data volumes and the demand for more complex and expensive IT systems and personnel. Though complete data collection is possible, it is unlikely that it would be kept or organised in such a way that it would be useful in the long run. Big data analytics has been widely employed to address most of the challenges in the power system, proving it to be a good and promising instrument for dealing with massive volumes of data [14].

Big data mining in the power sector and analysis of early detection of contingencies in the power sector can help plan for significant savings. This effort to save money on hardware would be possible since mining would reduce the computational complexity of the contingency analysis. A data transformation strategy is required for data mining in order to reduce the dimensionality of the data used in the mining process [15]. A hybrid approach to data transformation, combining data cleaning with principal component analysis, as discussed in [14]. Data mining performance indicators have few empirical studies. This study examined how data mining classification algorithms perform with larger inputs. The multi-layer perceptron (MLP), neural network, and nave bayes were tested with varied simulated data amounts, as discussed in [10]. Data classification is an essential part of the data mining process. It involves the extraction of models describing classes and the prediction of the appropriate class for individual data instances, as discussed in [16]. Multiple established classifiers can be used nowadays. Weka Explorer is used to apply various classification trees (decision stump, hoeffding tree, J48, LMT, random forest, and REP tree) to a variety of datasets discussed in [14]. A representative set of attributes to build a classification model is a central topic in machine learning. Machine learning's attribute selection difficulty is well known [17]. It offers probabilistic categorization and performs well on benchmarks. Attribute selection involves choosing a small group of features or attributes to predict target labels well. Attribute selection decreases the computational complexity of learning and prediction systems and saves on useless feature measurements. Attribute selection for machine learning uses regression analysis with forward selection, backward elimination, and quick reduction. AIC is used to evaluate proposed techniques [18]. Power system modelling and simulation have developed along with the expansion of power grids and the development of computational methods discussed in [19]. Data mining simplifies contingency analysis by using the mined data to classify contingency levels using the multi-class support vector machine (MCSVM) and multi-class relevance vector machine (MCRVM) discussed in [20]. Big data analysis helps remove faulty data from the system and transmit contingency data to the planning power engineer, as presented in [21]. The visualisation techniques are used to highlight the impact of features on outage occurrence, and association rule mining is used to uncover factors connected to each outage type as well as each other [22]. According to the presented survey, there has been sufficient work done in the areas of modelling and analysis, contingency ranking, critical bus ranking, and the incorporation of voltage collapse phenomena, as well as the development of FACTS device models. In the meantime, it has been noted that contingency condition prediction using data mining techniques is a focus area and can more accurately predict the severity of the system than traditional severity ranking methods.

This article has eight sections: i) Section 2 explains contingency analysis; ii) Section 3 calculates the line voltage stability index; iii) Section 4 discusses 4 proposed frameworks for contingency analysis; iv) Section 5 experimental results and discussion; and v) The section 6 concludes.

2. CONTINGENCY ANALYSIS

Both the active power flow limit and the reactive power limit, which has a substantial impact on the bus voltage, are subject to change during a transmission line contingency, making it crucial to predict both the power flow and the bus voltages in the aftermath of the event. Since a key part of any contingency analysis is running simulations of each potential scenario against the baseline model of the power grid, there are three significant challenges associated with this type of analysis. The primary challenge is the intricacy of creating a reliable model of the power grid. Secondly, the energy management system spends an inordinate amount of time computing the power flow and bus voltages, which is a problem because of the difficulty involved. Thirdly, it is reasonable to divide the online sensitivity analysis into three parts: defining the sensitivity, selecting the appropriate sensitivity measures, and evaluating the results. The definition of a contingency includes all the potential problems that could arise in a power system, as well as the steps taken to compile a list of solutions to those problems. The term "contingency selection" refers to the method of narrowing down a list of potential disasters by choosing only the most desperate scenarios that result in severe violations of safety constraints like maximum power flow and bus voltage. This system employs index calculations to rank the seriousness of potential events. The ranking of the contingency cases is determined by the outcomes of these index computations [1]. Next, the effect of the possible disruption is figured out, and the controls or security measures that need to be in place to stop more damage are put in place. Choosing which potential events will cause a breach in operational constraints is called "contingency selection." The performance Indices are a type of severity index that is then used to select the potential outcomes. Offline, these indices are computed using standard power flow algorithms for specific scenarios. The results are used to rank the contingencies, with the one having the highest PI value coming in first. The analysis is then performed, beginning with the highest-ranked contingency and continuing until no catastrophic contingencies remain.

