May 2019

Fault Diagnosis of HVDC Systems Using Machine Learning Based Methods

Yu Chen
University of Wisconsin-Milwaukee

FAULT DIAGNOSIS OF HVDC SYSTEMS USING MACHINE LEARNING BASED METHODS

by

Yu Chen

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

in Engineering

at

The University of Wisconsin-Milwaukee

May 2019

ABSTRACT

FAULT DIAGNOSIS OF HVDC SYSTEMS USING MACHINE LEARNING BASED METHODS

by

Yu Chen

The University of Wisconsin-Milwaukee. 2019
Under the Supervision of Professor Lingfeng Wang

With the development of high-power electronic technology, HVDC system is applied in the power system because of advantages in large-capacity and long-distance transmission, stability, and flexibility. Therefore, as the guarantee of reliable operating of HVDC system, fault diagnosis of the HVDC system is of great significance. In the current variety methods used in fault diagnosis, Machine Learning based methods have become a hotspot. To this end, the performance of several commonly used machine learning classifiers is compared in HVDC system. First of all, nine faults both in AC systems and DC systems of the HVDC system are set in the HVDC model in Simulink. Therefore, 10 operating states corresponding to the faults and normal operating are considered as the output classes of classifier. Seven parameters, such as DC voltage and DC current, are selected as fault feature parameters of each sample. By simulating the HVDC system in 10 operating states (including normal operating state) correspondingly, 20000 samples, each containing seven parameters, be obtained during the fault period. Then, the training sample set and the test sample set are

established by 80% and 20% of the whole sample set. Subsequently, Decision Trees, the Support Vector Machine (SVM), K-Nearest Neighborhood Classifier (KNN), Ensemble classifiers, Discriminant Analysis, Backward Propagation Neural Network (BP-NN), long Short-Term Memory Neural Network (LSTM-NN), Extreme Learning Machine (ELM) was trained and tested. The accuracy of testing is used as the performance index of the model. In particular, for BP-NN, the impact of different transfer functions and learning rules combinations on the accuracy of the model was tested. For ELM, the impact of different activation functions on accuracy is tested. The results have shown that ELM and Bagged Trees have the best performance in HVDC fault diagnosis. The accuracy of these two methods are 92.23% and 96.5% respectively. However, in order to achieve better accuracy in ELM model, a large number of hidden layer nodes are set so that training time increases sharply.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First of all, I want to express my sincere gratitude to my advisor Dr. Lingfeng Wang. During my whole graduate study at UW-Milwaukee, he gave me a lot of help both in my study and life. Also, his comments and suggestions enabled my work to progress smoothly. His patient answers to my questions were really helpful to my research. Also, Dr. Zhaoxi Liu's guidance is vital to my research. I would like to thank him for providing detailed suggestions on research directions.

Besides, I really appreciate the Scholarship from UWM and the financial assistance for research assistantship from my advisor. The financial support had a huge impact on my life in a different country, which enabled me to fully focus on my research instead of worrying about the living costs.

Moreover, I appreciate my parents for their support for my international student life and my emotion. Without their best effort, I can't imagine I could do what I have done today in my graduate study in US. I will try my best to do further, more in-depth research with their continued support. I dedicate this thesis to them and my advisor.

This work was supported in part by the National Science Foundation Industry/University Cooperative Research Center on Grid-connected Advanced Power Electronic Systems (GRAPES).

1 Introduction

1.1 Background of HVDC System

Electricity plays an inestimable role in the national economy as well as in people's lives. Understanding of electricity, application of electricity and development of power science all begins with direct current. French physicist M. Pieletz transfer a 1.5W, 1.5-2kV DC power to through a telegraph line to drive the water pump, which is the first DC transmission test in human history. Although power loss in this test is 78%, it is the beginning for high-voltage, long-range, large-capacity direct current transmission. Since the system in this test has power generator, transmission line, and power consumption equipment, it's considered as the first power system in the world.

The first DC power system which has whole voltage transfer structure was put into operating in 1954. The transmission line connects the Swedish Island of Gotland and the Swedish mainland to transfer a 100kV and 2000MW power. The distance between the two places is 96km. Mercury arc valve is used as the converter in this power system. After that, Thyristor is first used in Irkutsk DC transmission project in Canada in 1972. the 320MW,270KV transmission project used an SCR thyristor valve converter Instead of a mercury arc valve converter. In the 1960s, the appearance of SCR (Silicon Controlled Rectifier) opens up a new way for converters. In the past few decades, HVDC system related technologies are gradually maturing. HVDC transmission technology has become much better than ever. DC transmission construction costs, as well as operating energy consumption decline, Reliability is gradually improved in the meantime [1]. As a result, the HVDC system has been wildly applied in multiple fields in the power system.

1.1.1 Advantages

The role of the HVDC system is increasingly important in the power system. Compared with the traditional AC transmission mode, HVDC has several advantages, whether in terms of economy or performance. Mainly including the following aspects:

a. HVDC transmission technology has unique advantages in long-distance power transmission. In the AC power transmission, a significant phase difference will appear at each terminal of the transmission system. Although the frequency of AC power in the grid-connect systems is uniform to 50HZ, it often fluctuates. Because of these two factors, the AC system is not able to run synchronously. It's necessary to be adjusted with a complex and large compensation system and a strong comprehensive technology. Otherwise, a strong circulating current may be formed which is a throat to equipment or cause a power shutdown accident. However, no reactance is used in DC transmission so that the AC systems on both terminals don't have the need for synchronous operating. At the same time, since AC systems are connected through a DC system, frequency of one AC system can't affect another AC system so that HVDC transmission can connect two AC systems with different frequencies.

b. The line cost is low and the initial investment is small. AC transmission lines typically require 3 conductors, while DC transmission requires only 1 (unipolar) or 2 (bipolar). When conveying the same power, much less non-ferrous metals, steel, insulators, and other materials are used in DC transmission line. In addition, costs of transportation and installation are reduced.

c. The active loss of the line and the harm to the environment are smaller. When

the voltage level and the power being conveyed is the same, the active loss of HVDC transmission line is reduced by 1/3. Compared with the AC, no reactive power is lost in DC transmission. There is no capacitance current, no capacitance loss, no hysteresis loss and no eddy current loss in the steady operating state of HVDC transmission lines. Also, no parallel reactance compensation is required. The "space charge" effect of HVDC transmission lines results in less corona loss and radio interference, making it friendlier to the environment.

d. Different-frequency-network and asynchronous networking at the same frequency can be achieved. The DC system is theoretically applied in any transmission distance, and the transmission power will not be reduced due to the deterioration of the AC grid performance on both terminals, which can realize the “back-to-back” AC grid interconnection. Since DC transmission system is connected between AC systems, the AC power grids at both terminals can operate according to their respective frequencies and phases, without the need for synchronous adjustment. Two AC grids can also be standby and mutually supportive, thus improving the stability of the entire system after networking and the economic benefits of the power supply.

1.1.2 Disadvantages

Admittedly, although the HVDC system has the above advantages, in some ways, there are also weak [2]:

a. The equipment in the converter art in HVDC system cost much higher than the AC transmission system. As a result, cost of whole system and economic loss caused by serious faults increase.

b. Converters have a very high requirement on the reactive power even when the system is in a normal operating. So many components providing reactive power are in the HVDC system, which are expensive.

c. Requirements on capacity at the receiving terminal in AC system is high, because the DC system need a short circuit current from the AC system.

d. Since a large number of electronic components adopted in the converter station, harmonics are bound to be produced in the operating process. Further, voltage distortion caused by harmonic make the rectifier unstable, and a continuous phase change failure may occur in the inverter. Harmonics may lead to the relay protection equipment malfunction. In addition, harmonics will also have an adverse effect on the automatic devices and communication lines and direct current transmission in distribution networks.

1.1.3 Application of HVDC System

Based on the characteristics above, HVDC system is widely used in long-distance power transmission, power grid connection, transmission by submarine cable and power transmission by underground cable in large cities.

The types of HVDC power transmission application so far are:

a. Underwater cable [3] - Since the cable has a large capacity of capacitive charging reactive power, it is necessary to provide shunt reactor compensation in the middle of the line. However, it is impractical in AC power transmission. So HVDC is used when more than 30 km underwater cables are used in the transmission line. For example, the transmission line from Sweden FENNO to Finland SKAN's HVDC line uses

a 200km cable to spans the strait using HVDC.

b. An asynchronous connection between two AC systems - due to the different rated frequencies of AC systems, it is not appropriate to use an AC connection in this case. In addition, the gradual development of the two systems needs that they should be interconnected. Although they have the same frequency, and sometimes periods are not the same, using DC interconnection is also a common means. These two situations are pretty common in the United States and other countries such as India, Japan, Europe and etc.

c. Large-capacity long-distance overhead line transmission - When the distance is over 700km, HVDC is used instead of AC transmission, which is very competitive. BPA system in the U.S., Nelson River transmission system in Canada, Guiguang system in China and many other HVDC projects are all of this type.

1.2 Introduction of HVDC System

The HVDC system includes the converter, the DC transmission line, and the AC part at the converter station. If fault occurs in one part, the reliability of the HVDC transmission system and the safety of the equipment will be influenced.

There are usually two converter stations A/B and one power line in the HVDC system. One of the converter stations is rectifier station and the other one is inverter station.

The main equipment in the converter stations are rectifiers and inverters, which can realize the mutual transfer between alternating power and direct power. The DC transmission line is used to transmit DC current and power.

The main components of HVDC system are converters, AC breaker, DC smoothing reactors, harmonic filter, reactive power source, polar and DC transmission line [4].

- Converter - As mentioned earlier, the function of the converter is to realize the transformation between AC and DC, which is composed of Valve bridge and converter transformer. The valve component used in modern HVDC transmission systems is a thyristor with a rated voltage of 3-5 KV and a rated current of approximately 2.5 to 3 kA. In the converters, there are many converter bridges in series or parallel. The converter bridge used in HVDC system uses three-phase bridge circuit. The converter transformer transfer AC power to DC power or DC power to AC power, which connect the AC system and the converters.

- AC circuit breaker - It's is used to troubleshoot transformer faults and keep DC lines out of operating.

- DC smoothing reactors - The inductance of these reactors is up to one Henry. The smoothing reactors can reduce the harmonic DC voltage and DC current. Also, it can reduce the current in converters during the DC line-to-ground faults, and prevent the inverter from commutating faults.

- Harmonic Filter - The converters/inverters produce harmonic voltage and current on each terminal, which may increase the temperature of capacitor. Besides, the communication system might be disturbed at the same time. Hence, filters are equipped at the both terminals of DC transmission line.

- Filter Reactive power source - Due to the absorption of a large amount of passive reactive fluid inside the converter/inverter, it is necessary to provide reactive

power supply near the converter/inverter. Parallel capacitance compensation is usually used to achieve that. In addition, synchronous phase modulation or stationary reactive power compensator can also be used.

- Polar - Using earth as a neutral wire is used in most DC wiring. The grounding conductors should have a big cross-sectional area in order to carry higher currents.
- DC line – It's used to transmit DC current and power, which can be composed of overhead transmission lines or the various types of cables.

1.3 Fault Diagnosis Methods

Fault diagnosis technology is a comprehensive technology, which involves modern cybernetics, signal processing and pattern recognition, fuzzy theory, artificial intelligence, electronic technology, statistical mathematics, and other disciplines [5-9]. Modern fault diagnosis technology has more than 30 years of development history in which a lot of diagnostic methods form. Therefore, it has been applied successfully in Mechanic Engineering, Electrical Engineering, Aviation and many other fields.

1.3.1 Methods Based on Analytical model

This method uses on-line system identification technology to establish a mathematical model of the system. After faults occur in the system, the model will change. So, for the same input, the output will change. Therefore, by observing these changes of the model of the system, fault information can be obtained.

The analytic model of power system fault diagnosis is established by using the action information. The model can be described as an unconstrained 0-1 integer programming

problem, which can complete fault diagnosis when the protection information is incomplete. Analytic model can deeply study the dynamic properties of the essence of the system for real-time diagnosis.

Methods based on the analytical model can be divided into three types.

a. State estimation method - The state of the diagnosed process directly reflects the running state of the system, and fault diagnosis can be carried out by estimating the state of the system and combining with the appropriate model [10]. The basic idea of the state estimation method is that reconstructing the state of the controlled process first, constructing the residual sequence by comparing it with the measurable variable. Then the appropriate model can be constructed, and the fault is detected from the residual sequence by statistical test method at last. Therefore, using this method required the system to be observable or partially observable. The state estimation is usually completed by a variety of state observers or filters.

b. Parameter estimation - In many cases, there is a complex relationship between the model parameters and the actual physical object coefficients. Faults will affect the physical coefficients and then reflects parameters in the model. Since not all physical operating parameters are directly observed, the physical operating parameters can be calculated by estimating the model parameters. So, it's able to achieve fault detection, separation, and estimation. Compared with the method based on state estimation, residual sequence calculating is not necessary for parameter estimation method. What fault detection based on parameter estimation method relies on is statistical characteristics of the parameter change so that the method is easier to locate the fault and estimate the

amplitude of the fault. The least-square method is a famous parameter estimation method.

c. Parity Space method - Parity Space method uses the actual measured value of the input and output to test equivalence between current state and normal state. Then fault information can be obtained.

As industrial systems and devices have been becoming more complex, the use of methods based on the analytic model is greatly limited.

1.3.2 Methods Based on Signal Processing

In recent years, signal processing technology has developed rapidly. Spectrum analysis, time series analysis, correlation functions and other methods are able to extract the values of the feature signals. By analyzing these values, the system is operating in which state can be judged. New fault diagnosis methods are studied due to the complexity of diagnostic systems.

a. When the system runs into an abnormal operating state suddenly, the input and output signal will behave abnormally. Specifically, the signal will behave singularity. Wavelet transform is used to analyze the singular signal to find the extreme point corresponding to the abnormal operating state. Mathematical model of the system being diagnosed is not required in wavelet transform. Less limitation of the input signal and strong anti-interference make wavelet transform do the on-line fault detection. The shortcoming of this method is that there may be a certain amount of time delay during detection. In addition, both wavelet-artificial intelligence method and wavelet packet in wavelet transform make wavelet transform meet the need of fault diagnosis [11-14].

b. Method based on Adaptive Sliding Window - Establishing a model for the diagnosis process is not needed in this method. Its basic idea is that output data in the

sliding window is used to compare with input data to construct the residual sequence. When the system is in a different operating state, the residuals sequence can be greatly different. By making a hypothetical test of the residual sequence in normal operating and faults, which operating state the system is operating in can determined.

1.3.3 Pattern Recognition

Pattern recognition was created in the 1920s and developed into a discipline in the early 1960s. Pattern is a physical description of an object, expressed as information showing time or spatial distribution which is usually referred as a case or sample. A natural state or type associated with a concept or prototype of object is called a category. Samples are usually divided by the different degree of similarity. What pattern recognition actually do is to correctly classify a sample into a its category. Fault seepage is also a problem in pattern classification and identification. The purpose is that identify the state of the system (or equipment) and determine whether the system (or equipment) is faulty. Also, a clear fault type, cause, location, fault and severity, and other related information should be given. Then decisions can be made on the actual situation.

Pattern recognition generally includes two types - statistical recognition method and structure recognition. These methods all come from two procedures – design and realization. Based on the sample obtained from simulation or test, classifier can be designed. Then the classifier is trained and tested, which is the realization.

In general, there are four parts in the Pattern Recognition: Data Acquisition, Data processing, Feature Acquisition, Class Decision. The basic structure is depicted in Fig.1.1.

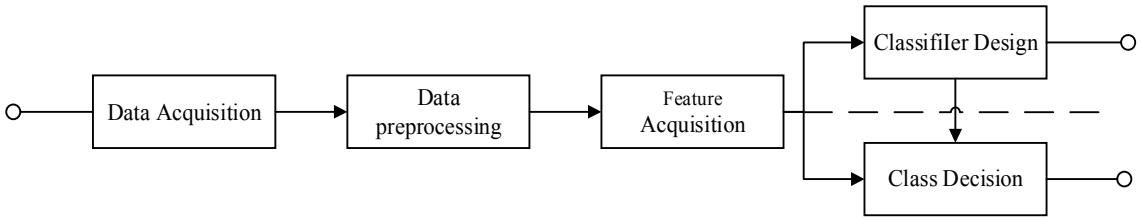

Fig.1.1 Basic Structure of Pattern Recognition

a. Data acquisition - the input waveform or graphics or other types of information is measured, sampled and quantized into matrix or vector representation which can be recognize and calculate by a computer.

b. Preprocessing - Noise is removed or reduced and useful information is enhanced. Error caused by input or measuring instruments or other factors is also recovered in this process.

c. Feature Extraction and Selection - The amount of data obtained from an image or waveform is too large to input to the designed model. In order to realize classification recognition effectively, it is necessary to transform the original data to get the characteristics that best reflect the essence of classification.

d. Classification Decision: Samples are classified into categories in the feature space by using statistical methods. The basic procedure is that a decision rule based on the sample training set is determined first, so that the classification of the identified object by such this rule has the least wrong recognition rate or the least loss. The pattern recognition can be understood as the transformation of data in different spaces. The flow chart of the data transformation is depicted in Fig.1.2.

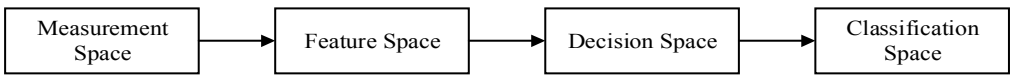

Fig.1.2 Understanding of Fault diagnosis Process

The measurement space is the space composed of the original data. The feature space is the space in which the original data is processed to obtain the feature for classification recognition. To improve diagnostic accuracy and decrease the number of class, data can be mapped into the feature space. Decision space and classification space are used for decision classification, and in most cases the two spaces are the same. Compared with the model-based method, establishing a model is not needed in pattern recognition. The only thing needed is representative training data, so the pattern recognition method is simple for calculating and flexible in analysis.

The commonly used methods of statistical pattern recognition are listed below.

a. Data Clustering: The goal of data clustering is to organize data into meaningful and useful sets using some similarity measurement methods[15]. In order to evaluate the similarity of internal data in the exploration stage of pattern recognition research, data clustering is necessary to extract meaningful data from a bunch of information that is not grouped. The most classic clustering algorithms is K-means[16].

b. Bayesian classification: Bayesian decision rules are obtained by using the Bayesian formula in probability theory. These rules are used to solve the problem of pattern classification, called Bayesian classification. Common decision rules used are the Minimum-error-rate Bayes classification and the Minimum-risk Bayes classification. In Bayesian classification, probability distribution of each class should be known, and number of classes of decision classification should be a certain number. The classic Naive Bayesian Classification Algorithm is improved to solve the complicate problems.

To achieve a real-time classification task and the accuracy of classification, the Bayesian Classification is improved by considering posteriori probability estimation and weight adjusting[17]. By using multiple linear regression model, the weight is calculated based on the correlation between attributes so the Naïve Bayesian classification algorithm is more accurate[18].

c. Decision Tree - Decision Tree, also known as multilevel classifier, is an effective method to classify in pattern recognition, especially multi-classification problems[19]. Decision Tree structure are rule-based and data-based[20]. The tree classifier can be used to transform a complex multi-category classification problem into a number of simple classification problems and solve. Samples are classified multiple categories not using one algorithm or one decision rule, but using multi-level to solve the problem gradually.

d. Fuzzy Pattern Recognition - In the 1965, Zadeh put forward the famous fuzzy set theory, and created fuzzy mathematics according to this theory[21-23]. Fuzzy pattern recognition is that using fuzzy ideas and methods in solving pattern recognition problems. Due to the lack of clarity and uncertainty in some of the states in the production process, such as the description - the voltage is "too high" or the vibration "severe" and so on.

e. Neural Network: NN (Neural Networks) is a very rapid development of marginal disciplines in the past more than 10 years. A large number of neurons are used in the network to simulate the learning process in human brain. By training the neural network, it can be used to solve several specific problems by using the experience in the

training. Nowadays, it's promoted in fault diagnosis combined with the existing signal processing, expert system and fuzzy technology, etc.[24-30].

f. Compared with the model-based fault diagnosis method and expert system, neural network shows outstanding advantages in fault diagnosis. However, at the same time, there are several shortcomings.

- High Requirements on training samples – Neural network has a strict requirement on the quantity of samples, which means that the distribution of sample data to be able to cover the failure states. Therefore, the number of samples is bound to be large. Besides, the quality of the sample is required to be high, which means that training sample must contain as many fault modes as possible. Also, there should be contradictions or conflicts between similar fault samples. However, in the most cases, these requirements are difficult to meet. The actual risk of learning machine includes two parts: the experience risk (training error) and confidence range, which is determined by the VC dimension and the number of the training samples.