3. LINE VOLTAGE STABILITY INDEX

In order to do a contingency analysis, the conventional alternating current flow solution provides active and reactive power flows as well as bus voltage magnitudes. The power system's line importance as well as its contingency ranking technique has been established. The ranking is accomplished through the application of a voltage stability index that is based on the results of the severity calculation. The NR method is utilized in order obtain the load flow solutions to study voltage stability index making use of each scenario of contingency load and to investigate NR [11]. The Figure 1 shows single line diagram of two bus system.

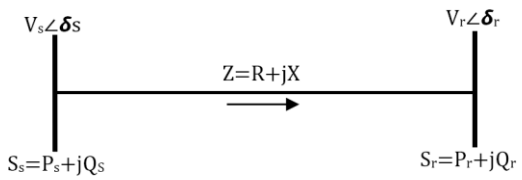

Figure 1. Single line representation of two bus system

4. PROPOSED FRAMEWORK FOR CONTINGENCY ANALYSIS

The proposed framework for a contingency study is a structured approach to assessing and preparing for potential future events or situations that may disrupt normal operations or plans. It provides a systematic methodology for identifying risks, evaluating their potential impact, and developing strategies to mitigate or manage them effectively. The proposed framework for the contingency analysis of the power system model includes various stages such as data collection, data processing, training the machine learning model, and prediction of contingencies based on the training model, as given in Figure 2.

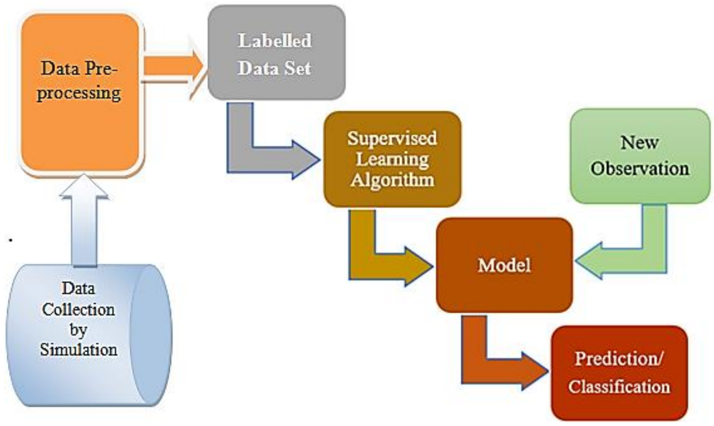

Figure 2. Data mining process applied to contingency study of power system

4.1. Data collection

Power systems are being operated in a stressed condition mainly due to the ever-increasing load demand, depleting energy resources, and environmental constraints on transmission line expansion. The system stability is one of the major concerns for the power engineers to operate the system in its rated capacity. In order to overcome some of these problems and to enhance the system performance in many power systems the flexible AC transmission system (FACTS) devices are being used [8]. The system studies with respect to the contingencies are to be revaluated due to the connection of FACTS devices in the system. In the analysis of contingency study, the following data were considered: i) System data such as bus number, bus code, voltage magnitude, angle in degree, load in MW, and MVAr, generators data like MW, MVAr, Q_{min} and Q_{max}, injected MVAr; and ii) Line data such as line number, resistance and reactance of the line, half line charges, transformer details. In this case IEEE 30 bus system is considered as case study and the data sets are generated under various operating line outages of the power system network. As the simulations results in huge data and to enable the system planner to arrive at the useful information from this huge data.

4.2. Data preprocessing

One of the significant steps for data mining is data preprocessing, which transforms the collected raw data into a form suitable for training the data mining models. Label encoding is one such pre-processing step, which converts the labels of an attribute that are in human readable form in the given data set, into numbers [18]. The data mining methods will later decide on how to operate on these numbers by converting them into machine-readable form. Table 1 shows how label encoding transforms the attribute namely ‘severity condition’ in this work from human-readable form into numbers.

Table 1. Label encoding of “severity condition”

Labels before encoding	Critical	Semi-critical	Non-critical
Numbers after encoding	0	1	2

4.3. Training the machine learning models

4.3.1. Decision tree

Decision trees are frequently used in data mining applications for predicting a target variable which is discrete or continuous in nature. The internal/core nodes of a decision tree stand for the qualities/attribute test conditions being tested, the branches for the results of those tests, and the leaf nodes (terminal nodes) for the target labels [23]. In order to learn a tree, the source set must be partitioned into subsets with values for the attributes serving as the dividers. This method (called recursive partitioning) is applied to each newly derived subgroup. Each node provides an opportunity to partition the prediction into subsets whose members share a common value for the target variable [19]. The decision on whether to divide a subgroup further or not is based on the traditional impurity measures such as entropy and Gini index from information theory. The entropy of a set S, with n samples and n_c number of distinct values of the target class is given by (1).

$$Entropy(S) = -\sum_{i=1}^{n_c} p_i \; log_2 p_i \tag{1}$$

Where p_i is the probability of the ith class in S.

For a data set with two distinct values of the target class, entropy of a group/partition will be the maximum value, which indicates the decision about the target for this group is totally unclear. Hence the decision tree induction algorithm splits this group further into smaller and pure partitions based on another attribute test condition [24]. On the other hand, if the entropy of the set is the minimum, zero, the algorithm ends in a clear decision about the target variable. Algorithms for pruning the decision trees also help to avoid training over fitted and under fitted models. Apart from this, pruning also helps to speed up the inference and reduces the storage size of the models.

4.3.2. Bagging based ensemble method

In spite of being simple to train and use in inference, decision trees suffer from the problem of instability. Small variations in the training data, will generate a completely different decision tree. This problem is mitigated by training multiple decision trees in an ensemble learner, where the features, and samples are sampled randomly with replacement and used for training the ensemble learners. The training and testing phase of the ensemble technique namely bagging is given in Figures 3 and 4. The Following is pseudo code.

Bagging (D, n, k, T):
Input: D–the training data set, n–the no. of samples and k---the no. of base learners, T---the test data set
Output: An ensemble of decision trees
Begin
    Using sampling with replacement on D, create multiple data sets D_i for i=1 to k
    Train k no. of base learners using the data set D_i for the ith learner, where i=1 to k
    For each record t in the test data set T, find the predicted output of this test data t by all the base learners
    Apply majority voting on the predicted class labels of t, to find the ensemble output C^*
End

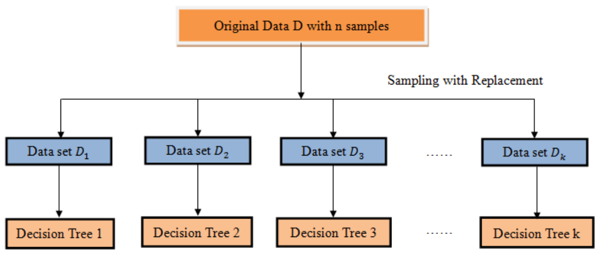

Figure 3. Training phase of bagging

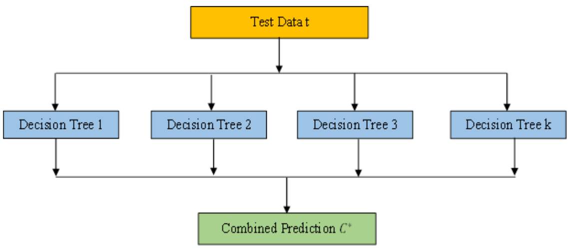

Figure 4. Testing phase of bagging

4.3.3. AdaBoost method

This class of ensemble methods also create multiple data sets D_i from the original data set D where i=1 to k, no. of base learners. However, unlike bagging, the base learners are trained in a sequential manner and the samples are also assigned weights at the end of each iteration. First a base learner is trained using D_1 which is created using sampling with replacement from D. This base learner is used to predict the class of the training instances. All samples that are wrongly predicted by this learner increase in weight and those that are correctly predicted will decrease in weights. The next data set D_2 is created for the next learner using sampling with replacement on the newly assigned weights of samples. Same process is repeated until all k base learners are trained. Updating the weights of the samples at the end of each round will make the wrongly predicted samples become more and more prevalent in subsequent iterations. The prediction error rate of each base learner is also used to perform weighted majority voting of the final ensemble output C^* for each test data. One of the commonly used boosting based ensemble technique is the AdaBoost method. The pseudocode of the AdaBoost algorithm is given in Table 2.