- If training dataset is poor, the larger VC dimension of the model, the greater the confidence range and the greater the possible difference. This is the reason why over learning occurs, resulting in poor generalization performance of the fault diagnosis model. When a fault diagnosis model is established, the empirical risk should be minimized, the VC dimension should be also minimized to narrow the confidence range to achieve a smaller actual risk

- It's hard to optimize the structure and parameters. To make fault diagnosis

model more accurate, traditional fault diagnosis method based on data learning has a "trial" process for the selection of fault diagnosis model and algorithm, which can be understood as the process of adjusting the confidence range. If one of the models is suitable for the samples, better results can be achieved. However, because of the lack of theoretical guidance, this choice can only rely on prior knowledge and skills.

1.3.4 Support Vector Machines

The traditional statistical pattern recognition methods are studied under the premise that the number of samples is enough. In some cases, especially in the fault diagnosis, the number of samples is often limited. In these cases, the traditional pattern recognition method is difficult to achieve an acceptable effect. Since 1960s, researchers represented by Vapnik began to study machine learning in the cases that have limited samples. By the middle of the 90, Statistical Learning Theory (SLT) was created, and a new pattern recognition method - Support Vector Machine (SVM), was developed based on this theory. It is considered to be the best theory for the research of statistical estimation and prediction with small samples.

With statistical learning theory as the background, SVM has a higher theory and math background. Compared with other methods such as neural networks, SVM has the following characteristics[31-33]:

a. In SVM method, minimized risk is the guarantee of the generalization of the SVM.

b. By using the theory of kernel function, nonlinear data can be transferred to linear data by using the high dimensional mapping. Then the function is set up in the high dimensional space. Hence, solution of nonlinear is simplified, and the computational

complexity does not increase.

c. The SVM algorithm is finally transformed into a convex optimization problem. After that, global optimal solution is obtained theoretically, which avoids the local optimal problem in neural network.

d. The VC dimension is used to solve the problem that the complexity of the algorithm and the dimension of input vector are closely related.

e. SVM has lower requirement on the size of data so that it can effectively deal with the tasks in which there are only small samples.

f. can effectively solve the problem with small samples learning.

1.4 Structure of Thesis

In Chapter 1, it's mainly introduced that the HVDC background and the basic information of algorithm applied in fault diagnosis. In Chapter 2, the HVDC model and common faults in AC system and DC system in HVDC are introduced first. Ten operating states and seven fault features are selected. Then the waveform of all 10 operating states are depicted and analyzed. In Chapter 3, machine learning based methods are introduced. In addition, results of fault diagnosis and accuracy is depicted. In Chapter Conclusion, whole results of this thesis are recapitulated.

2 Faults in HVDC System and Simulation

2.1 HVDC System Model and Simulation Settings

The model of HVDC system used is a 500kV/2kA system depicted in Fig.1, so the maximum transmission power is 1000 MW. The frequency is 50 Hz. 12-pulse bridge is used in the converters. In this system, a 60 Hz AC system with 500kV, 5000MVA and a 50 Hz AC system with 345kV, 10000 MVA are connected by a DC transmission line. The length of the DC transmission line is 300 km. There are two 0.5 H reactors at the both terminals of the distributed parameter line. There are 9 breakers in the HVDC system to simulate 9 faults. Three breakers are set in the middle of the DC transmission line to simulate the impact the grounding location has on system.

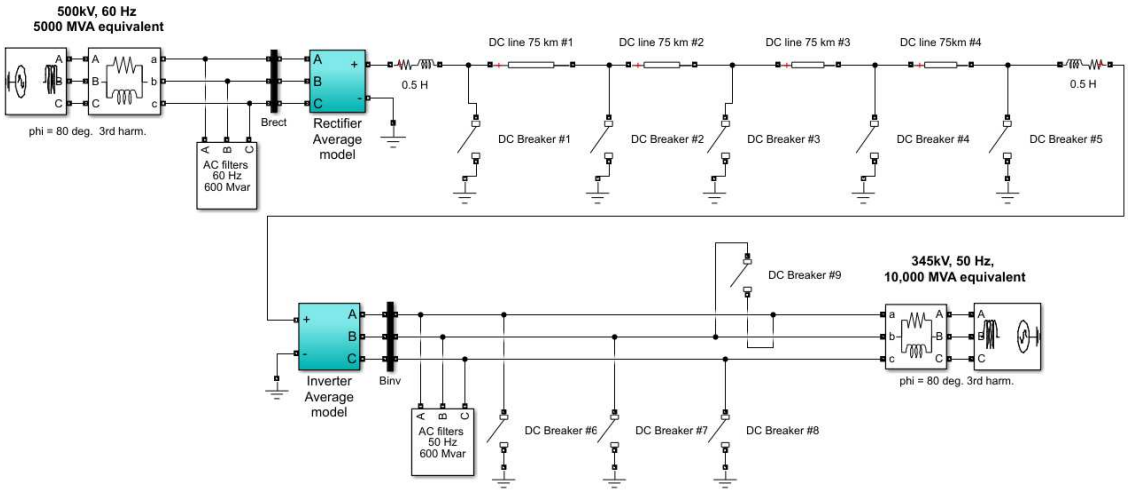

Fig.2.1 HVDC System Model

Distributed resistance (Ohms/km) is 0.015. Distributed inductance (H/km) is 0.792e-3. Distributed capacitance (F/km) is 14.4e-9. The tap ratio (N primary / N secondary) is 0.9 and 0.96 at rectifier and inverter. the Mvar can be supplied by capacitors because of no harmonic generation in this model.

The time step of the simulation is set to 50 us. Whole simulation time is 1.4s while all

faults are set during 0.7s – 0.8s. Therefore, data of 20,000 time points are available to be obtained during the entire simulation.

2.2 Faults in HVDC System

HVDC system includes the AC part of converter, HVDC transmission line and converter station, in which the converter is the most important part. Converter realizes the interchange of AC and DC. Because the converter as the main component, the common faults of HVDC system includes internal faults and external faults. Internal fault refers to the failure caused by converter itself, and the external fault includes the faults of the DC line and the AC system. The causes of these two faults and their impact on the system are analyzed separately below[34-42].

a. Internal Fault of Converter

The converter includes a rectifier and an inverter, and the types of failures are substantially the same.

- Common Faults and Effects in Rectifier
  - Mis-opening – it will occur on the bridge arm of the rectifier bridge. Valve in the most of blocking period withstand the forward voltage. If the effect of excessive forward voltage rising, or valve control pole causing circuit fault, the valve bridge mis-opening may cause. There is less chance that the opening circuit fault occurs in rectifier. Even if it occurs, it is same as it open early, which is little disturbance to normal operating.
  - Not-opening – it will occur on the bridge arm, because of the loss of the trigger pulse or the valve pole control circuit fault. This fault will cause the DC

voltage dropping.

- Component Fault - When faults of the valve components occur, the voltage of other normal will increase.
- Short-circuit in Bridge Arm - the internal or external insulation of the inverter bridge arm is damaged or short-circuited. A short circuit in the bridge arm causes an increase in the AC current and a decrease in the DC current.
- Short-circuit in DC Bus: The short-circuit faults occur between DC outgoing line and reactor in DC line, which is a relatively serious rectifier fault. When this type of failure occurs, the AC current surges and the DC voltage drops to zero.

- Common Faults and Effects in Inverter

Mis-opening, not-opening, component fault and bridge arm short-circuit fault will also occur in inverter. Fault location, fault cause and influence are the same as the rectifier. However, unlike rectifiers, the most common failure of the inverter is a phase change failure. When the Commutation between the two bridge arms is over, the valve just exiting the guide fails to restore the blocking capacity under the effort of reverse voltage, or commutation isn't completed during reverse voltage period, the valve voltage will be positive. The rear trigger valve will be scheduled to exit the guide valve switching phase, which is called phase change fault. This fault will cause the DC voltage dropping, the DC current increasing, AC voltage dropping, trigger forward angle being too small and fixed shutdown angle being too small may cause phase change fault.

b. External Fault of Converter

The faults at rectifier terminal and AC system terminal are of same type, mainly including two phase-to-ground faults and phase-to-phase faults.

- Phase-to-ground Faults

Phase-to-ground faults includes single-phase grounding, double-phase grounding, and triple-phase grounding.

  - Single-phase Grounding - grounding in any one of three phases will lead to voltage of the short-circuit phase to dropping. DC voltage and DC current will decrease correspondingly. Non-characteristic harmonics will increase[42].
  - Double-phase Grounding – voltage of fault phase will drop when grounding fault occur in any two of the three AC phases. The DC voltage and DC current will decrease correspondingly, and the non-characteristic harmonic will increase.
  - Triple-phase Grounding: When the three AC phases are grounded, the AC voltage is symmetrically reduced, and the DC voltage and DC current drop accordingly. Also, the non-characteristic harmonics increase.

- Phase-to-phase Faults

The Interphase fault includes two types of faults which are triple-phase-to-phase short-circuit fault and double-phase-to-phase short-circuit fault. When the Phase-to-phase faults occurs in the AC system in inverter terminal, the decrease of AC voltage may cause the commutation fault.

Whether the above various phase-to-phase short-circuit faults occur on the rectifier

terminal or the inverter terminal, the impact reflected on the DC voltage and DC current are the same so that it's hard to distinguish. However, it is easy to distinguish short-circuit fault occur in which side and which phase by the three-phase AC voltage waveform on both terminals.

The faults on the DC line are mainly Polar-grounding short-circuit faults. Polar faults occur only in cases where there is considerable physical damage that causes bipolar conductors to touch each other, so this type of fault is not common. Short-circuit faults in DC Line will leads to the DC current surging and voltage over-limitation. It is also a more serious fault.

2.3 Fault Waveform and Feature Extraction

In this thesis, ten representative states are considered, including normal, DC line-to-ground fault at rectifier terminal, DC line-to-ground fault at inverter terminal, single-phase grounding, double-phases grounding, triple-phases grounding, short-circuit between two phases and DC Line-to-Ground Faults which are divided to grounding at 75km, grounding at 150km, grounding at 225km according the fault location.

When simulating single-phase grounding and double-phases grounding faults, choosing either phase (A, B, C) or any two phases (A and B, B and C, C and A) of the AC three-phase will have the same impact on DC voltage and DC current. It is easy to distinguish which phase fault occur in by AC waveform. Therefore, phase A, phase A/B and phase A/B are chosen as fault phases in single-phase grounding, double-phases grounding and short-circuit between two phases fault simulation.

Whole simulation time is 1.4s while all faults are set during 0.7s – 0.8s.

The faults feature selected are DC voltage V_{dcr} at rectifier terminal, DC current I_{dcr} at rectifier terminal, DC voltage V_{dci} at inverter terminal, DC current I_{dci} at inverter terminal, AC current I_A in phase-A, AC current I_B in phase-B and AC current I_C in phase-C. AC voltage V_A, V_B and V_C are used to show which phase the grounding fault occurs in.

2.3.1 Normal Operating

As depicted in Fig.2.2 and Fig.2.3, when HVDC system operates in normal state, DC current I_{dcr} at rectifier terminal is the same as DC current I_{dci} at inverter terminal, and DC voltage V_{dcr} at rectifier terminal equals to DC voltage V_{dci} at inverter terminal because of no grounding point in DC line.

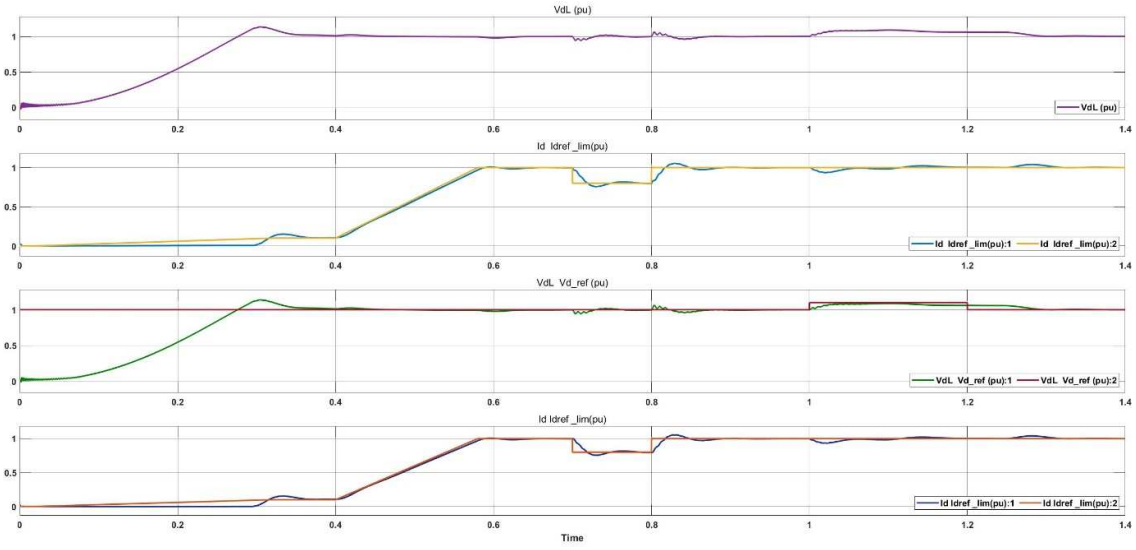

Fig.2.2 DC Voltage and Current at Rectifier and Inverter Terminals in Normal Operating

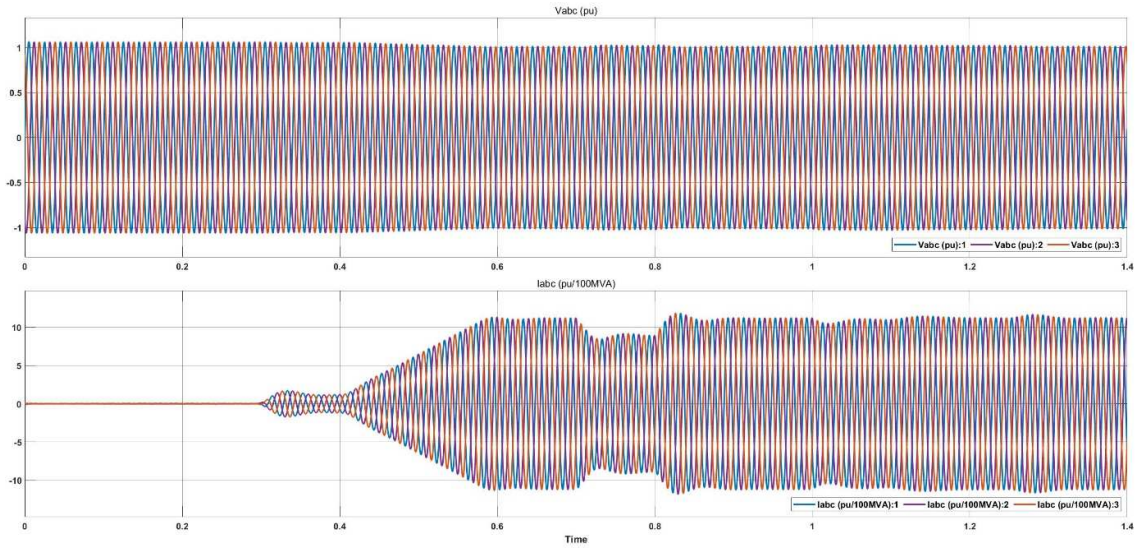

Fig.2.3 AC Three-phase Voltage and Current in Normal Operating

2.3.2 DC Line-to-ground Fault at Rectifier Terminal

As depicted in Fig.2.4 and Fig.2.5, when HVDC system operates in DC Line-to-ground Fault at Rectifier Terminal fault, DC current I_{dcr} at rectifier terminal is the same as DC current I_{dci} at inverter terminal. When grounding fault occurs at rectifier terminal at 0.7s, DC current at inverter terminal and AC current will drop to 0V. DC current surges at the same time.

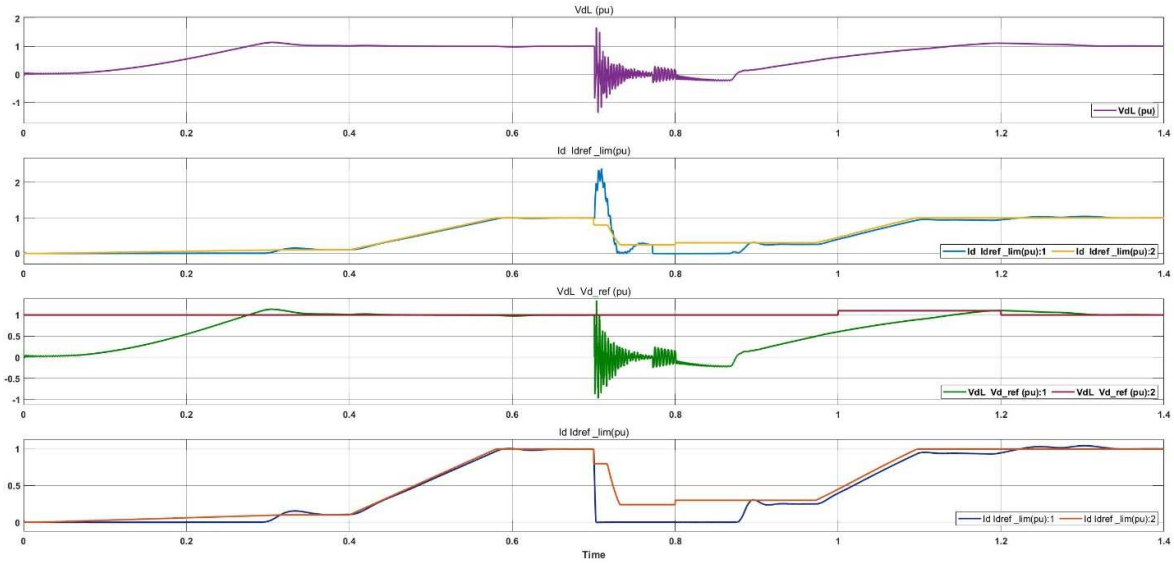

Fig.2.4 DC Voltage and Current at Rectifier and Inverter Terminals in DC Line-to-ground Fault at Rectifier Terminal

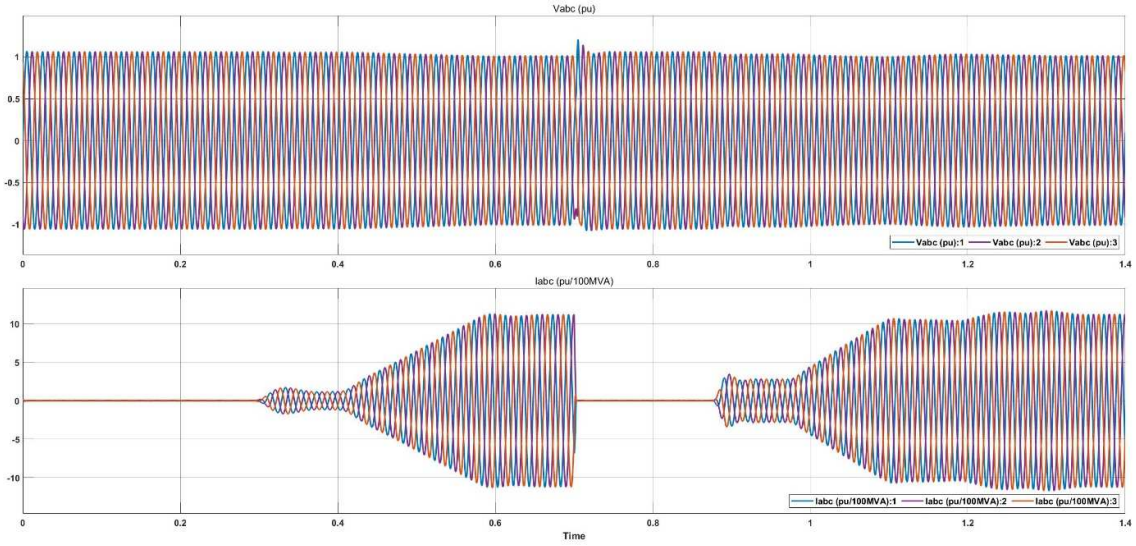

Fig.2.5 AC Three-phase Voltage and Current in DC Line-to-ground Fault at Rectifier Terminal

2.3.3 DC Line-to-ground Fault at Inverter Terminal

As depicted in Fig.2.6 and Fig.2.7, besides the voltage at rectifier, other DC features are similar to the features when DC line-to-ground fault occurs at rectifier terminal. Few oscillations occur in the DC current I_{dci} at inverter terminal after the fault occurs. The waveforms belong to AC system are the same as waveforms above.

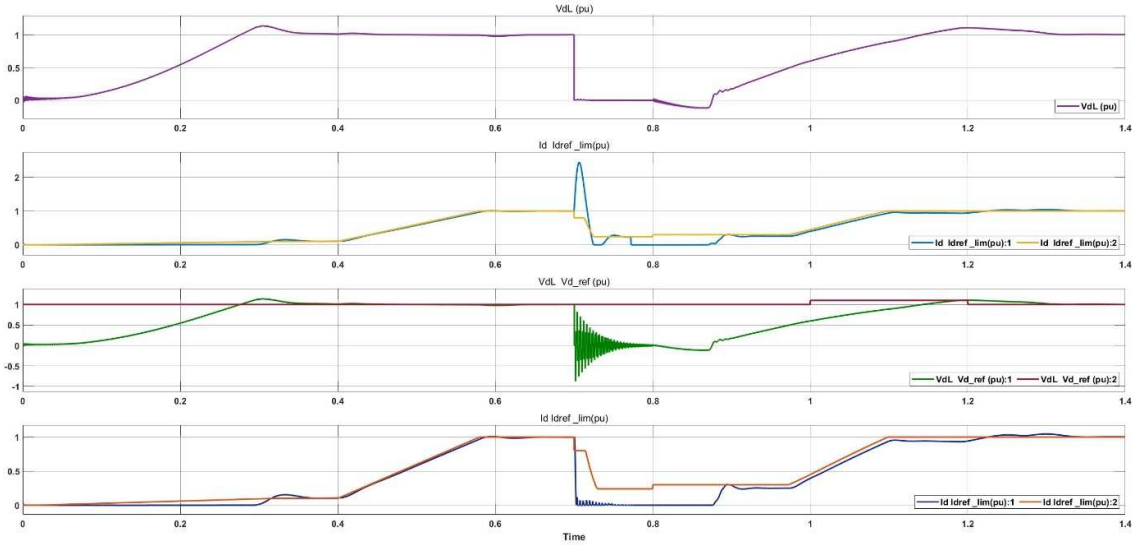

Fig.2.6 DC Voltage and Current at Rectifier and Inverter Terminals in DC Line-to-ground Fault at Inverter Terminal

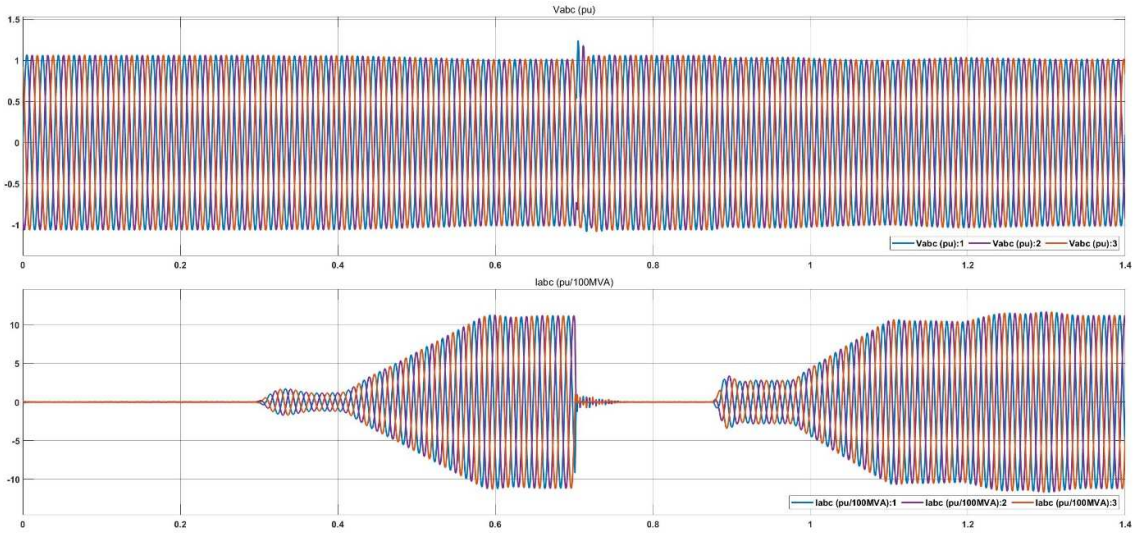

Fig.2.7 AC Three-phase Voltage and Current in DC Line-to-ground Fault at Inverter Terminal

2.3.4 Single-phase Grounding Fault

As depicted in Fig.2.8 and Fig.2.9, all the feature in DC system - DC voltage Vdcr at rectifier terminal, DC current Idcr at rectifier terminal, DC voltage Vdci at inverter terminal, DC current Idci at inverter terminal decrease slightly after the fault occurs. After phase A is grounded, voltage of phase A V_A drops to 0.

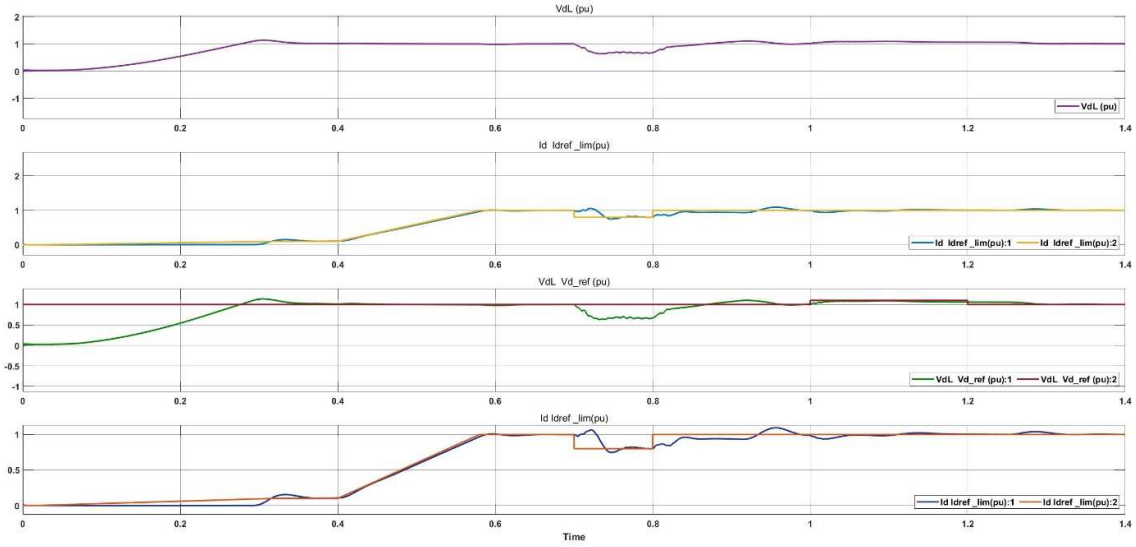

Fig.2.8 DC Voltage and Current at Rectifier and Inverter Terminals in Single-phase Grounding Fault

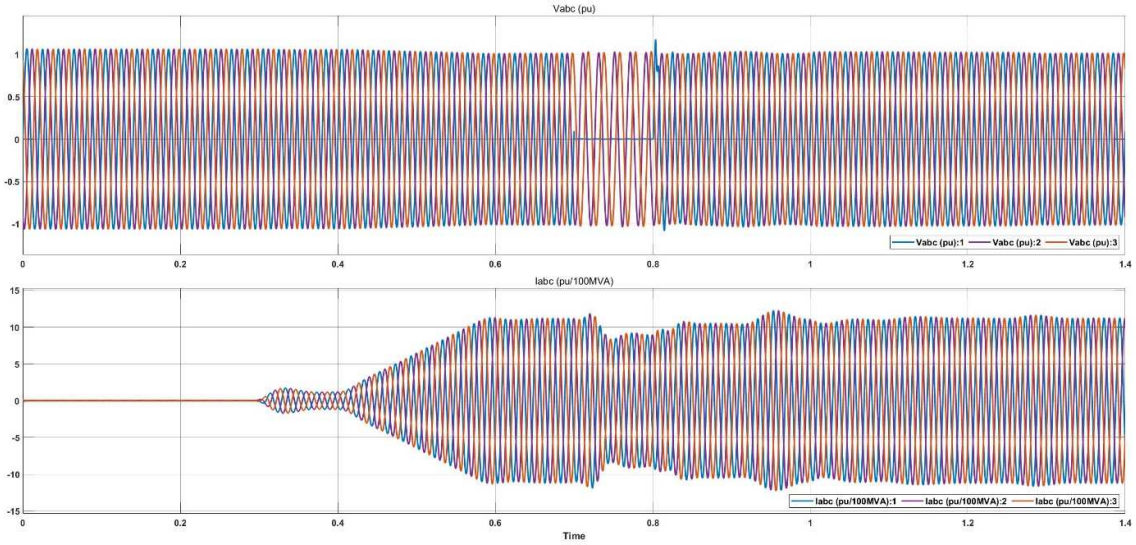

Fig.2.9 AC Three-phase Voltage and Current in Single-phase Grounding Fault

2.3.5 Double-phases Grounding Fault

As depicted in Fig.2.10 and Fig.2.11, DC voltage V_{dcr} at rectifier terminal and DC voltage V_{dci} at inverter terminal decrease significantly. DC current I_{dcr} at rectifier terminal and DC current I_{dci} at inverter terminal drops sharply. After phase A and phase B are grounded, V_A and V_B drop to 0.

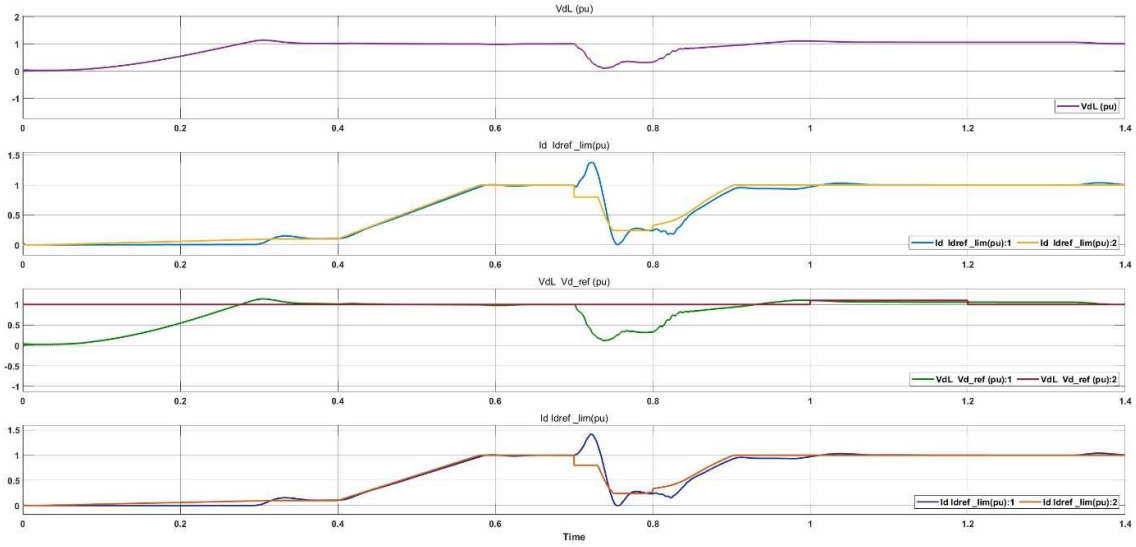

Fig.2.10 DC Voltage and Current at Rectifier and Inverter Terminals in Double-phases Grounding Fault

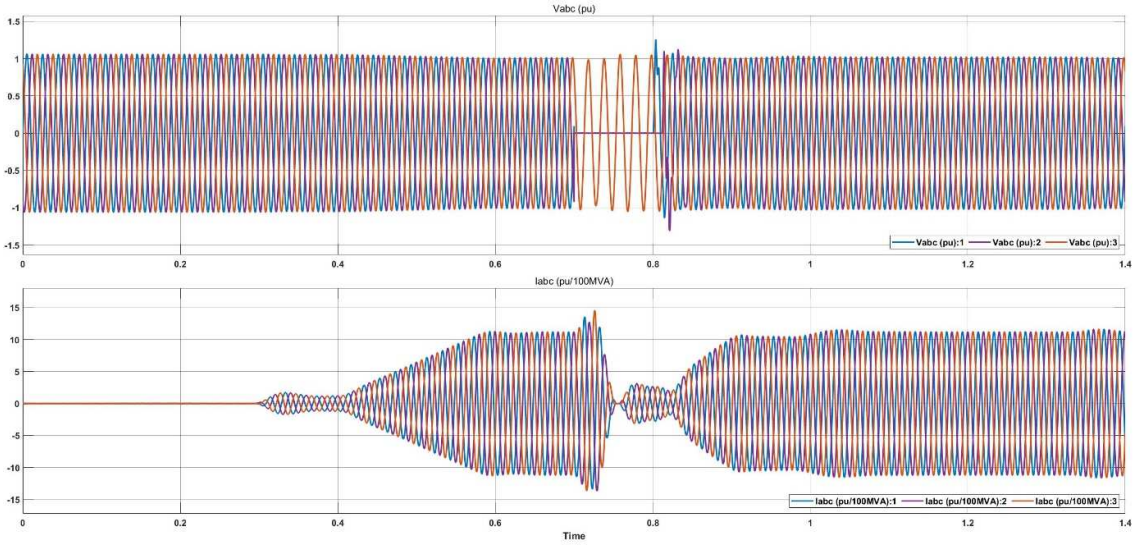

Fig.2.11 AC Three-phase Voltage and Current in Double-phases Grounding Fault

2.3.6 Triple-phases Grounding Fault

As depicted in Fig.2.12 and Fig.2.13, DC voltage V_{dcr} at rectifier terminal and DC voltage V_{dci} at inverter terminal decrease significantly to nearly -0.5pu. DC current I_{dcr} at rectifier terminal and DC current I_{dci} at inverter terminal drops sharply to 0pu. After phase A, phase B and phase C are grounded, V_A, V_B and V_C drop to 0.

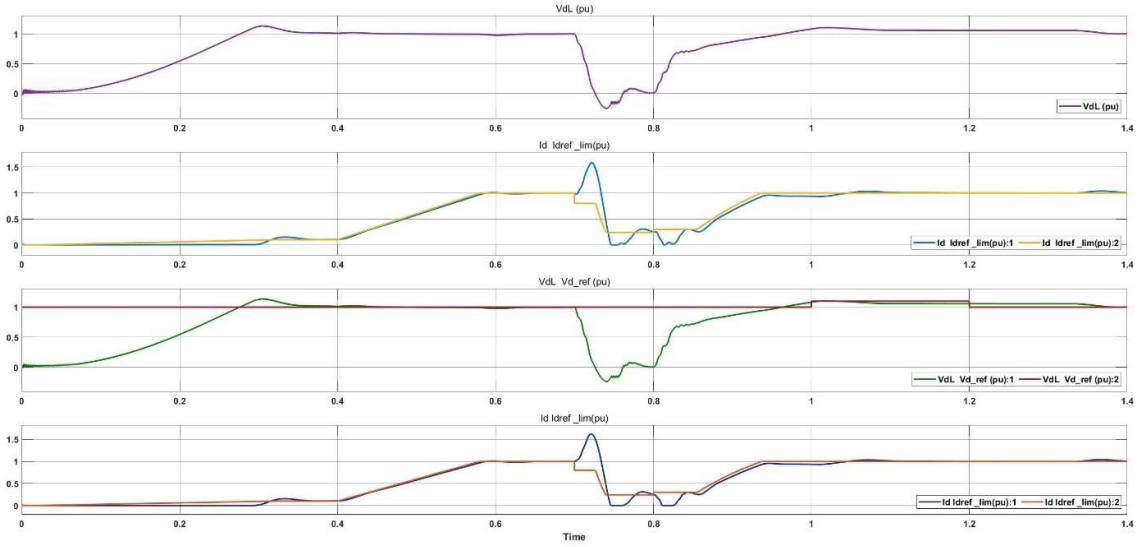

Fig.2.12 DC Voltage and Current at Rectifier and Inverter Terminals in Triple-phases Grounding Fault

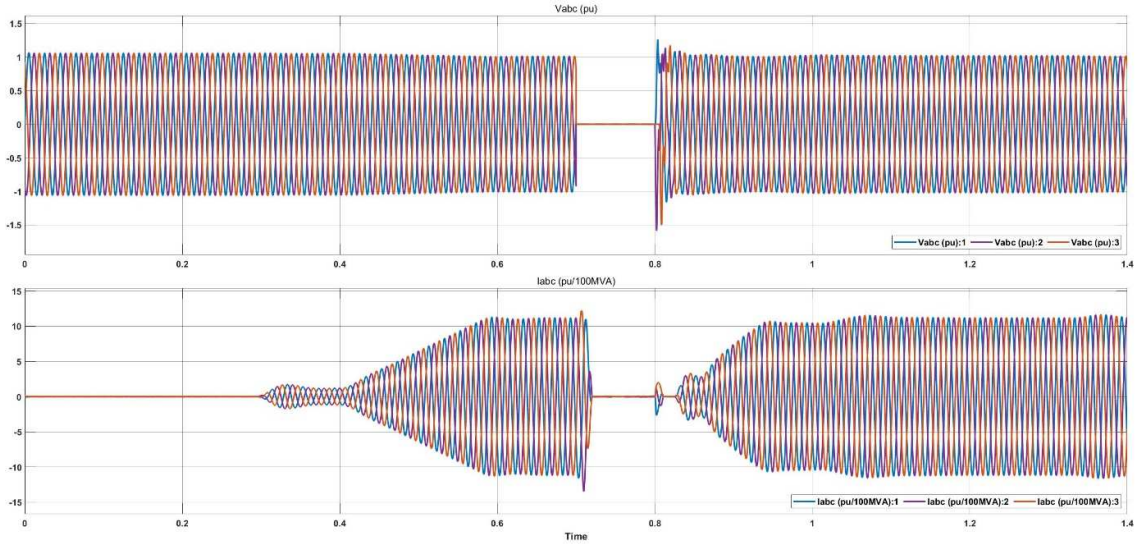

Fig.2.13 AC Three-phase Voltage and Current in Triple-phases Grounding Fault

2.3.7 Short-circuit between Two Phases in AC System

As depicted in Fig.2.14 and Fig.2.15, DC voltage V_{dcr} at rectifier terminal and DC voltage V_{dci} at inverter terminal slowly rise after a sharp decrease to 0pu. There is overvoltage in DC current I_{dcr} at rectifier terminal and DC current I_{dci} at inverter terminal, rising first and then decreasing to 0. After phase A and phase B are connected, V_A, V_B reduce by half.