Table 2. The pseudocode of the AdaBoost algorithm

AdaBoosting (D, n, k, T): Input: D–the training data set, n–the no. of samples and k---the no. of base learners, T---the test data set
Output: an ensemble of decision trees Begin Step 1. Initialize the weight of all training samples as 1/n. Step 2. Repeat the following steps for i=1 to k   2.1. Create the bootstrap sample B_i for the base learner C_i

2.2. For all samples in D, find the predicted output by this learner C_i
2.3. Calculate the error rate of this learner as

$$E_i = \begin{cases} \frac{1}{n} \sum_{j=1}^{n} \begin{cases} w_j \text{ if this sample is correctly predicted} \\ 0 \text{ if this sample is wrongly predicted} \end{cases} \end{cases},$$

2.4. Calculate the weight of this classifier as α_i=(½) ln 〚$(1-Ei)/Ei$〛
2.5. Increase the weight of wrongly predicted samples and decrease the weight of correctly predicted samples.

$$W^{next\ round} = \frac{W^{current\ round}}{z} \begin{cases} exp^{-\alpha_i \text{ if it is wrongly predicted}} \\ exp^{\alpha} \text{ if it is correctly predicted} \end{cases}$$

Step 3. For each record in the test set T, find its predicted output

$$C^* = argmax_y \sum_{j=1}^{k} \begin{cases} \alpha_j \text{ if it is correctly predicted} \\ 0 \text{ otherwise} \end{cases}$$

where y is the set of all class labels

4.4. Performance metrics for predictive accuracy

The performance of these classifiers are measured based on the various metrics based on the confusion matrix such as predictive accuracy, precision, F1 score. Table 3 summarizes the various metrics used for evaluating the performance of the trained classifiers where TP, TN, FP, and FN are the true positives, true negatives, false positives, and false negatives respectively [25].

Receiver operating characteristics curve (ROC) is another metric used for evaluating the performance of the classifier. It is a plot between the true positive rate and false positive rate of the classifier. The area under the receiver operating characteristics curve (AUC) of the classifier is calculated using the trapezoidal rule. It is a value between 0 to 1 and for an ideal classifier it is exactly 1.

Table 3. Various performance metrics

Metric	Formula	Definition
Accuracy	$\left[\frac{(TP+TN)}{(TP+FP+TN+FN)}\right]$	Accuracy defines the number of correct predictions made by the classifier out of all predictions
Precision	$\left[\frac{(TP)}{(TP+FP)}\right]$	Precision specifies the ability of a classification model to predict only the samples of a particular class
Recall	$\left[\frac{TP}{(TP+FN)}\right]$	Recall specifies the ability of a classification model to predict all samples of a particular class
F1 score	$\left[2 * \left(\frac{Precision * Recall}{Precision + Recall}\right)\right]$	It combines precision and recall. It is mainly used for evaluating classifiers trained with data sets having imbalanced class distribution

5. EXPERIMENTAL RESULTS AND DISCUSSION

The IEEE 30 bus system is considered for the system study. This system is consists of 1-slack buses, 5-generator buses, 24 load buses, and 41 transmission lines. The total active load on the system is 283.400 MW and the total reactive power on the system is 126.20 MVAr. In this case load flow analysis is carried on base load condition without any line outage and without incorporating unified power flow controller (UPFC) to the system. Power flow solution is achieved by using the newton-raphson method. The maximum power mismatch is considered as 7.54898×10^-07, the system is converged at 4th iteration and the time taken for the computing is 1.2406×10^-04. The Total active power loss in the system is 17.5985 MW and the total reactive power loss in the system is 22.2444 MVAr. The transmission lines are classified into three categories like critical, semi-critical, and non critical by estimating the line voltage stability severity index. The MATLAB software was used to carry out the simulation work and generated data and applied the proposed frame work as mentioned in the section four. The sample data of line voltage stability index is shown in the Table 4.

The data from the simulations with MATLAB were converted to a structured format with line number, compensator, and load condition and Lmn value as independent variables and severity condition as dependent variable. Table 5 shows the first 10 samples of this structured data. Values of ‘compensator’ are either ‘with UPFC’ or ‘without UPFC’. Values of the ‘severity condition’ are either ‘critical’, ‘semi critical’ or ‘non critical’. These two variables ‘compensator’ and ‘severity condition’ are pre-processed using the label encoding technique in scikit-learn library. The preprocessed data set is shown in Table 6. The decision tree classifier for predicting the severity condition was trained with the pre-processed data set using ‘entropy’ as the splitting criteria of a node. The bagging based ensemble method was trained on the pre-processed data set using classification and regression trees (CART) as the base learners and 100 such base learners. The Boosting based ensemble method was trained on the pre-processed data set and Figure 5 shows the confusion matrices of the three classifiers namely decision tree, bagging and AdaBoosting classifier respectively.