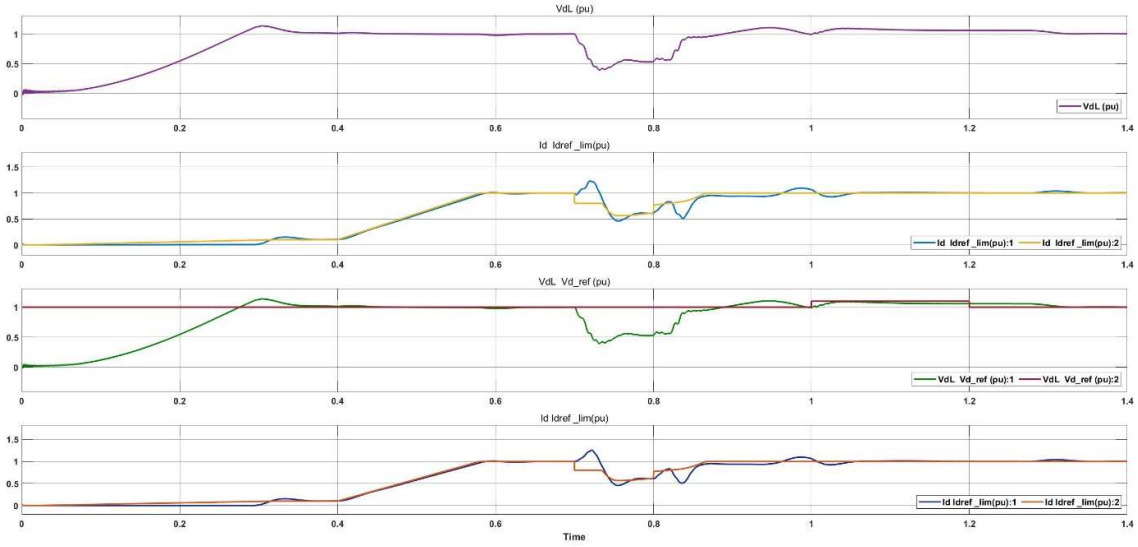

Fig.2.14 DC Voltage and Current at Rectifier and Inverter Terminals in Short-circuit between Two Phases in AC System

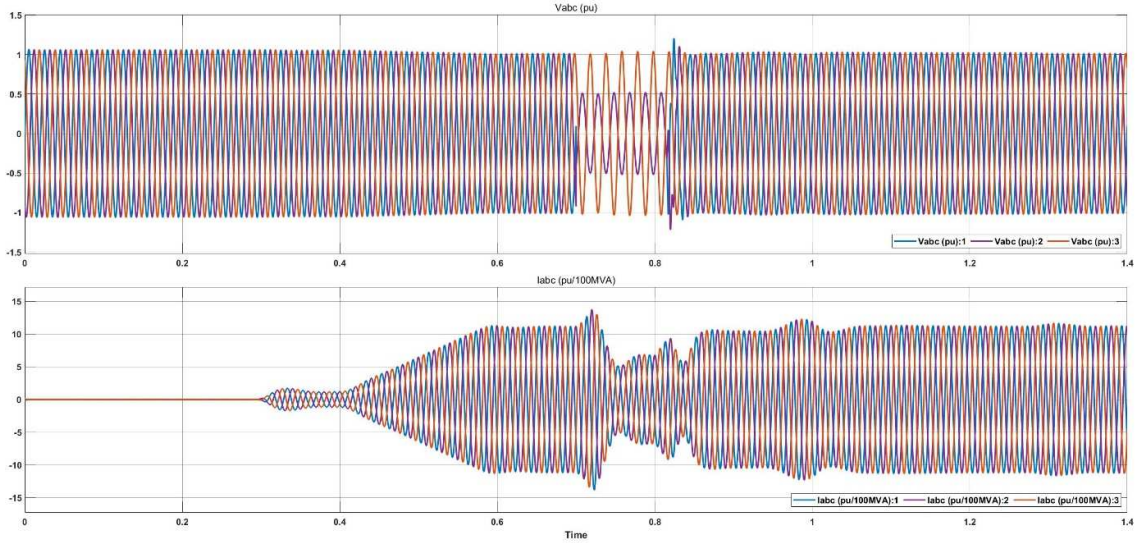

Fig.2.15 AC Three-phase Voltage and Current in Short-circuit between Two Phases in AC System

2.3.8 Line-to-Ground Faults at Middle of DC Transmission Line

As depicted in Fig.2.16 - Fig.2.21,different fault locations have little impact on the feature, depicted as figures. After phase A, phase B and phase C are grounded, V_A, V_B and V_C drop to 0. From the naked eye, the difference between fault waveform of different locations is very small. Distinguishing these DC line grounding fault at different locations is going to be a tough challenge for classifiers.

a. Grounding at 75km from Rectifier Terminal

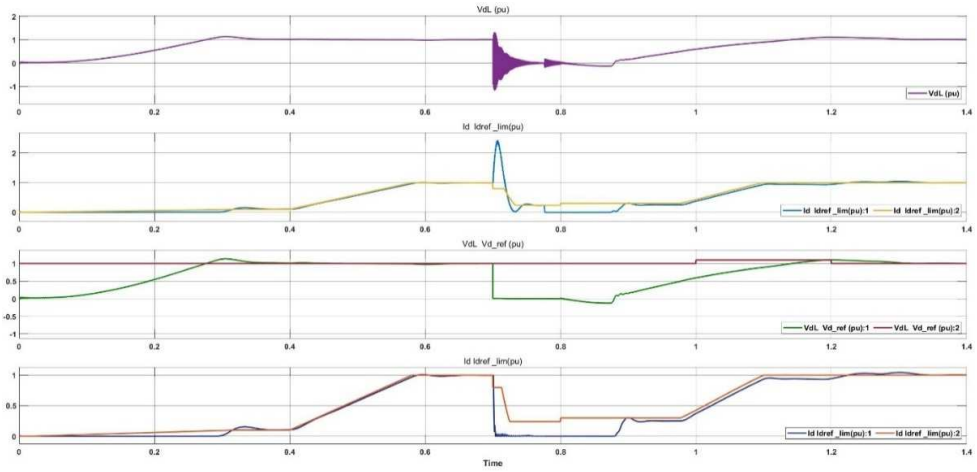

Fig.2.16 DC Voltage and Current at Rectifier and Inverter Terminals in Line-to-Ground Faults at 75km from Rectifier Terminal

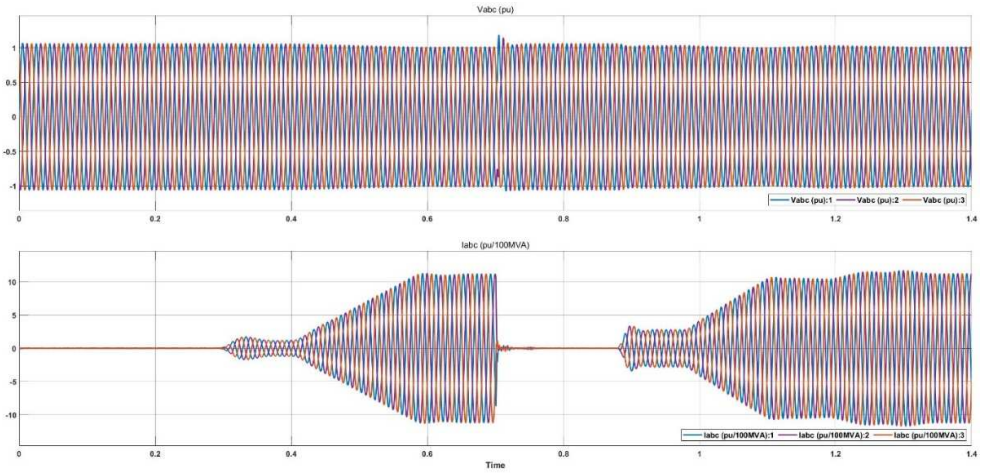

Fig.2.17 AC Three-phase Voltage and Current in Line-to-Ground Faults at 75km from Rectifier Terminal

b. Grounding at 150km from Rectifier Terminal

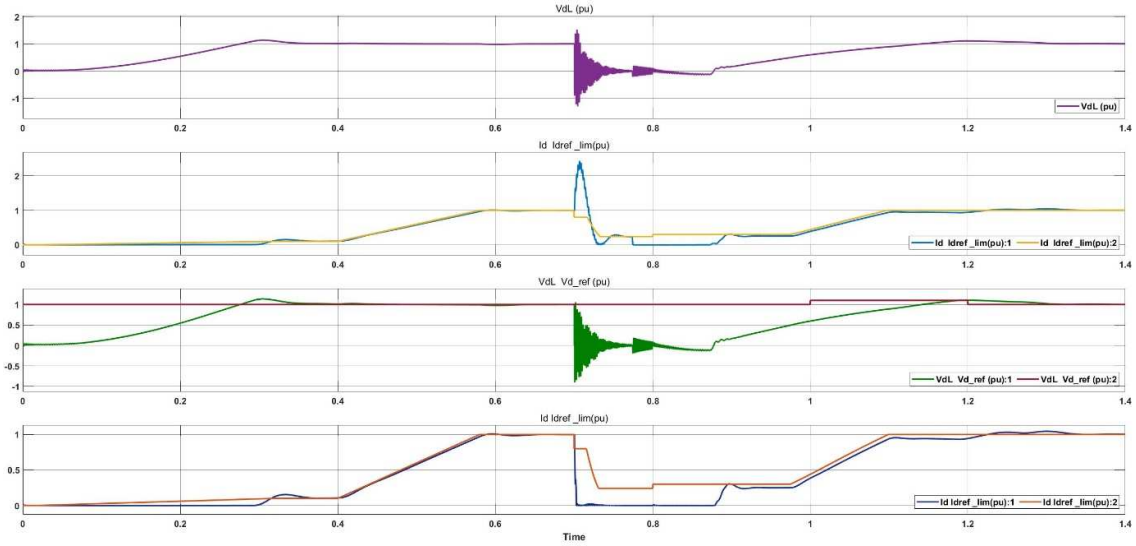

Fig.2.18 DC Voltage and Current at Rectifier and Inverter Terminals in Line-to-Ground Faults at 150km from Rectifier Terminal

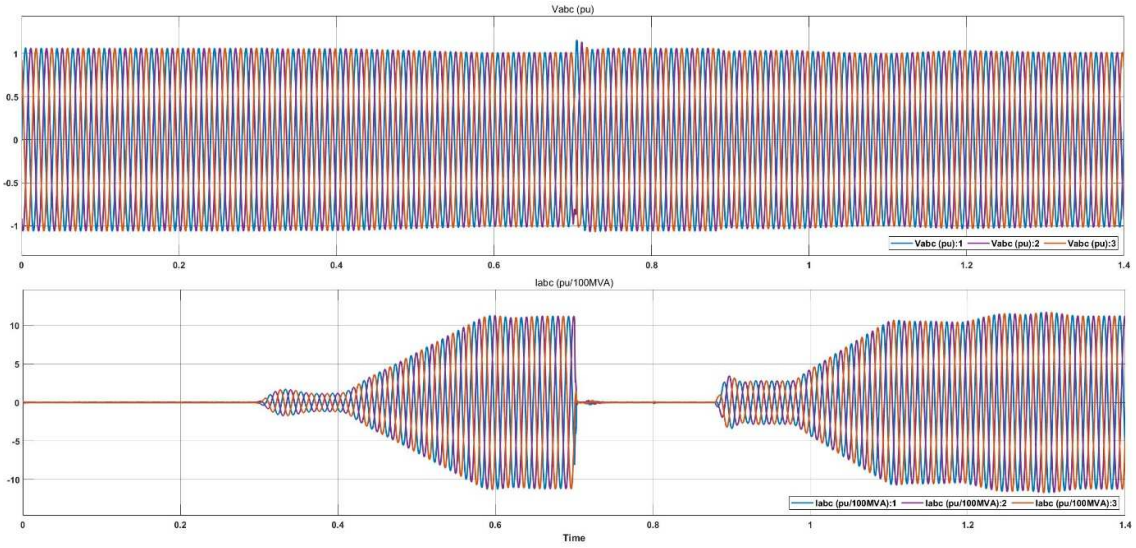

Fig.2.19 AC Three-phase Voltage and Current in Line-to-Ground Faults at 150km from Rectifier Terminal

c. Grounding at 225km from Rectifier Terminal

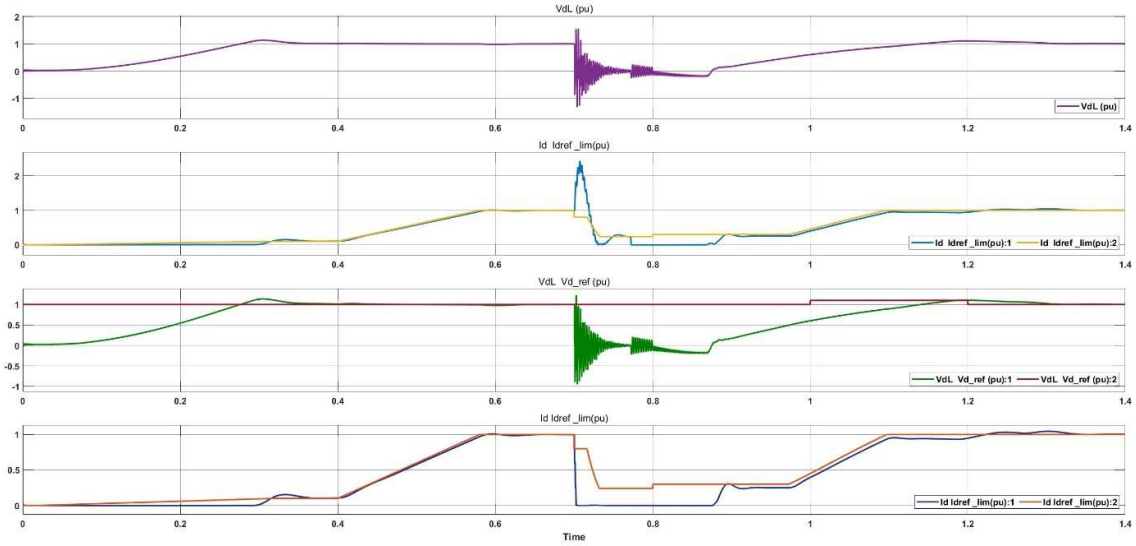

Fig.2.20 DC Voltage and Current at Rectifier and Inverter Terminals in Line-to-Ground Faults at 225km from Rectifier Terminal

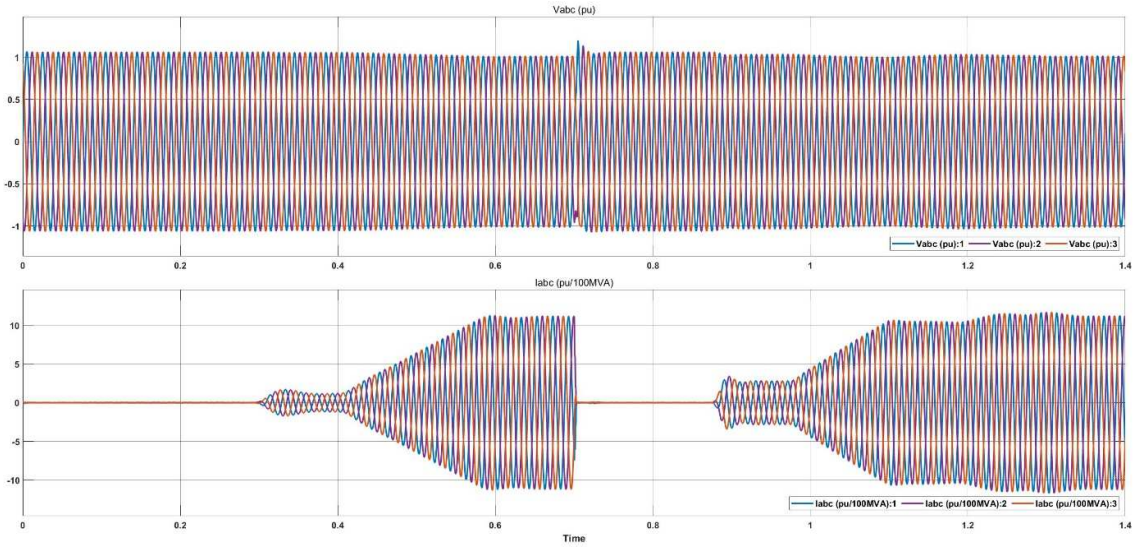

Fig.2.21 AC Three-phase Voltage and Current in Line-to-Ground Faults at 225km from Rectifier Terminal

3 Faults Diagnosis of HVDC System

3.1 Decision Trees

Decision trees are often used to deal with classification. As the name, a tree structure that arranges samples from the root node to the leaf node, which can classify the samples. Each leaf node is one classification to which the sample belongs.

The classification process starts with the root node. The direction of the next move is determined according to the value of sample and node the leaf node is reached.

General process of the decision tree:

a. Data Collection

b. Data Preparing - The tree construction algorithm is only applicable to the nominal data, so the numerical data must be discretized.

c. Data Analysis - The construction figure of the tree should be checked whether it's appropriate.

d. Training - Structure of the tree is constructed.

e. Testing - the error rate is calculated by using experience tree.

The more widely used decision tree algorithm is ID3 algorithm and C4.5 algorithm, both of which use top-to-bottom greedy search.

a) ID3

ID3 algorithm is based on greedy algorithm, originated from the concept learning system. The decline of information is chosen as the standard of test properties. In each node, the property which has not been used and has the highest information gain is selected as classification criteria. This process is repeated until the generated decision tree can perfectly

classify the training examples.

The purpose of dividing a dataset is to change data from an unordered state to an ordered state. Information gain is used to judge whether the model is the best in models using variety data dividing rules. Information gain is the gain of information obtained after each attribute partition after dividing the dataset. When the information gain is biggest, the division result is the best. entropy is the information expectation, the equation is:

$$H = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \quad (1)$$

$P(x_i)$ is the probability of selecting classification i. n is the number of all classification attributes.

Then, the information gain of a property A relative to Dataset S can be calculated using equation(2).

$$Gain(S, A) \equiv \mathrm{H(S)} - \sum_{veValue(A)} \frac{|S_v|}{|S|} H(S_v) \quad (2)$$

$Value(A)$ is a collection of all possible values in attribute A. $S(v)$ is a subset of attribute A of which value is V in collection S.

According to the different attribute division, different information gain can be obtained, and then the dataset is divided according to the attribute with the greatest information gain.

By iterating this process, when the tree can fully classify the samples, or all properties are used, a complete tree is constructed.

b) C4.5

Compared with ID3, C4.5 is another edition of ID3. The different between these two methods is that C4.5 uses information gain rate to select attributes. To prevent overfitting, the decision is pruned. C4.5 can be used to process non-discrete data and incomplete data.

Spilt information can be calculated by using equation (3):

$$SplitInfo(S,A) = -\sum_{j=1}^{v} \frac{|S_j|}{|S|} log_2 \frac{|S_j|}{|S|} \quad (3)$$

Split information is the information generated when the sample set S is divided into a division of the *v* possible values corresponding to attribute A. Hence, information gain can be calculated by using equation (4):

$$GainRatio(S,A) = \frac{Gain(S,A)}{SplitInfo(S,A)} \quad (4)$$

The confusion matrix is a matrix of n rows and n columns. Every column is one prediction class. So, the sum of the samples in each column is the quantity of samples which are classified in this class. Each row in confusion matrix represents actual category. So, each row is quantity of samples in this class.

3.1.1 Fine Tree

The confusion matrix of the Fine Trees classification result is depicted in Fig.3.1. Fine Trees can distinguish the first 7 operating states commendably. However, it works not well in distinguishing the DC line-to-ground at different location. The accuracy of this model is 87.4%.

Fig.3.1 Confusion Matrix of Fine Trees Model

3.1.2 Medium Tree

The confusion matrix of the Medium Trees classification result is depicted in Fig.3.2. Medium Trees works not well in distinguishing operating state 5,6,8,9 from other states. Especially, 95% samples in operating state 10 are wrongly assigned in other operating states. Therefore, the accuracy of this model is 73.4%.

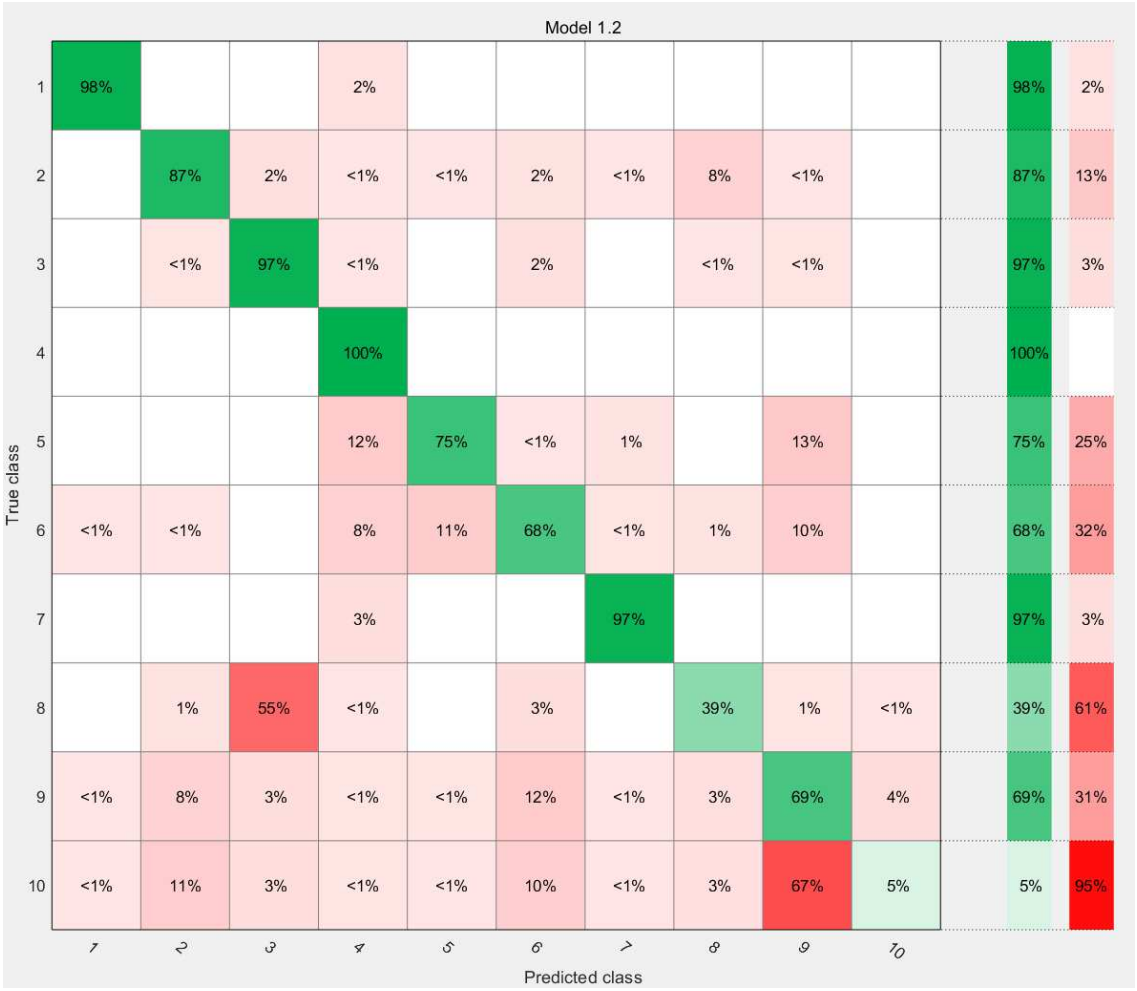

Fig.3.2 Confusion Matrix of Medium Tree Model

3.1.3 Coarse Tree

The confusion matrix of the Coarse Trees classification result is depicted in Fig.3.3. Coarse decision tree works perfectly in states 1,3,4 and it works not well in operating state 7. It fails to distinguish operating states 2,5,6,8,9,10 from others. Therefore, the accuracy of this model is only 38.0%.