Table 4. Sample simulation data of line voltage stability index for different contingencies

C No	Line No	Lmn1	Lmn2	Lmn3	Lmn4	Lmn5	Lmn6	Lmn7	Lmn8	Lmn9	Lmn10
1	2	0.1014	0	0.10693	0.07164	0.15671	0.00341	0.02228	0.02176	0.03014	0.0051
2	3	0.0868	0.01234	0	0.06941	0.14933	0.06033	0.01711	0.04948	0.02997	0.0048
	4	0	0	0	0	0	0	0	0	0	0
3	5	0.1010	0.03630	0.05811	0.08990	0	0.04055	0.03105	0.03094	0.10896	0.0303
4	6	0.0414	0.02858	0.12520	0.09097	0.04776	0	0.04754	0.03738	0.01999	0.0099
5	7	0.0779	0.06002	0.05673	0.07571	0.1897	0.03736	0	0.03575	0.02404	0.0079
6	8	0.0594	0.03446	0.0309	0.10002	0.07205	0.00661	0.0172	0	0.03690	0.0197
7	9	0.0659	0.03204	0.0189	0.05326	0.05345	0.00385	0.0146	0.06312	0	0.0196
8	10	0.1052	0.01753	0.02348	0.041312	0.070609	0.026938	0.02286	0.059703	0.034308	0

Table 5. Sample data before preprocessing

Sl. No	Line number	Compensator	Load condition	Lmn value	Severity condition
0	2	Without UPFC	100	0.322331	Critical
1	3	Without UPFC	100	0.255147	Critical
2	5	Without UPFC	100	0.307364	Critical
3	6	Without UPFC	100	0.2594	Critical
4	7	Without UPFC	100	0.21755	Semi critical
5	8	Without UPFC	100	0.250169	Semi critical
6	9	Without UPFC	100	0.194066	Non critical
7	10	Without UPFC	100	0.192362	Non critical
8	14	Without UPFC	100	0.323621	Critical
9	17	Without UPFC	100	0.192517	Non critical

Table 6. Sample pre-processed data set

Sl. No	Line number	Compensator	Load condition	Lmn value
0	2	1	100	0.322331
1	3	1	100	0.255147
2	5	1	100	0.307364
3	6	1	100	0.259400
4	7	1	100	0.217550
..	…	…	…	…
..	…	…	…	…
115	23	0	150	0.338358
116	24	0	150	0.369273
117	25	0	150	0.375541
118	26	0	150	0.388223
119	27	0	150	0.360587

$$\underset{5a}{\begin{bmatrix} 6 & 0 & 1 \\ 0 & 15 & 1 \\ 4 & 3 & 6 \end{bmatrix}} \quad \underset{5b}{\begin{bmatrix} 6 & 1 & 0 \\ 0 & 16 & 0 \\ 3 & 3 & 67 \end{bmatrix}} \quad \underset{5c}{\begin{bmatrix} 3 & 0 & 4 \\ 0 & 15 & 1 \\ 1 & 3 & 6 \end{bmatrix}}$$

Figure 5. Confusion matrix of the trained classifiers

The performance metrics that are based on these confusion matrices for all three classifiers are shown in Table 7. The experimental results have shown that the bagging-based ensemble methods outperform the other two classifiers in terms of all the performance measures. The Figures 6 to 8 shows the ROC analysis for the three classifiers. It can be seen from these results the area under the ROC curve (AUC) is the maximum for the bagging-based ensemble method than the other two classifiers. The ideal case of AUC being 1 is achieved for class 1 in the case of bagging classifier.

Table 7. The performance comparison of decision tree, bagging, and AdaBoost classifier

Data mining model	Classification accuracy	Precision	Recall	F1Score
Decision tree classifier	0.75	0.76	0.75	0.74
Bagging classifier	0.85	0.85	0.81	0.79
AdaBoost classifier	0.67	0.66	0.67	0.65

These experimental results have shown that the performance of the classifiers in predicting the severity condition is more with the ensemble-based method namely bagging with classification and regression trees as the base learners. The bagging-based ensemble models provide scope for training the base learners in parallel and hence speed up the training phase for prediction.

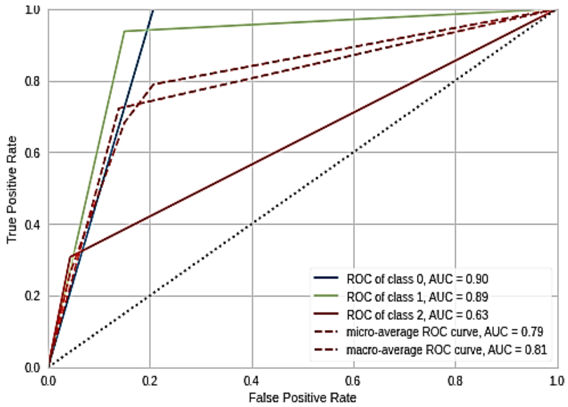

Figure 6. ROC curve for the decision tree-based severity prediction model

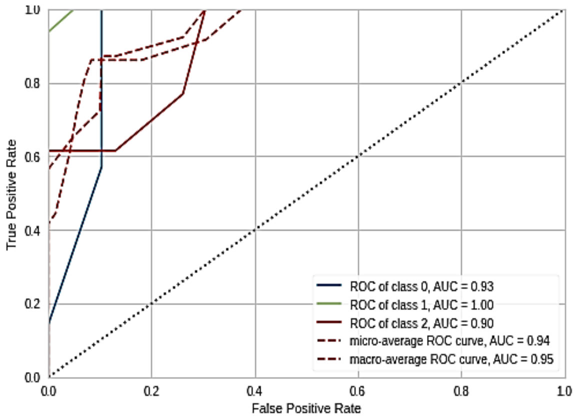

Figure 7. ROC curve for the bagging-based severity prediction model

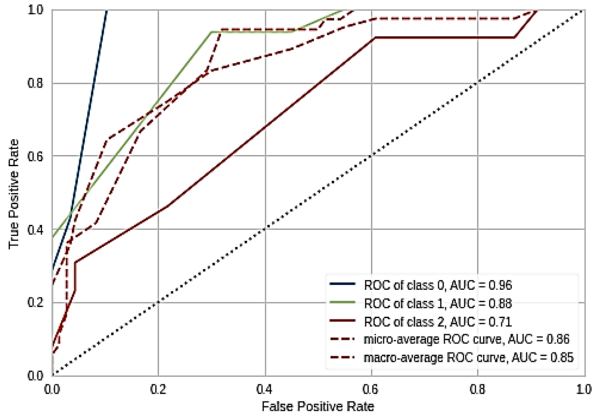

Figure 8. ROC curve for the boosting based severity prediction model

6. CONCLUSIONS

The outcomes of the simulation yield a sizable dataset with a variety of attributes to assess the contingency analysis. The contingency prediction has been carried based on the different classification methods from a data mining perspective. The decision tree classifier, bagging classifier, and AdaBoost classifier classification methods are employed and have given accurate predicted results compared to the manual classification. The decision tree classifier predicted the severity condition with 75% of accuracy, the bagging classifier predicted severity condition of with 85% of accuracy and the AdaBoost classifier is predicted the severity condition with 67% of accuracy based on the trained data set for different load conditions and contingency conditions. The severity of the line/ was predicted as critical, semi critical or non-critical by the trained models. The bagging classifier was found to perform well compared to other two classifiers.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to the administrations of both Presidency University in Bengaluru and Mohan Babu University in Tirupati for their constant support and the availability of facilities essential to the completion of this study.

REFERENCES

[1] M. Jain, P. S. Venkataramu, and T. Ananthapadmanabha, "Critical bus ranking under line outage contingencies," in *2007 Proceedings of the IASTED International Conference on Energy and Power Systems*, 2007, pp. 69–74.

[2] L. M and M. S. Nagaraj, "Principle component analysis based data mining for contingency analysis on IEEE 30 bus power system," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2, pp. 878–882, Dec. 2019, doi: 10.35940/ijitee.B7116.129219.

[3] T. U. Badrudeen, F. K. Ariyo, A. O. Salau, and S. L. Braide, "Analysis of a new voltage stability pointer for line contingency ranking in a power network," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3033–3041, Dec. 2022, doi: 10.11591/eei.v11i6.4266.

[4] P. Venkatesh and N. Visali, "Investigations on hybrid line stability ranking index with polynomial load modeling for power system security," *Electrical Engineering & Electromechanics*, no. 1, pp. 71–76, Jan. 2023, doi: 10.20998/2074-272X.2023.1.10.

[5] F. Hussian, G. R. Goyal, A. K. Arya, and B. P. Soni, "Contingency ranking for voltage stability in power system," in *2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Jul. 2021, pp. 1–4, doi: 10.1109/CONECCT52877.2021.9622724.

[6] P. K. Gouda and P. K. Hota, "Impact of FACTS devices in optimal generation in deregulated power system," *International Journal of Applied Power Engineering (IJAPE)*, vol. 4, no. 3, pp. 118–125, 2015.