Fig.3.3 Confusion Matrix of Coarse Tree Model

3.2 Support Vector Machine

By seeking the minimum structure risk, the generalization ability of SVM can be better. Therefore, experience risk and confidence range can be minimized. So, even if quantity of statistical sample is small, a good statistical rule can be achieved.

a) Linear SVM

If the spatial dimension is not considered, such linear function is collectively referred as a super plane. However, it is clear that the line which can divide samples is not the only one. There are countless lines. Linear SVM is the line which can correctly divide the samples. At the same time, the distance γ from individual classifications is max. The distance γ equals to the heterogeneous support vector projections on linear equation *w.* It can be calculated using following equation (5):

$$\gamma = \frac{(\vec{x}_+ - \vec{x}_-)\cdot\vec{w}^T}{\|\mathrm{W}\|} = \frac{\vec{x}_+\cdot\vec{w}^T - \vec{x}_-\cdot\vec{w}^T}{\|\mathrm{W}\|} \tag{5}$$

$\vec{x}_+$ and $\vec{x}_-$ are positive support vectors and negative support vectors. Since $\vec{x}_+$ and $\vec{x}_-$ meet the function $y_i(w^T x_i + b) = 1$, equation (7) can be obtained.

$$\begin{cases} 1 * (w^T x_+ + b) = 1, y_i = +1 \\ -1 * (w^T x_- + b) = 1, y_i = -1 \end{cases} \tag{6}$$

$$\begin{cases} w^T x_+ = 1 - b \\ w^T x_- = -1 - b \end{cases} \tag{7}$$

Combined with equation (5), equation (6) can be obtained.

$$\gamma = \frac{1-b+(1+b)}{\|w\|} = \frac{2}{\|w\|} \tag{8}$$

In SVM, the distance γ is maximized by using equation (9).

$$\max_{w,b} \frac{2}{\|\mathrm{W}\|}, s.t. y_i(w^T x_i + b) \geq 1 (i = 1,2,\ldots,m) \tag{9}$$

Obviously, when the $\frac{2}{\|\mathrm{W}\|}$ goes to maximum, $\|\mathrm{W}\|$ will go to minimum. So, function can be transferred to equation (10).

$$\min_{w,b} \frac{1}{2} \|W\|^2, s.t. y_i(w^T x_i + b) \geq 1(i = 1,2, \dots, m) \quad (10)$$

Equation (10) is the basic equation of support vector machine.

b) Nonlinear SVM

As for the nonlinear problems, linear support vector machine is not effective. It is necessary to use non-linear models to classify solve these problems, depicted in Fig.3.4.

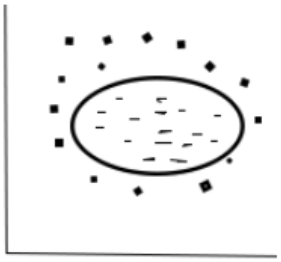

Fig.3.4 Non-linear Sample Space

It is clear that using a straight line does not separate the two types of samples. However, elliptic curve (non-linear model) can perfectly solve nonlinear problems. Solving nonlinear problems is very hard, so if liner SVM solution can be used in solving nonlinear classification problem, it will be much easier. To realize this idea, linear problems should be transferred to linear problems.

The training sample is mapped from the original space to a high-dimension space, so the sample will be divided linearly in this space. If the dimension of original space is limited that is, the attribute is limited, then there must be a high dimensional feature space in which the samples can be divided. Since feature vector after mapping of x is set as $\varphi(x)$, The model of partition of the hyperplane in the feature space can be represented as equation (11)

$$f(x) = w^T \emptyset(x) + b \quad (11)$$

So, the minimum function is:

$$\min_{w,b} \frac{1}{2} \|w\|^2, s.t. y_i(w^T \emptyset(x_i) + b) \geq 1(i = 1,2, \dots, m) \quad (12)$$

The duality problem is:

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \emptyset(x_i)^T \emptyset(x_j) \quad (13)$$

$$s.t. \sum_{i=1}^{m} a_i y_i = 0, a_i \geq 0, i = 1,2, \dots, m \quad (14)$$

In equation (13), $\emptyset(x_i)^T, \emptyset(x_j)$ are very difficult to be calculated directly when solving the equations, because the dimensions of the feature space might be extremely high or even infinite. $\emptyset(x_i)^T, \emptyset(x_j)$ are the inner projections of samples x_i and x_j after these vectors are mapped to feature space. The inner product of samples equal to the results of equation (15).

$$k(x_i, x_j) = < \emptyset(x_i), \emptyset(x_j) > = \emptyset(x_i)^T \emptyset(x_j) \quad (15)$$

Therefore, function (16) can be transferred to

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (16)$$

$$s.t. \sum_{i=1}^{m} a_i y_i = 0, \alpha_i \geq 0, i = 1,2, \dots, m \quad (17)$$

The result of function is equation (18):

$$\begin{aligned} f(x) &= w^T \emptyset(x) + b \\ &= \sum_{i=1}^{m} \alpha_i y_i \emptyset(x_i)^T \emptyset(x_j) + b \\ &= \sum_{i=1}^{m} \alpha_i \, y_i k(x_i, x_j) + b \end{aligned} \quad (18)$$

$k(x_i, x_j)$ is the kernel function. According to the features of the samples and choices of parameters, quantities of kernel functions can be used. The kernel function being used generally are depicted below.

a) Linear Kernel Function:

$$k(x_i, x_j) = x_i^T x_j \quad (19)$$

b) Polynomial Kernel Function:

$$k(x_i, x_j) = (x_i^T x_j)^d \quad (20)$$

In the function, *d* is the idempotent number of a polynomial. When *d* equals to 1,

this function transfers to Linear Kernel Function.

c) Gaussian Kernel Function ($\sigma > 0$)

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{21}$$

d) Laplace Kernel Function ($\sigma > 0$)

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma}\right) \tag{22}$$

d) Sigmoid Kernel Function ($\sigma > 0, \theta > 0$)

$$k(x_i, x_j) = \tanh\left(\beta x_i^T x_j + \theta\right) \tag{23}$$

What's more, kernel functions can be combined as a new kernel function.

3.2.1 Linear SVM

The confusion matrix of the Linear SVM classification result is depicted Fig.3.5. Linear support vector machine works not well in most operating states, especially in operating states 3,8,9,10. The accuracy is only 57.6%. By using SVM in large data classification, simulation time of all method used in this thesis become very long. It has higher requirements on the hardware performance when the size of data is large.

Fig.3.5 Confusion Matrix of Linear SVM Model

3.2.2 Quadratic SVM

The confusion matrix of the Quadratic SVM classification result is depicted in Fig.3.6. Quadratic SVM works not well in distinguishing operating state 3,8,10 from other states. Especially, 83% samples in operating state 9 are wrongly assigned in other operating states. Therefore, the accuracy of this model is 78.1%.

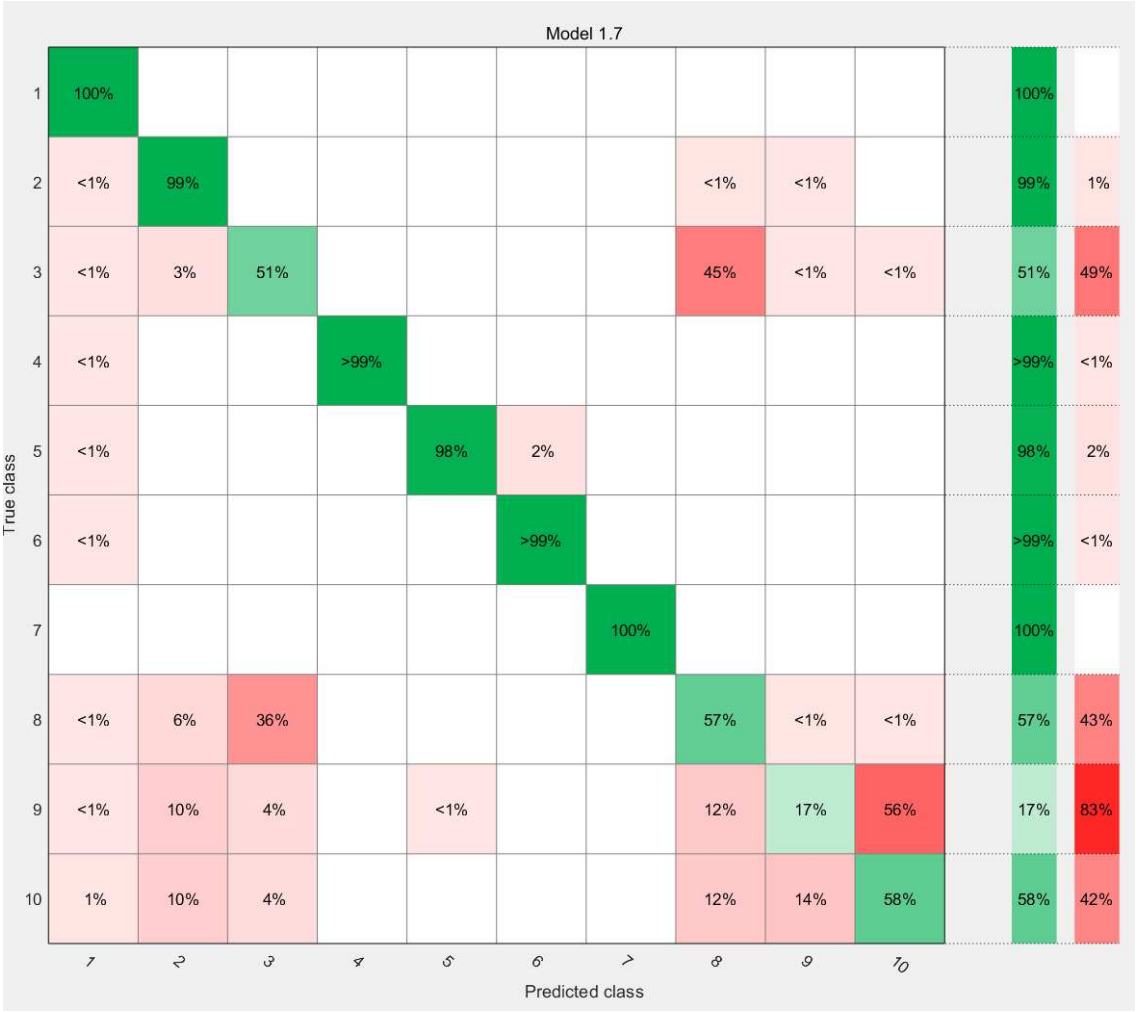

Fig.3.6 Confusion Matrix of Quadratic SVM Model

3.2.3 Cubic SVM

The confusion matrix of the Cubic SVM classification result is depicted Fig.3.7. Cubic SVM works not well in distinguishing operating state 2,8,10 from other states. Especially, classifier works worse in operating states 3,9. Therefore, the accuracy of this model is 72.8%.

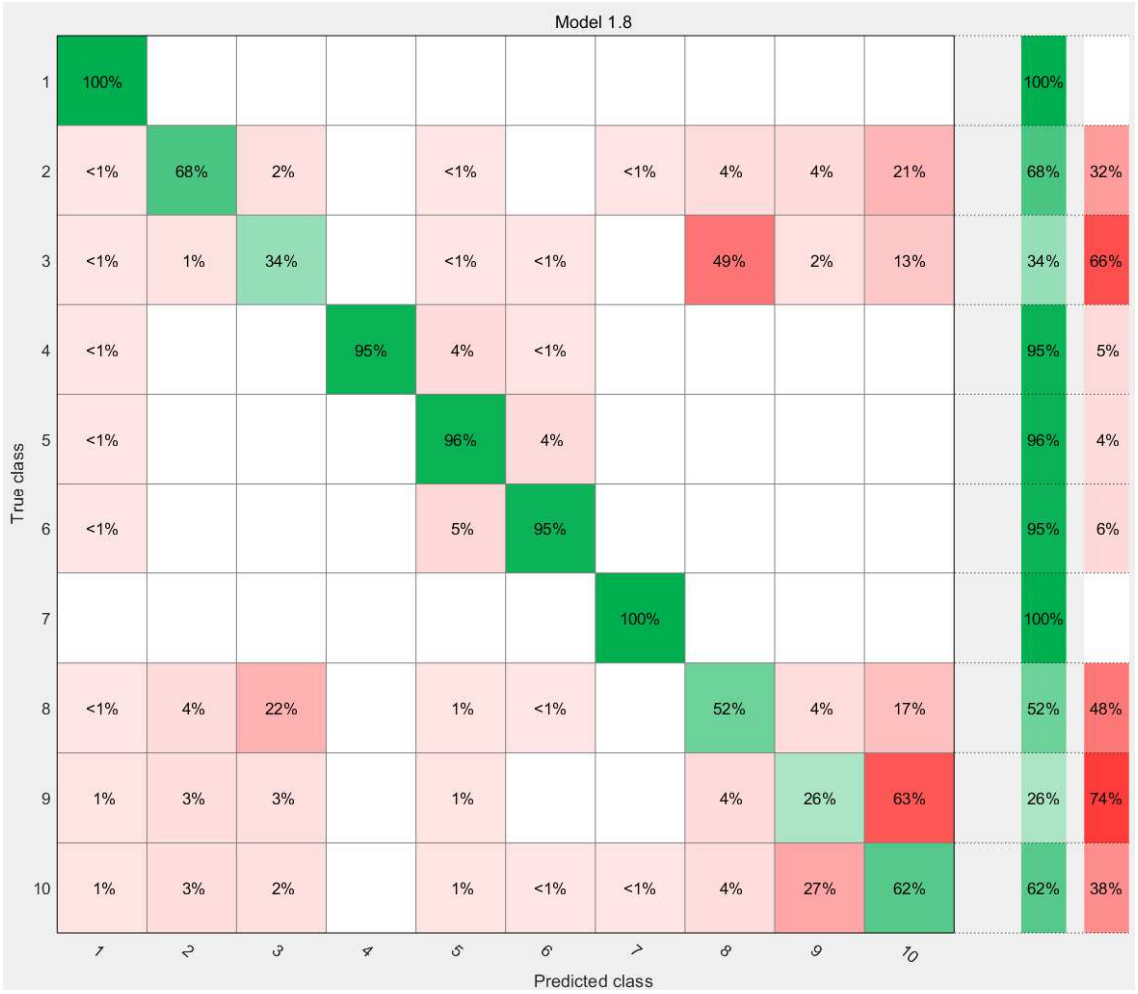

Fig.3.7 Confusion Matrix of Cubic SVM Model

3.2.4 Fine Gaussian SVM

The confusion matrix of the Fine Gaussian SVM classification result is depicted in Fig.3.8. Fine Gaussian SVM works not well in distinguishing operating state 8,9,10 from state 3. In particular, the accuracy of states 3,9,10 are all pretty low. Therefore, the accuracy of this model is 83.6%.

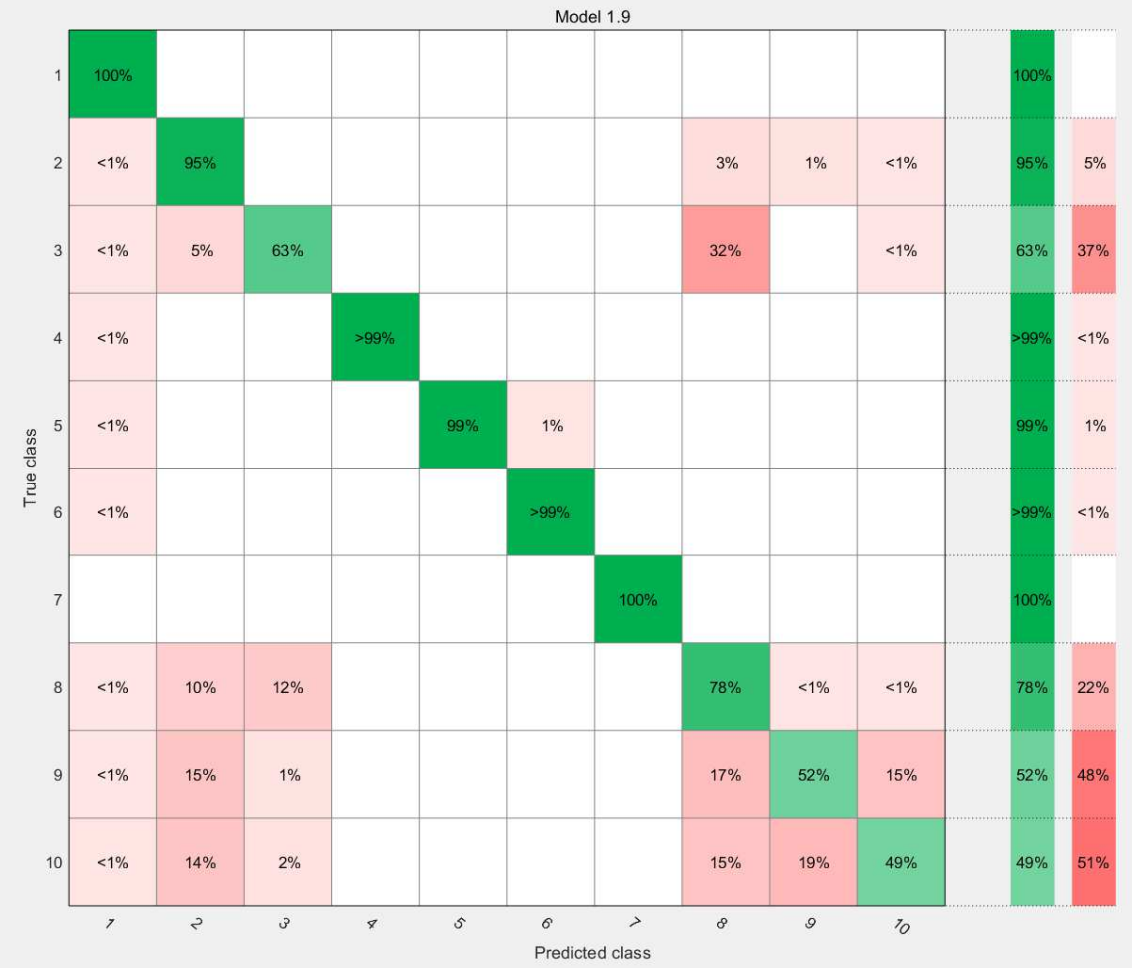

Fig.3.8 Confusion Matrix of Fine Gaussian SVM Model

3.2.5 Medium Gaussian SVM

The confusion matrix of the Medium Gaussian SVM classification result is depicted in Fig.3.9. Medium gaussian SVM works not well in distinguishing operating state 3 from other states. Especially, classifier works worse in operating states 8,9,10 which are the DC Line-to-ground fault. Therefore, the accuracy of this model is 74.7%.

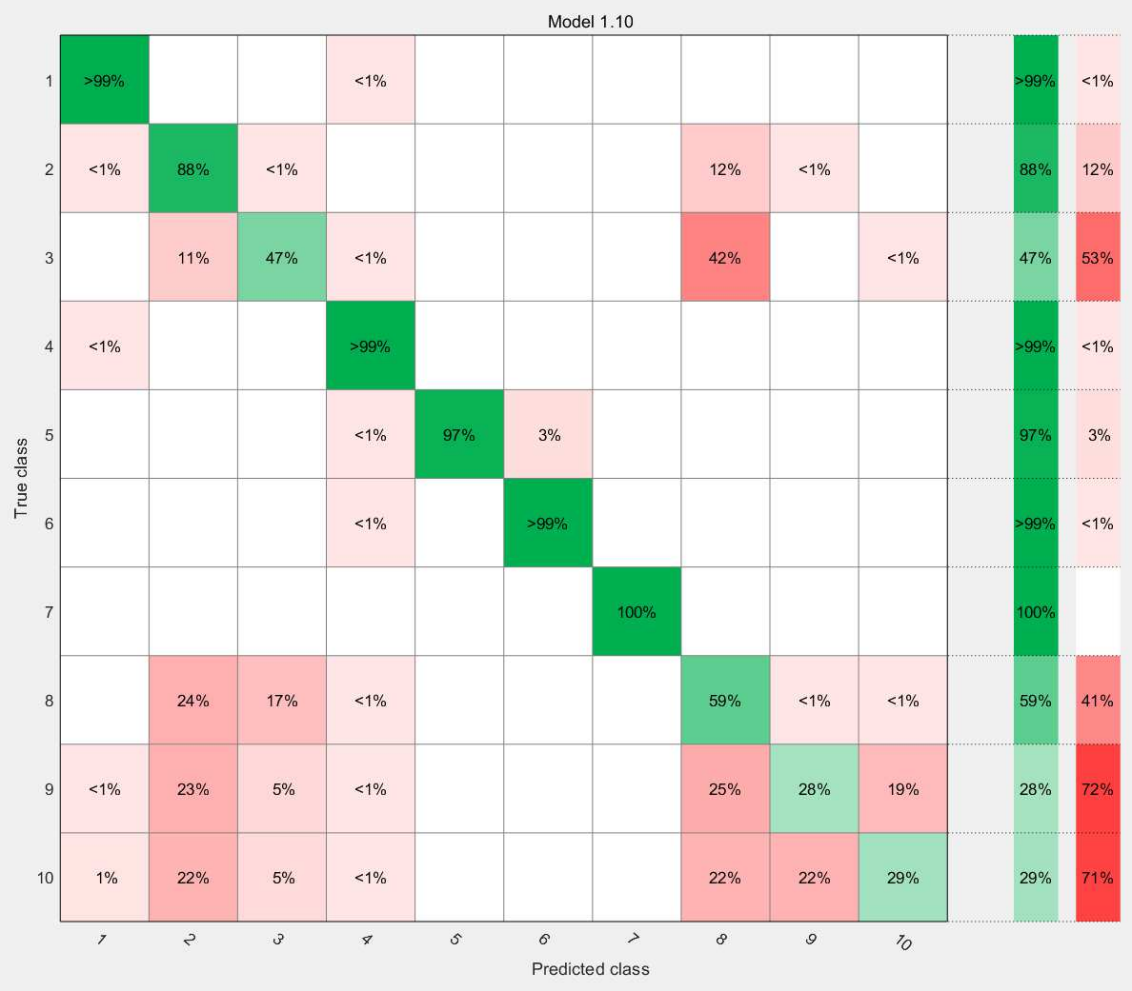

Fig.3.9 Confusion Matrix of Medium Gaussian SVM Model

3.2.6 Coarse Gaussian SVM

The confusion matrix of the Coarse Gaussian SVM classification result is depicted Fig.3.10. Coarse Gaussian SVM works badly in distinguishing operating state 3,8,9,10 from other states. Also, the most samples in these operating states are assigned into operating state 2. Therefore, the accuracy of this model is 63.1%.

Fig.3.10 Confusion Matrix of Coarse Gaussian SVM Model

3.3 Nearest Neighborhood Classifiers

K-Nearest Neighbor (KNN) is an algorithm used in classifying samples and regression analysis, proposed by Cover and Hart in 1968. The input is the vector consisting of the data point in the feature space. The output of the K-Nearest Neighbor is the multiple categories of the samples. In K-Nearest Neighbor algorithm, a training dataset is assumed that the classification of some samples has been determined. When new samples are input in the classifier, the classification of new samples is determined through majority vote of classification according K Nearest Neighbor point training samples. Therefore, there is not explicit learning process in KNN algorithm[43].

Basic elements of KNN are selection of k, distance metrics, classification decision rules. In distance metrics, mostly used methods include European distance, Manhattan distance, Chebyshev distance, Minkowski distance. European distance between $\vec{a}(x_{11},\ x_{12},\cdots x_{1n})$ and $\vec{b}(x_{21},\ x_{22},\cdots x_{2n})$ can be calculated by using equation (24).

$$d = \sqrt{\sum_{k=1}^{n}(x_{1k} - x_{2k})^2} \tag{24}$$

Algorithm process (perform on each unknown point):

a. The distance from unknown points to all the points that the categories is known is calculated.

b. Categories is sorted by distance (ascending).

c. The first k points that are closes to the unknown point are selected. The number of the categories of these k points is counted.

d. The category which appears most frequently in the above K points is the category of the unknown point.

According to the algorithm process, the idea of KNN is simple and effective. However, whole data is needed to be input in KNN so that requirement on the hardware is high when the size of data is huge. It may take much time as distance between each unknown point and all known points is calculated. Also, the classification result is sometimes hard to understand.

3.3.1 Fine KNN

The confusion matrix of the Fine KNN classification result is depicted in Fig.3.11. Fine KNN works well in operating state 1,2,4,5,6,7 classification. However, it works not very well in the three DC line-to-ground faults and operating state 3. As we can see from the confusion matrix, these four states are easily confused. Therefore, the accuracy of this model is 89.9%.

Fig.3.11 Confusion Matrix of Fine KNN Model

3.3.2 Medium KNN

The confusion matrix of the Medium KNN classification result is depicted Fig.3.12. Medium KNN works well in operating state 1,2,4,5,6,7 classification. However, it has the same problem that model works not very well in the three DC line-to-ground faults and operating state 3 as Fine KNN. As we can see from the confusion matrix, operating states 1,2,3,8,9,10 are confused. Therefore, the accuracy of this model is 85.5%.

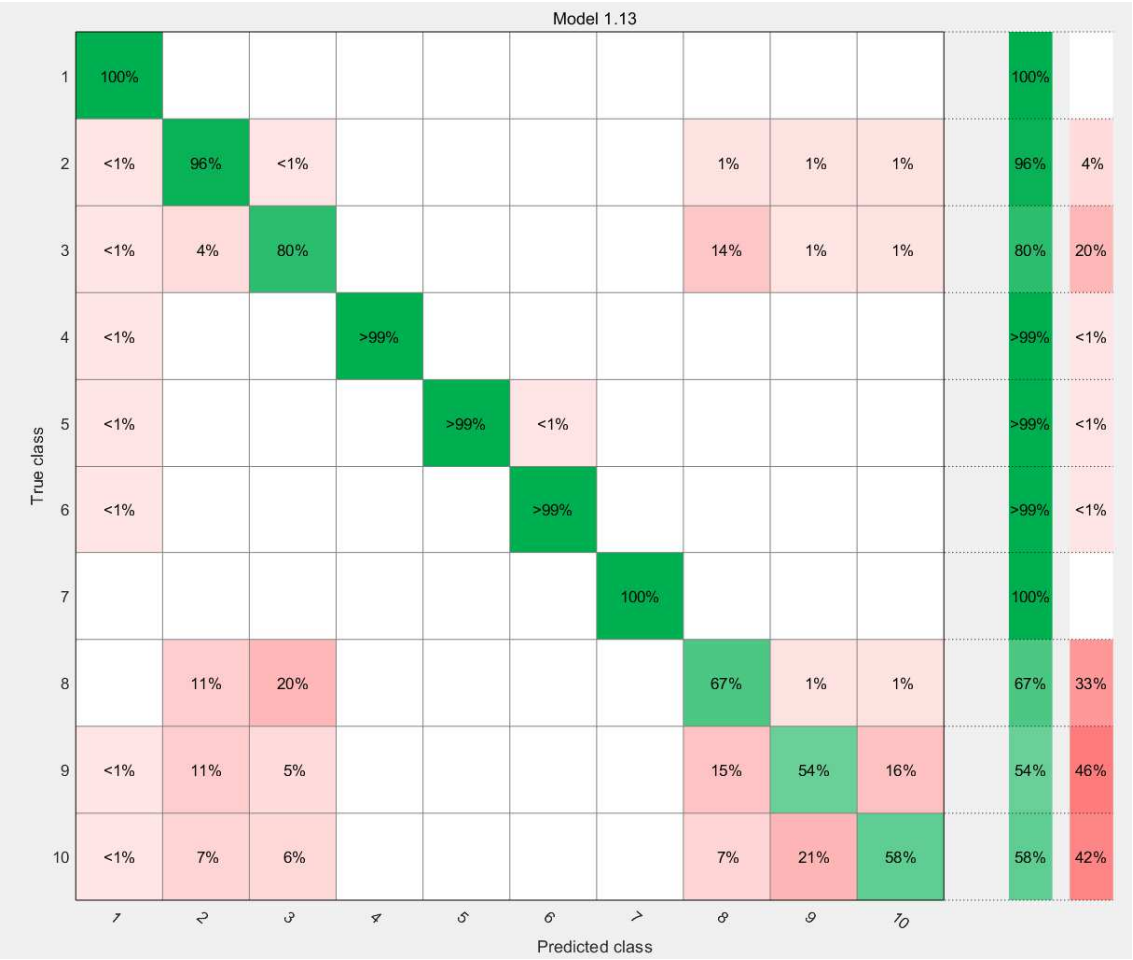

Fig.3.12 Confusion Matrix of Medium KNN Model

3.3.3 Coarse KNN

The confusion matrix of the Coarse KNN classification result is depicted Fig.3.13. Coarse KNN works badly in distinguishing operating state 2,3,8,9,10 from other states. Except operating state 7 which is perfectly classified, all operating Tate's are not well distinguished. Therefore, the accuracy of this model is 71.7%.

Fig.3.13 Confusion Matrix of Coarse KNN Model

3.3.4 Cosine KNN

The confusion matrix of the Cosine KNN classification result is depicted in Fig.3.14. Cosine KNN works well in this classification. The only problem is that operating states 8,9,10 which represent DC Lin-to-ground fault at different location are not only confused with each other, but also confused with the first three operating states, especially the third operating state. The accuracy of this model is 84.9%.

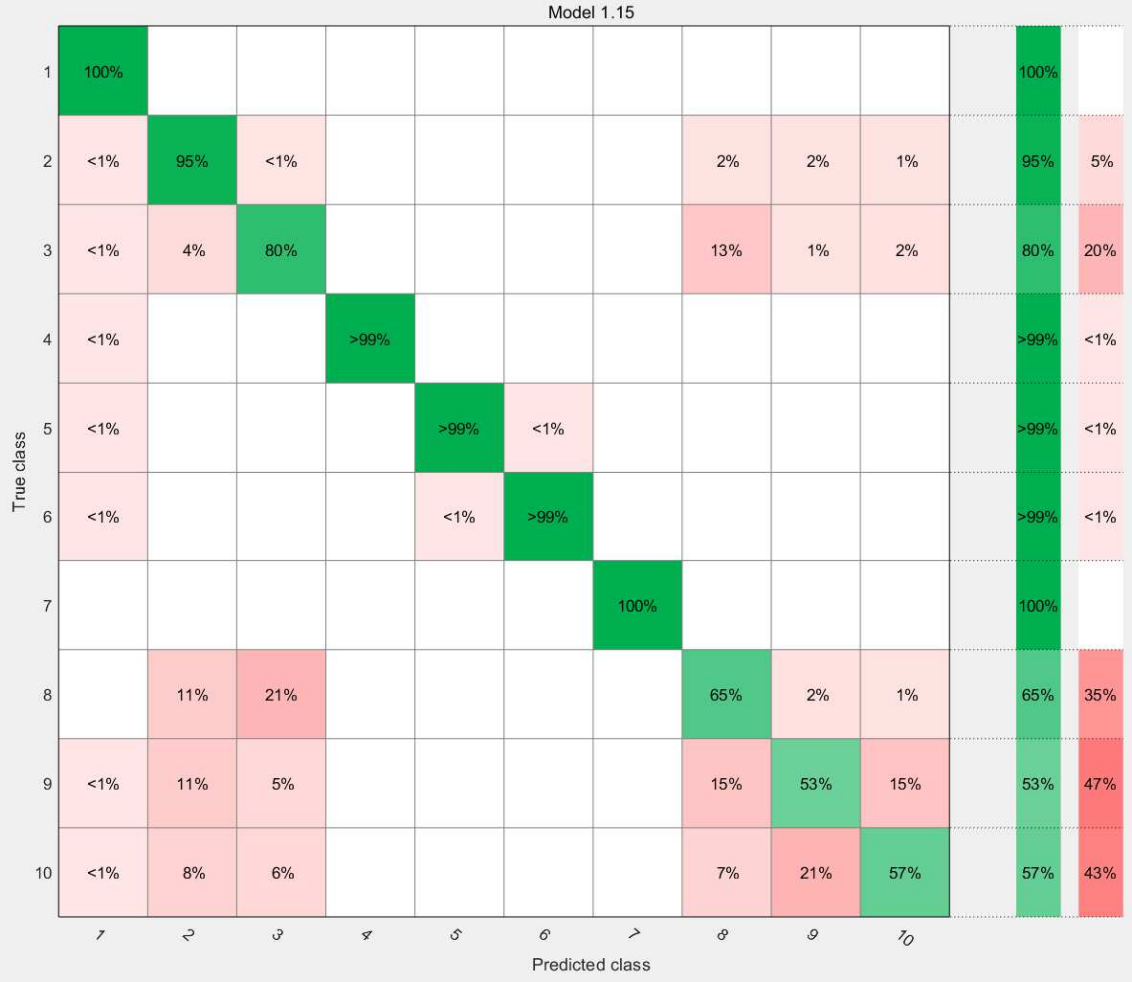

Fig.3.14 Confusion Matrix of Cosine KNN Model

3.3.5 Cubic KNN

The confusion matrix of the Cubic KNN classification result is depicted in Fig.3.15. Confusion mastic of Cubic KNN is quite similar to the one of Cosine KNN. Operating states 3,8,9,10 are confused with each other. The accuracy of this model is 84.6%.

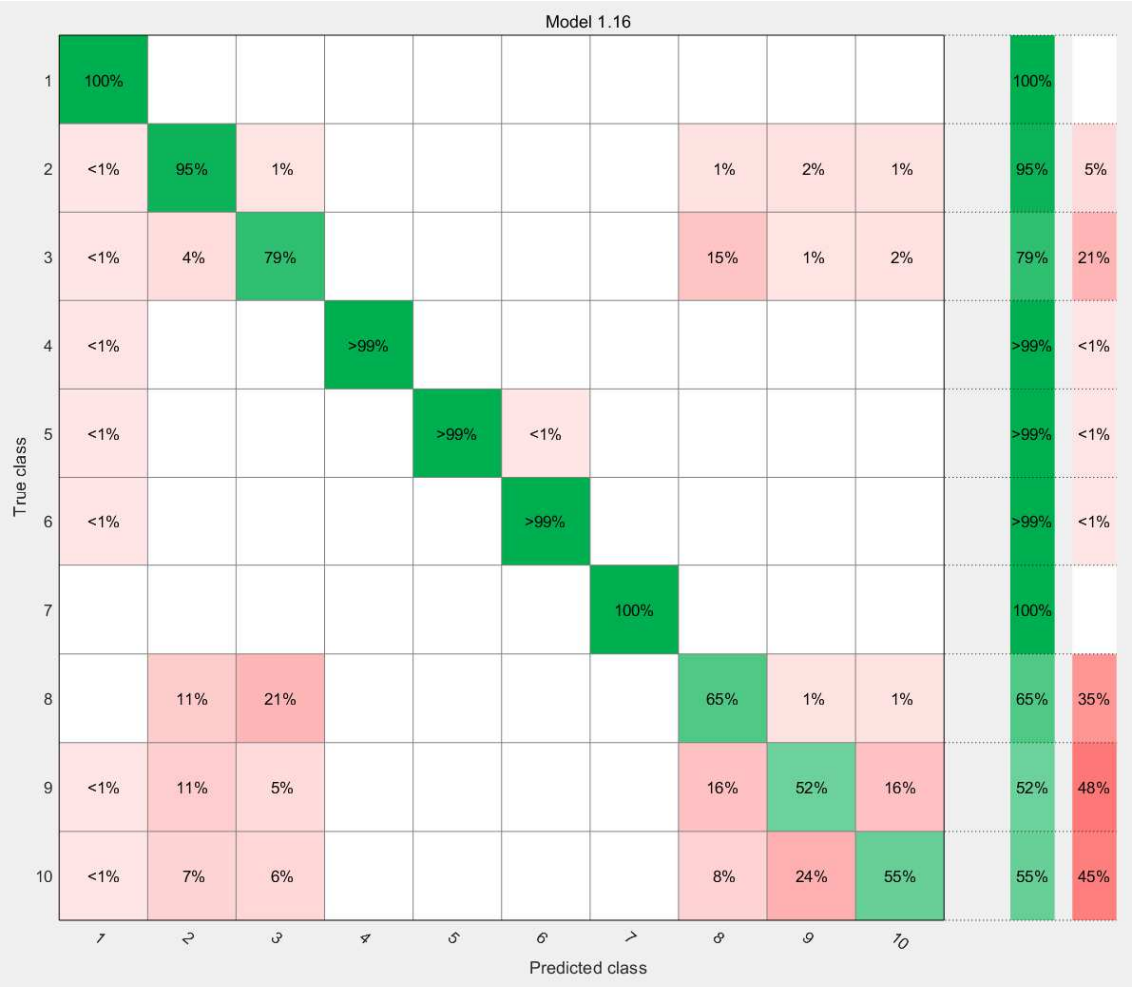

Fig.3.15 Confusion Matrix of Cubic KNN Model

3.3.6 Weighted KNN

The confusion matrix of the Weighted KNN classification result is depicted in Fig.3.16. Weighted KNN works well in most operating states. However, it's hard for this model to handle with the classification of DC line-to-ground faults at different locations. The accuracy is 88.9%.

Fig.3. 16 Confusion Matrix of Weighted KNN Model

3.4 Ensemble Classifiers

Ensemble Learning is the method that appropriately combines a variety of weak learners with lower performance to form a high-performance learner.

a. Bagging Learning

In Bagging Learning, a virtual training sample is generated by selecting *n* samples from the original dataset $\{(x_i, y_i)\}_{i=1}^{n}$. Duplicate selection is allowed. Therefore, several training sample set with a slight difference compared with the original sample set are obtained. These new sample set are input in week learner φ_j. Then a strong learner *f* is obtained by calculating average of all week learners $\{\varphi_j\}_{j=1}^{b}$ as the quation (25).

$$f(x) \leftarrow \frac{1}{b}\sum_{j=1}^{b}\varphi_j(x) \tag{25}$$

By using the above method, a number of slightly different weak classifiers can be obtained from a large set of training samples. Then these classifiers are integrated, and a stable and reliable classifier can be obtained.

b. Boosting Learning

The basic idea of Boosting Learning is that the weights of the wrong-classified-samples are increased and the weights of the right-classified-samples are declined during the iteration. The most standard weighting method is the AdaBoost algorithm.

All weight $\{w_i\}_{i=1}^{n}$of the training samples $\{(x_i, y_i)\}_{i=1}^{n}$ are set to $\frac{1}{n}$. The weight value of strong classifier is set to 0.

$$w_1, \cdots, w_n \leftarrow \frac{1}{n}, f \leftarrow 0 \tag{26}$$

For $j = 1, \cdots, b$, the following calculations are repeated.

a. For the current weight $\{w_i\}_{i=1}^{n}$, the weak classifier with the lowest misclassified rate train the samples.

$$\varphi_j = argminR(\varphi\ ) \tag{27}$$

3.4.1 Boosted Trees

The confusion matrix of the Boosted Trees classification result is depicted in Fig.3.17. Boosted tree works not well in most operating states. Especially, classifier works worse in operating states 3,9,10. Samples under operating state 8 are well classified. However, 46% samples under operating state 3 are wrongly classified into operating state 8. Therefore, the accuracy of this model is 76.6%.