[7] P. Duraisamy and A. Ponnusamy, "Power system performance improvement by optimal placement and sizing of SVC using genetic algorithm," *International Journal of Applied Power Engineering (IJAPE)*, vol. 6, no. 2, p. 55, Aug. 2017, doi: 10.11591/ijape.v6.i2.pp55-62.

[8] R. V. Angadi, S. B. Daram, P. S. Venkataramu, and V. J. Manohar, "Data mining and machine learning technique in contingency analysis of power system with UPFC," *Distributed Generation & Alternative Energy Journal*, pp. 1305–1328, May 2022, doi: 10.13052/dgaej2156-3306.3751.

[9] A. E. Airoboman, P. James, I. A. Araga, C. L. Wamdeo, and I. K. Okakwu, "Contingency analysis on the Nigerian power systems network," in *2019 IEEE PES/IAS PowerAfrica*, Aug. 2019, pp. 70–75, doi: 10.1109/PowerAfrica.2019.8928883.

[10] K. Pal, S. Sachan, F. Gholian-Jouybari, and M. Hajiaghaei-Keshteli, "An analysis of the security of multi-area power transmission lines using fuzzy-ACO," *Expert Systems with Applications*, vol. 224, p. 120070, Aug. 2023, doi: 10.1016/j.eswa.2023.120070.

[11] R. V Angadi, S. Babu Daram, and P. S. Venkataramu, "Analysis of power system security using big data and machine learning techniques," in *2020 IEEE 17th India Council International Conference (INDICON)*, Dec. 2020, pp. 1–8, doi: 10.1109/INDICON49873.2020.9342458.

[12] S. Rongrong, L. Qing, S. Xin, N. Baifeng, and W. Qiang, "Application of big data in power system reform," in *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, Apr. 2021, pp. 1340–1342, doi: 10.1109/IPEC51340.2021.9421337.

[13] P. Venkatesh and N. Visali, "Machine learning for hybrid line stability ranking index in polynomial load modeling under contingency conditions," *Intelligent Automation & Soft Computing*, vol. 37, no. 1, pp. 1001–1012, 2023, doi: 10.32604/iasc.2023.036268.

[14] R. V. Angadi, S. B. Daram, and P. S. Venkataramu, "Role of big data analytic and machine learning in power system contingency analysis," in *Smart Electrical and Mechanical Systems*, Cambridge, MA, USA: Academic Press, 2022, pp. 151–184, doi: 10.1016/B978-0-323-90789-7.00004-X.

[15] N. Minh Khoa and L. Van Dai, "Detection and classification of power quality disturbances in power system using modified-combination between the stockwell transform and decision tree methods," *Energies*, vol. 13, no. 14, p. 3623, Jul. 2020, doi: 10.3390/en13143623.

[16] R. Samya, "Attribute selection using machine learning technique," *International Journal of Computational Intelligence and Informatics*, vol. 7, no. 2, pp. 122–132, 2017.

[17] K. Chen, Z. He, S. X. Wang, J. Hu, L. Li, and J. He, "Learning-based data analytics: moving towards transparent power grids," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 1, pp. 67–82, Mar. 2018, doi: 10.17775/CSEEJPES.2017.01070.

[18] H. Akhavan-Hejazi and H. Mohsenian-Rad, "Power systems big data analytics: an assessment of paradigm shift barriers and prospects," *Energy Reports*, vol. 4, pp. 91–100, Nov. 2018, doi: 10.1016/j.egyr.2017.11.002.

[19] Y. Chen, Z. Huang, S. Jin, and A. Li, "Computing for power system operation and planning: then, now, and the future," *iEnergy*, vol. 1, no. 3, pp. 315–324, Sep. 2022, doi: 10.23919/IEN.2022.0037.

[20] S. Prasher, L. Nelson, A. S. Sindhu, S. Sumathi, and M. Jagdish, "MCSVM and MCRVM based contingency classification model," in *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)*, Mar. 2023, pp. 440–443, doi: 10.1109/ICEARS56392.2023.10085481.

[21] S. B. Daram, V. Ak, R. V Angadi, and K. Sesiprabha, "An approach to contingency severity prediction using data analytic techniques in a power system," in *2022 International Conference on Smart and Sustainable Technologies in Energy and Power Sectors (SSTEPS)*, Nov. 2022, pp. 63–67, doi: 10.1109/SSTEPS57475.2022.00028.