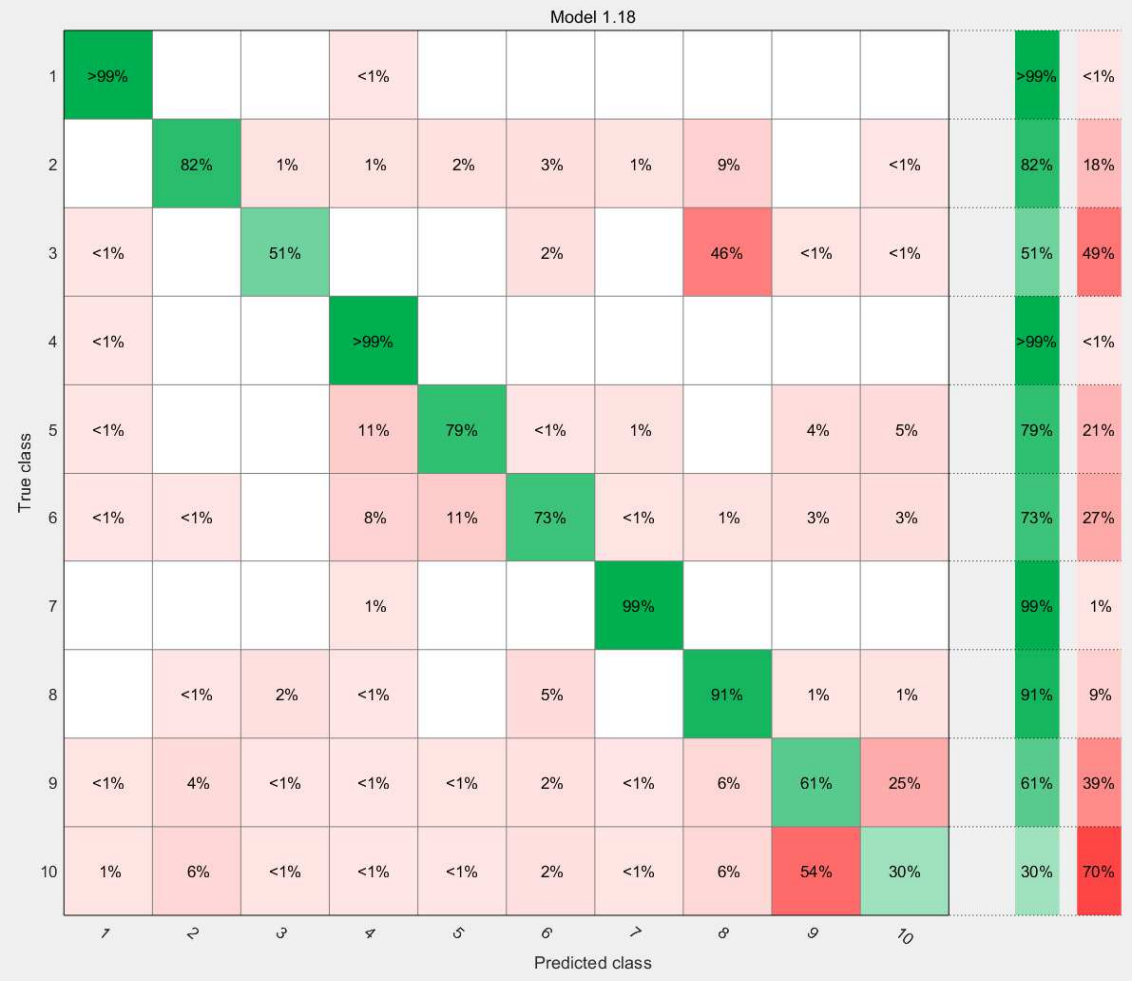

Fig.3.17 Confusion Matrix of Boosted Trees Model

3.4.2 Bagged Trees

The confusion matrix of the Bagged Trees classification result is depicted in Fig.3.18. By using bagged tree, most samples are classified perfectly into right operating states category. Still, when it comes to DC line-to-ground at different locations faults, performance is not perfect but acceptable, because only the last two fault states are confused. The fault locations in last two fault state is close so that the difficulty of finding the actual grounding location is much easier that the scenarios that more than two non-related operating states are confused. The accuracy is 96.5%.

Fig.3.18 Confusion Matrix of Bagged Trees Model

3.4.3 Subspace Discriminant

The confusion matrix of the Subspace Discriminant classification result is depicted in Fig 3.19. Subspace Discriminant fails in most operating state classification. The accuracy is only 42.1%.

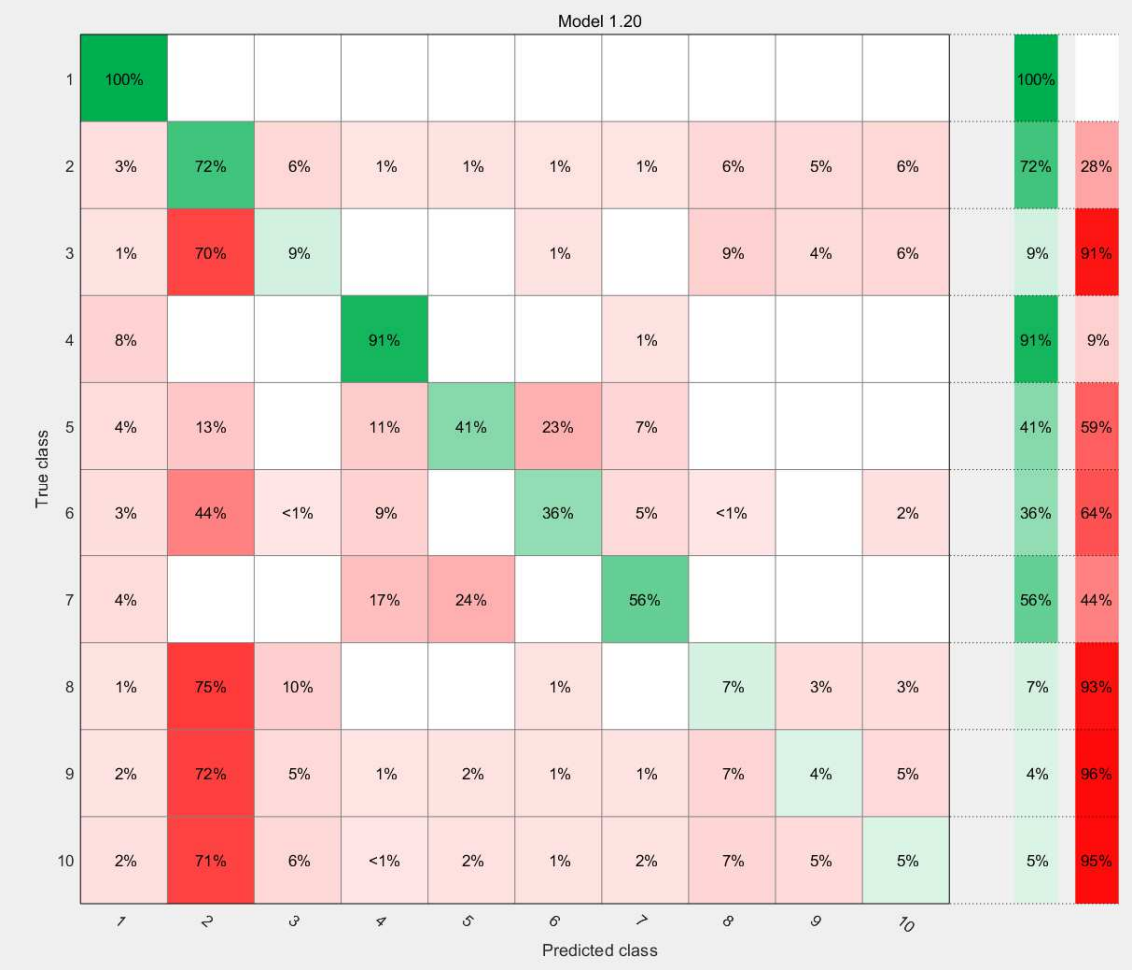

Fig.3.19 Confusion Matrix of Subspace Discriminant Model

3.4.4 Subspace KNN

The confusion matrix of the Subspace KNN classification result is depicted in Fig.3.20. Subspace KNN work well in most operating states. However, like the other KNN methods used in this thesis, it can hardly distinguish operating states 3,8,9,10, especially operating states 3 and 8. The accuracy of this model is 87.9%

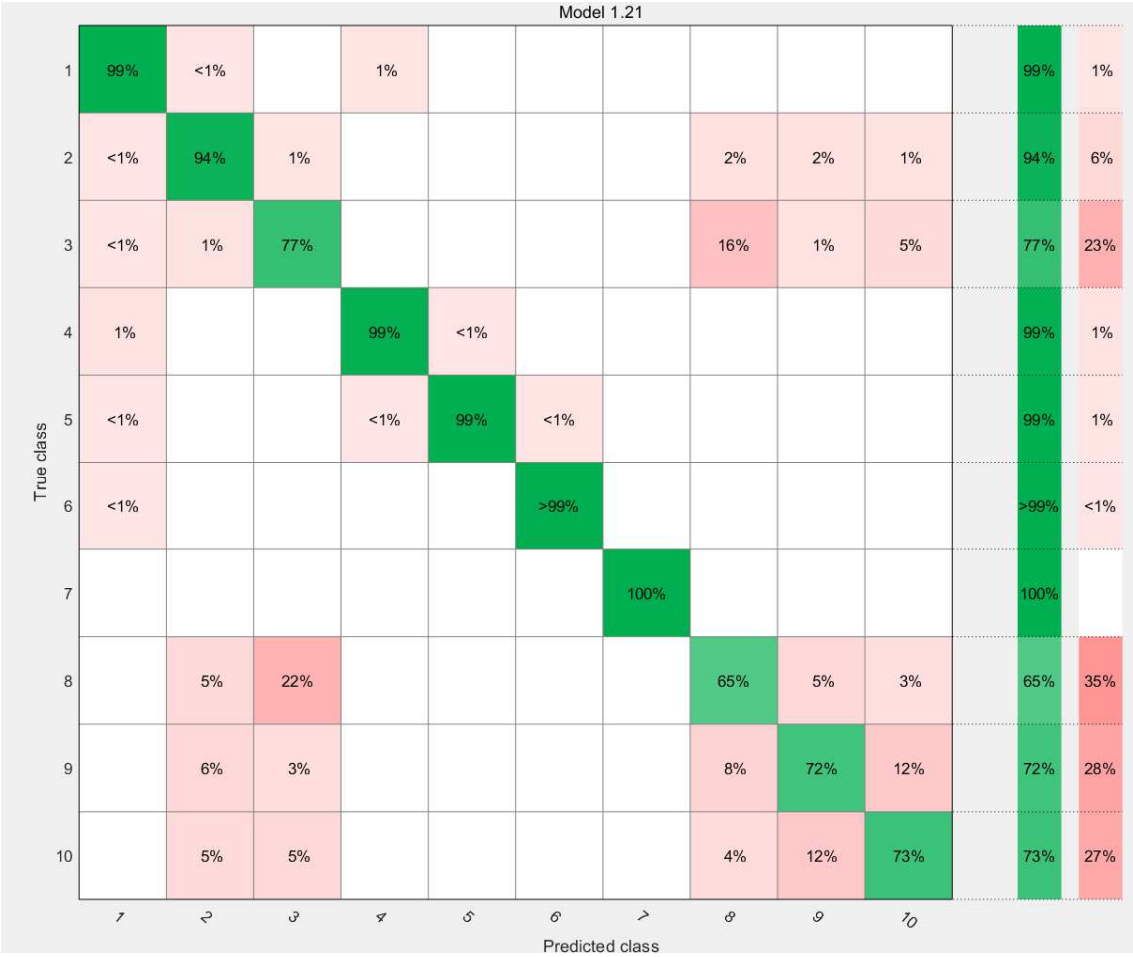

Fig.3.20 Confusion Matrix of Subspace KNN Model

3.4.5 RUSBoosted Trees

The confusion matrix of the RUSBoosted Trees result is depicted in Fig.3.21. RUSBoosted tree works well in some of the operating states, but the performance of this model in other operating states is poor. Moreover, it cannot distinguish the last operating state. The accuracy is 73.4%.

Fig.3.21 Confusion Matrix of Subspace KNN Model

3.5 Discriminant Analysis

When the categories some samples are known and new samples are needed to be classified, Discriminant Analysis is good to mine the largest value of the data. It is usually necessary to give a descriptive statistical model to measure the proximity of new samples to

each known category. The model is the discriminant function. Moreover, a discriminating rule is needed to judge the attribution of the new sample. The discriminant rules can be deterministic or statistical, corresponding to Fisher discrimination and Bayes discrimination.

The basic idea of Fisher discrimination is projection, which is to project multidimensional samples on one dimension. The expectation of projection is that the different sample points of each whole are separated as far as possible, and the sample points from the same whole are concentrated as far as possible. The projection function derived from the idea of unary variance is the discriminant function. It can also be said that Fisher is looking for such a space in which the distance between points in same class is the smallest, the distance between classes is the largest.

Linear Discriminant Analysis is proposed by Fisher (1936) for flower classification. First, Fisher linear discriminant function is calculated, Then, discriminant boundary values are calculated. At last, the discriminant rule is set up.

The confusion matrix of the Linear Discriminant result is depicted in Fig.3.22. Except the first operating state, Linear Discriminant can hardly complete the classifier work. It doesn't work in the DC line-to-ground at different locations. The accuracy of this model is 42.7%.

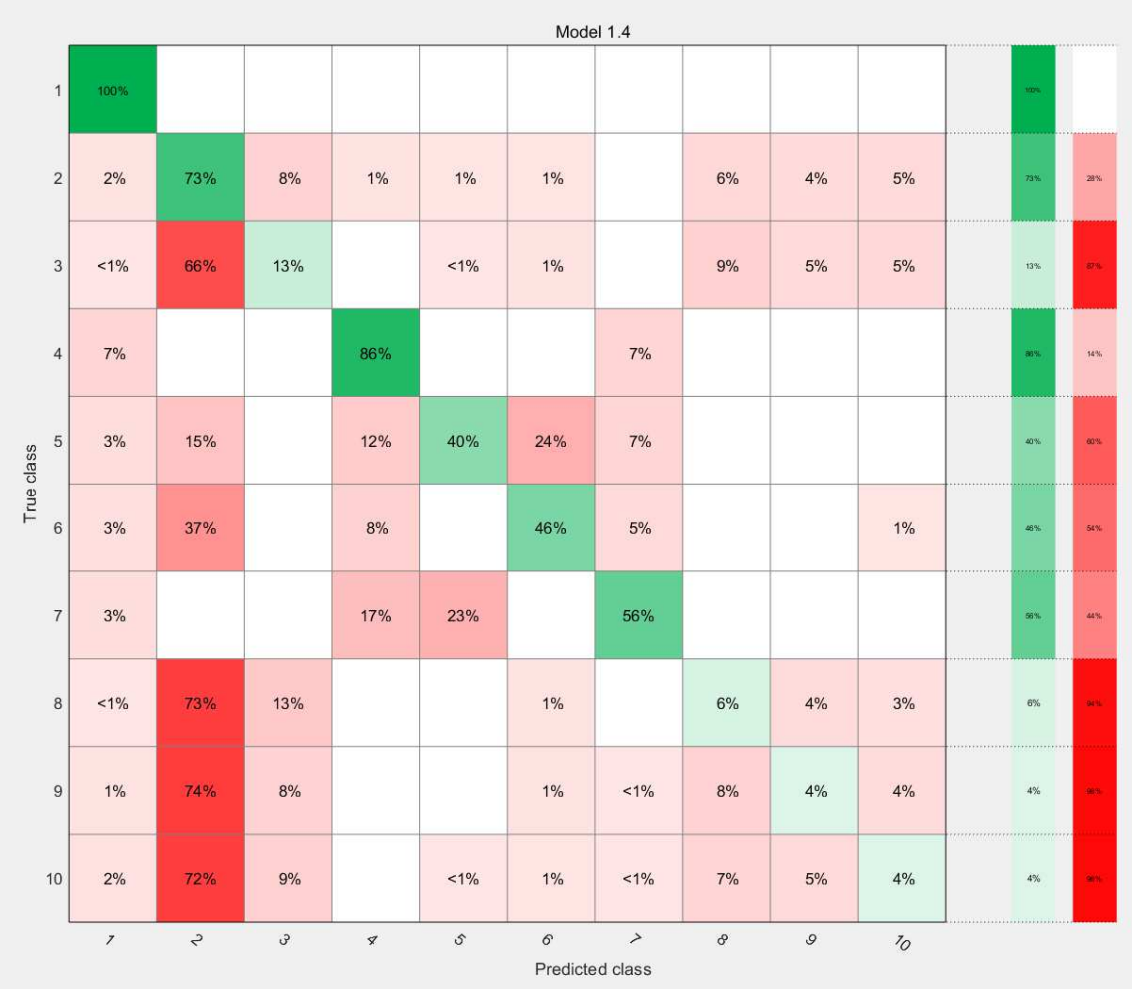

Fig.3.22 Confusion Matrix of Linear Discriminant Model

3.6 Backward Propagation Neural Network

Neural Network mimic the process of human neuronal activation and transmission. Taking three-layer neural network as an example, BP neural network includes input layer, hidden layer and output layer, depicted in Fig3.23. The input layer receives the data and the output layer outputs the data. The previous layer of neurons connects to the next layer of

neurons. Each neuron collects the information from the previous layer of neurons, and passes the value to the next layer after activation. BP Neural network is one of the supervised learnings. BP Neural Network is the most basic neural network. Its output is transmitted by forward propagation, and the error is transmitted by back propagation.

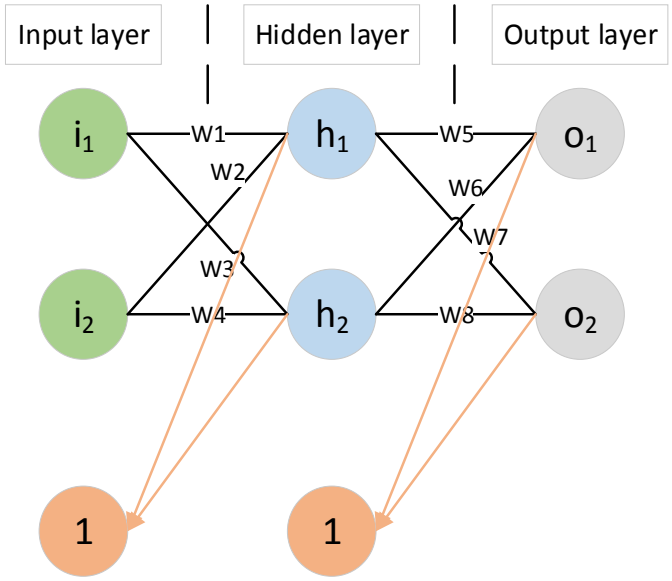

Fig.3.23 BP-NN Model Structure

The learning rate affects the speed of convergence. Also, it's a key parameter for the convergence. The low learning rate setting contributes the convergence of the network, but the convergence will spend more time. On the contrary, the high learning rate setting may result in none convergence, which cause a bad classification performance. Considering this situation, learning rate is set as 0.01.

Besides that, the number of hidden layer nodes has little effect on the recognition rate, but too many nodes will increase the computation, which makes the training slower. The quantity of hidden layer nodes Q is mainly calculated by experience equation (28) – (30).

$$Q = m + n + t \tag{28}$$

$$Q = \sqrt{m + n} + a \tag{29}$$

$$Q = \sqrt{mn} \tag{30}$$

m-dimensions of input vector, n-dimensions of output vector, t-a random number

between 2 and 6, *a*-a random number between 1 and 10. By using the equation (28), *t* is set as 7 so the hidden layer nodes are set as 24.

The activation function has a significant effect on both the recognition rate and the convergence speed. When approaching the high quadratic curve, the S-shaped function is much more accurate than the linear function, but the computation is much larger. Several combinations of common learning rule and transfer function are tested. The result is depicted in Table 3.1-3.4. The dataset used in this test do not include the data from DC line-to-ground at different faults, which makes the accuracy better. As the result depicted, Levenberg-Marquardt (trainlm) is used as the learning rule and Hyperbolic tangent S-type transfer function (tansig) is used as the transfer function.