[22] M. S. Bashkari, A. Sami, and M. Rastegar, "Outage cause detection in power distribution systems based on data mining," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 640–649, Jan. 2021, doi: 10.1109/TII.2020.2966505.

[23] S. Manimaran, I. AlBastaki, and J. A. Mangai, "An ensemble model for predicting energy performance in residential buildings using data mining techniques," *ASHRAE Transactions*, vol. 121, pp. 402–410, 2015.

[24] P. Sarang, "Decision tree," in *Thinking Data Science*, Cham, Switzerland: Springer, 2023, pp. 75–96, doi: 10.1007/978-3-031-02363-7_4.

[25] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug. 1996, doi: 10.1007/BF00058655.

BIOGRAPHIES OF AUTHORS

Ravi V. Angadi received his B.E. in Electrical and Electronics Engineering from VTU, Belagavi, Karnataka (India) in 2010, and M. Tech. degree in Power Electronics from JNTUA, Anantapur (India) in 2014 and Ph.D. from Presidency University, Bengaluru in 2023. He is currently working as an Assistant Professor in the Department of Electrical and Electronics Engineering at Presidency University, Bengaluru, Karnataka, (India). He has guided UG students' projects sponsored by KSCST, DST and VTU-RGS and one project has been applied for Patent. He has published many papers in national/international journals/conferences/book chapter. He was a Governing Council Member at SSCE, Bengaluru during the AY 2017-18. Mr. Ravi is a life member of IE (I) and ISTE, MIEEE. He can be contacted at email: raviangadi4045@gmail.com.

J. Alamelu Mangai received her Ph.D. from BITS Pilani, Dubai Campus in 2015. She is working as Professor in the Department of Computer Science and Engineering at Presidency University, Bangalore. Her research interests include data mining and machine learning algorithms, and applications. She can be contacted at email: alamelu.jothidurai@presidencyuniversity.in.

Dr. V. Joshi Manohar currently working as Professor and HoD at Presidency University, Itgalpur, India. He received his Ph.D. in Electrical Drives from Jawaharlal Nehru Technological University, Anantapur, India in 2015. M. Tech. in Power Electronics from VTU, Belgaum, KA, India in 2004 and B. Tech. degree in Electrical and Electronics Engineering from Nagarjuna University, Guntur, AP, India, 2000. His research area includes the control of multi-level inverters using soft computing techniques. Reactive power compensation at low switching frequency and AI-based electrical drive control. He's a life member of the ISTE, Senior IEEE Member and Fellow Institute of Engineers. He can be contacted at email: joshivmanohar@presidencyuniversity.in.

Dr. Suresh Babu Daram received his B. Tech. in Electrical and Electronics engineering from JNTU Hyderabad (India) in 2006, M. Tech. in Power Systems Engineering from Acharya Nagarjuna University (India) in 2009 and Ph.D. in Power Systems from Visvesvaraya Technological University, Belgaum (India) in 2018. He was Assistant Professor in the Dept. of Electrical & Electronics at GGITM Bhopal from 2009-2015. Currently he is Professor in Department of Electrical and Electronics at Sree Vidyanikethan Engineering College, Tirupati (A.P), India. He has received Best Teacher Award from MPCST in 2014 and has best paper award in International Conference "Dr. M. H. Rashid Best paper award" in 2016, "National Conference best paper award" in 2016, "National Techno Conference best paper award" in 2020. He has published more than 55 national/international journal/conference papers/book chapters. His research interests include energy management systems, power system optimization, and voltage instability studies incorporating FACTS controllers' power system security analysis, data analytics and machine learning. Dr. Suresh is a member of IEEE, AMIE (India), IAENG, CSTA, IACSIT, IRED and student member-ASTM. He can be contacted at email: sureshbabudaram@gmail.com.

Paritala Venkateswara Rao is pursuing his fourth-year Bachelor's degree in the stream of Computer Engineering at Presidency University, Itgalpura, Rajankunthe, Yelahanka, Bangalore-560 064. He is a polyglot programmer with experience in Java, C, and Python. He is a voracious reader with a special interest in data analytics, data mining, and cybersecurity. An active member and serving as Vice-President in the ROTARACT Club at Presidency University, Bangalore as well as an active member in the National Service Scheme (NSS). He enjoys programming, playing badminton, photography, and loves to travel. He has an interest in researching technologies like artificial intelligence, machine learning, deep learning, and affiliated fields. He can be contacted at email: venkateshparitala4@gmail.com.