Table.3.1 Testing Results of BP-NN Model Using Trainlm as Learning Rule

Learning Rule	Levenberg-Marquardt (trainlm)		
Activation Function	Sigmoid Function (logsig)	Hyperbolic tangent S-type transfer function (tansig)	Flexible maximum transfer function (softmax)
Performance	138 iterations	137 iterations	163 iterations
Accuracy	97.536%	98.107%	96.607%

Table.3.2 Testing Results of BP-NN Model Using Traingd as Learning Rule

Learning Rule	Steepest Gradient Descent (traingd)		
Activation Function	Sigmoid Function (logsig)	Hyperbolic tangent S-type transfer function (tansig)	Flexible maximum transfer function (softmax)
Performance	0.14	0.122	0.118
Accuracy	37.214%	39.25%	32.893%

Table.3.3 Testing Results of BP-NN Model Using Traingdx as Learning Rule

Learning Rule	Momentum and Dynamic Adaptive Learning Rate Gradient Descent Algorithm (traingdx)

Activation Function	Sigmoid Function (logsig)	Hyperbolic tangent S-type transfer function (tansig)	Flexible maximum transfer function (softmax)
Performance	0.047	0.0456	0.0442
Accuracy	78.929%	78.429%	76.929%

Table.3.4 Testing Results of BP-NN Model Using Traingda as Learning Rule

Learning Rule	Dynamic Adaptive Learning Rate Gradient Descent Algorithm (traingda)		
Activation Function	Sigmoid Function (logsig)	Hyperbolic tangent S-type transfer function (tansig)	Flexible maximum transfer function (softmax)
Performance	0.07	0.0583	0.0553
Accuracy	75.571%	77.643%	72.250%

The number of epochs has a slight impact on the accuracy, depicted in Fig3.24. When the epoch is increasing, the accuracy is well improved and then remain around 75%. However, the time needed to training is increase sharply, so the requirements on hardware increase at the same time. The maximum epochs are set as 200.

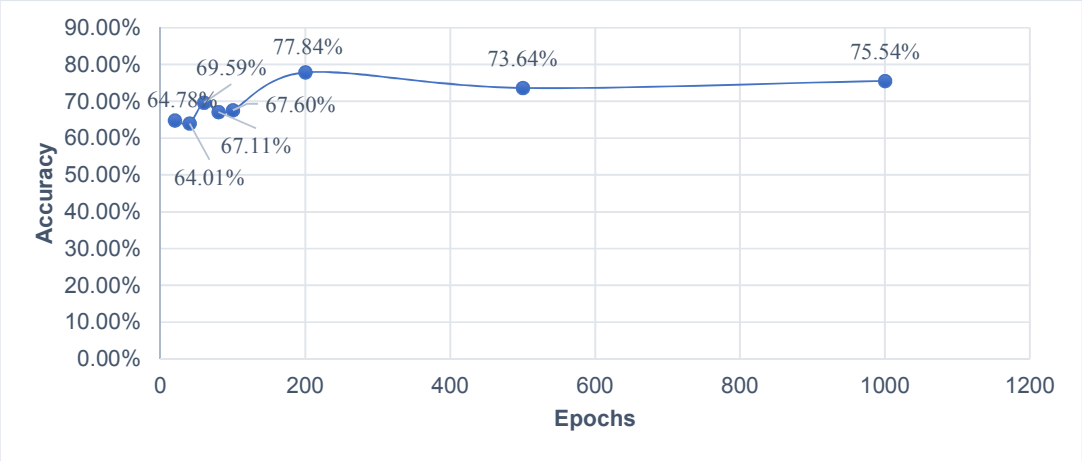

Fig.3.24 Relationship between Accuracy and Epochs

The Mean Square Error of the BP-NN model is depicted in Fig3.25. Mean Square Error, also called Quadratic Loss and L2 Loss, is a measure of the degree of difference between the model and original data. The performance goal is set as 0.01, but the raining performance

only remain 0.0294 when epoch reach the default value.

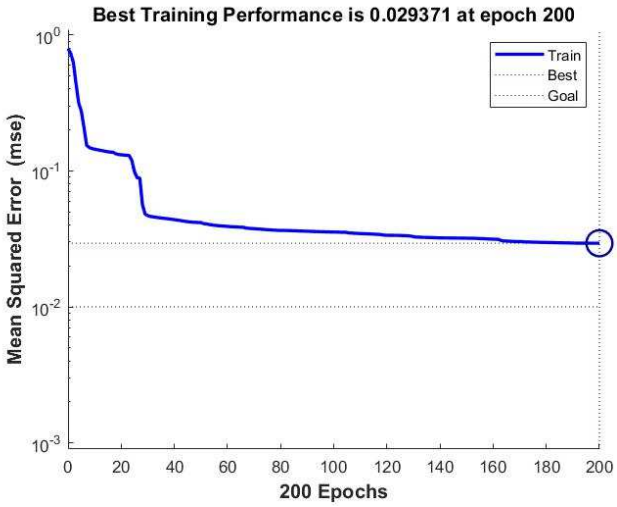

Fig.3.25 Mean Square Error of the BP-NN

Finally, the accuracy of BP-NN model is 77.84%.

3.7 Long Short-Term Memory Neural Network

The structure of the LSTM-Neural Network is different from that of the traditional neural network. There is a ring pointing in it, which is used to pass the information processed at the current time. Also, it is possible that the information in the past time can be used now. The structure of ring is depicted in Fig.3.26.

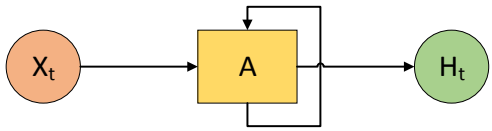

Fig.3.26 Simple Structure of LSTM-NN

X_t - Input Signal, A – Processing Part, H_t – Output Signal

When the whole structure is expanded in Fig.3.27, LSTM-NN is a chain neural network. It can be considered as a multiple replication of simple neural network.

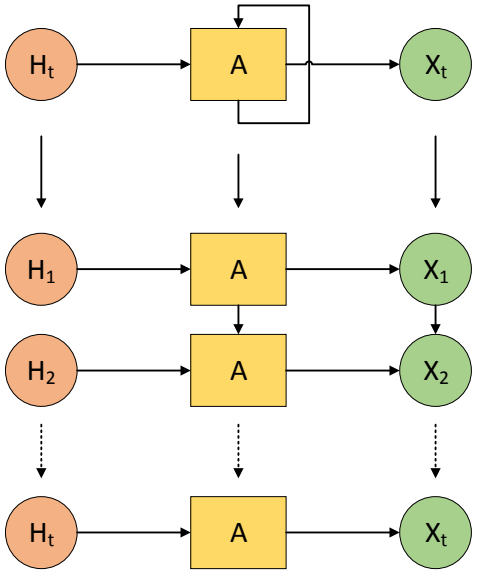

Fig.3.27 Expanded LSTM-NN Structure

There is addition of valve nodes of various layers in LSTM-NN except the RNN structure. There are 3 types of valves: The Forget Gate, the Input Gate and the Output Gate. These valves can be turned on or off to determine whether the output of the previous memory state of the network model has reached a threshold in that layer. The sigmoid function is commonly used as an activation function in LSTM, which can obtain the status of network.

The weights of each layer are gradually modified during the backward propagation. The memory function of the LSTM model is realized by these valve nodes. When the valve is opened, the training results of the previous model are input into the current model calculation. If the valve is off, previous results will not have an impact on next model[44].

The results are depicted in Fig 3.28, Fig3.29. Fig.3.30 is the illustrations of the accuracy waveform. Obvious oscillations can be seen in the accuracy waveform. The highest accuracy during whole testing is around 70%, and the accuracy remains 62.5% at the end of testing.

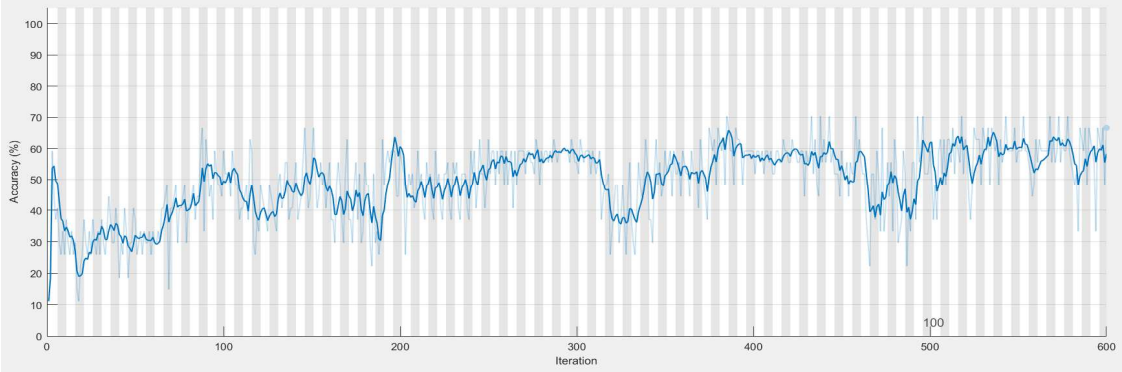

Fig.3.28 Accuracy Waveform of LSTM Model

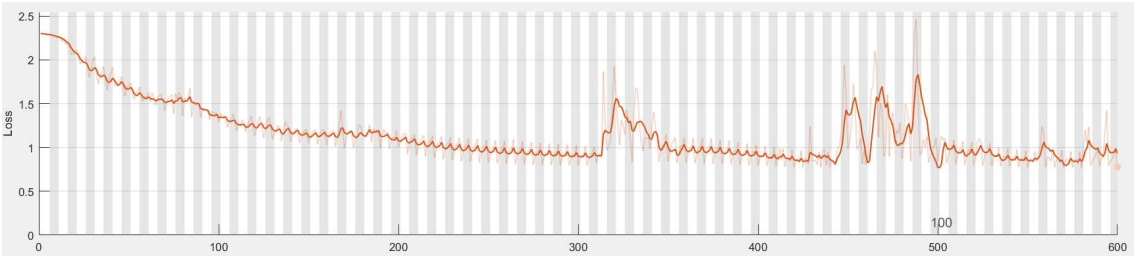

Fig.3.29 Loss Waveform of LSTM Model

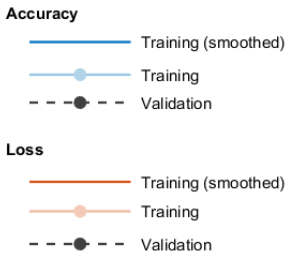

Fig.3.30 Illustrations of The Accuracy and Loss Waveform

3.8 Extreme Learning Machine

The Extreme Learning Machine (ELM) was proposed to solve the problem of low efficiency and complicated parameter setting of backward probation. ELM usually uses Single-Layer Feedforward Neuron Network as the structure (SLFN). SLFN includes input layer, hidden layer and output layer. The output equation of hidden layer is calculated by using equation (31).

$$f_L = \sum_{i=1}^{l} \beta_i h_i(x) = h(x)\beta \quad (31)$$

x-Input of neural network, β-weight of output, $h(x)$-activation function which is used to map data in input layer to feature space. Function $h(x)$ can be calculated by using equation

(32).

$$h(x) = G(a_i, b_i, x) \tag{32}$$

a_i-input weights

a_i and b_i are the parameters of feature mapping, which is also called node parameters. Since the mapping is random or artificially determined and it's not going to be modified during the procedure, the mapping can be any piecewise continuous function. The common piecewise continuous functions are:

a. Trigonometric Function

$$G(a_i, b_i, x) = \cos(a \cdot x + b) \tag{33}$$

b. Gaussian Function

$$G(a_i, b_i, x) = \sqrt{\|x - a\| + b^2} \tag{34}$$

c. Radial Basis Function

$$G(a_i, b_i, x) = \exp(-b \cdot \|x - a\| \tag{35}$$

d. Sigmoid Function

$$G(a_i, b_i, x) = \frac{1}{1+\exp(a \cdot x + b)} \tag{36}$$

e. Hyperbolic Sine Function

$$G(a_i, b_i, x) = \frac{1-\exp(a \cdot x + b)}{1+\exp(a \cdot x + b)} \tag{37}$$

f. Hard Limit Function

$$G(a_i, b_i, x) = \begin{cases} 1, & \text{if } a \cdot x + b \leq 0 \\ 0, & \text{otherwise} \end{cases} \tag{38}$$

Only the output weight is needed to solve in ELM, which is essentially linear-parameter model, so its learning process is easy to converge on the minimum[45].

The structure of ELM is depicted in Fig.3.32.

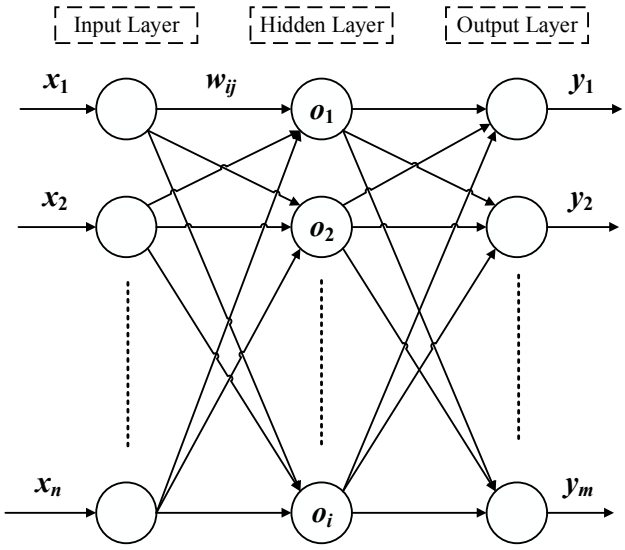

Fig.3.31 Structure of ELM

Learning process of ELM with N group of data, L nodes in hidden layer and M output layer is depicted below.

a. Assigning Node Parameters Randomly - At the beginning of the calculation, the node parameters of the SLFN are randomly generated, that is, the node parameters are independent of the input data. The random generation can use any continuous probability distribution.

b. Hidden Layer Output Matrix Calculation - Randomly assign node parameters: The number of rows is the number of training data entered. The number of columns is the number of implicit layer nodes. Therefore, the size of the implicit layer output matrix is N-M. The output matrix is essentially the result of mapping N input data to L nodes.

c. Solving the Output Weights - Unlike other algorithms, the output layer can have no error nodes in the ELM algorithm. The core of the ELM algorithm is to solve the output weight so that the error function is minimal.

L1 loss function used in ELM is:

$$\min_{\beta \in R^{L \times m}} \|H\beta - T\|^2 \tag{39}$$

H-Output matrix, *T*- training goal, ||- Frobenius norm of the matrix element

After the introduction of L1, the function is rewritten as:

$$\min_{\beta \in R^{L \times m}} \frac{1}{2}\|\beta\|^2 + \frac{C}{2}\|H\beta - T\|^2 \tag{40}$$

C- regularization coefficient.

So, the problem is the same as Ridge Regression problem which is calculated by using the function (41).

$$\beta^* = \left(H^T H + \frac{1}{C}\right)^{-1} H^T T \tag{41}$$

Except that, single value decomposition (SVD) also can be used to solve the weight function:

$$H\beta = \sum_{i=1}^{N} u_i \frac{d_i^2}{d_i^2 + C} u_i^T X \tag{42}$$

u_i- feature vector of HH^T, d_i-feature value of *H*.

Studies have shown that relatively small weight coefficients can improve the stability and generalization ability of SLFN, so regularization of ELM is necessary in complex problems[46].

The number of hidden nodes has a significant effect on the accuracy, as shown in Fig.3.32. With the nodes increasing, the accuracy is first rise sharply but then growth rate has gradually slowed down. To make sure training time is acceptable, 10000 hidden layer nodes is set. The accuracy of ELM is 92.23%.

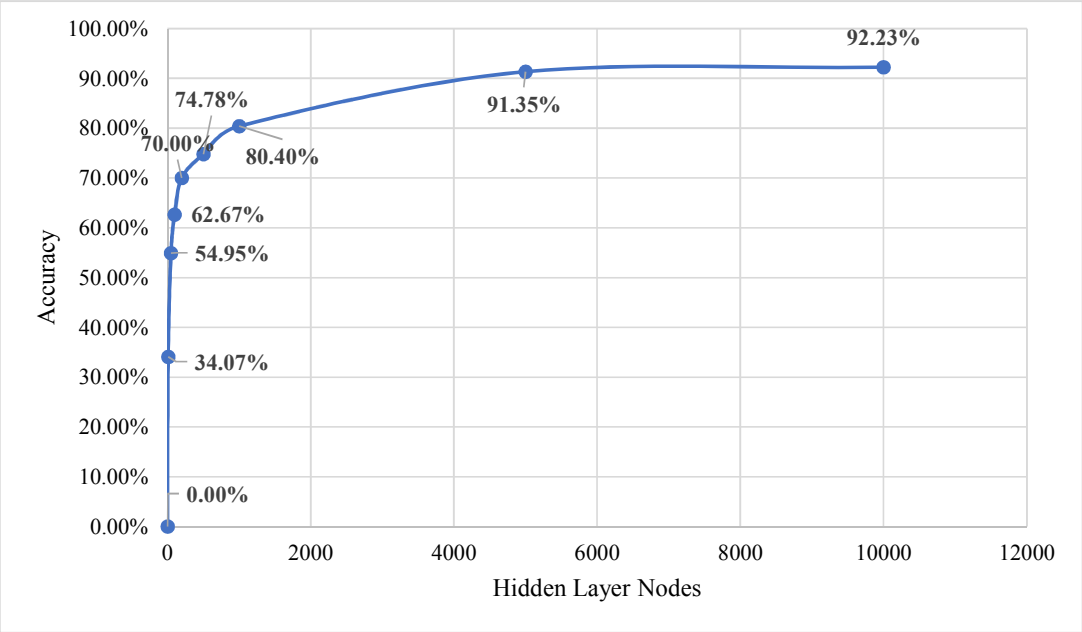

Fig.3.32 Relationship between Accuracy and Hidden Layer Nodes

3.9 Comparison and Analysis

The accuracy of all methods is aggregated in Table.3.5. Coarse Trees, Subspace Discriminant and Linear Discriminant have the worst performance in HVDC system fault diagnosis. Moreover, Medium Trees, Linear SVM, Quadratic SVM, Cubic SVM, Medium Gaussian SVM, Coarse Gaussian SVM, Coarse KNN, Boosted Trees, RUSBoosted KNN, BP-NN and LSTM-NN are not accurate enough for the fault diagnosis. Except methods above, accuracy of other methods is all higher than 80%. Bagged Trees has the highest accuracy of 96.5%. Although there is only a little difference between the waveform of DC line-to -ground at three different locations, it can well distinguish it for each other. It can overcome the hardest problem for other methods. In addition, the accuracy of ELM also looks good. However, it is result from 10000 hidden layer nodes setting. So, the training time is much longer than Bagged tree.

Table.3.5 Heat Table of Machine Learning Based Methods Accuracy

Methods	Medium Trees	Coarse Trees	ELM
Accuracy	73.40%	38.00%	92.23%
Methods	Linear Decision Tress	Linear SVM	Quadratic SVM
Accuracy	87.40%	57.60%	78.10%
Methods	Cubic SVM	Fine Gaussian SVM	Medium Gaussian SVM
Accuracy	72.80%	83.60%	74.70%
Methods	Coarse Gaussian SVM	Fine KNN	Medium KNN
Accuracy	63.10%	89.90%	85.50%
Methods	Coarse KNN	Cosine KNN	Cubic KNN
Accuracy	71.70%	84.90%	84.60%
Methods	Weighted KNN	Boosted Trees	Bagged Trees
Accuracy	88.90%	76.60%	96.50%
Methods	Subspace Discriminant	Subspace KNN	RUSBoosted KNN
Accuracy	42.10%	87.90%	73.40%
Methods	Linear Discriminant	BP-NN	LSTM-NN
Accuracy	42.70%	77.84%	62.50%

4 Conclusion and Future Work

In this thesis, few Machine learning based methods are tested in HVDC system fault diagnosis. Considering the DC line-to-ground might occur at any location in the DC line, three fault locations are set in the middle of the DC line, so that 9 fault operating states corresponding to 7 common faults and normal operating are considered in this thesis. To distinguish the different operating states, for each sample, 7 parameters are selected as the fault features. Then, waveforms of 7 faults features can be obtained after HVDC system is simulated in Simulink. According to the time step in simulation, 20000 samples can be obtained within the fault period, which are grouped into training sample set and testing dataset. After different classifiers are trained and tested, as the accuracy of each classifier is selected as the index of its performance, performance of classifiers are compared. Hence, the most appropriate classifier is selected in HVDC fault diagnosis. Moreover, during the model training, the impact of different transfer functions on the accuracy and learning rules combinations in BP-NN and the impact of different activation functions in ELM on the accuracy are also considered and tested. The combination and activation function with the best accuracy are respectively selected in the BP-NN model and ELM model. The accuracy of Coarse Tree, Subspace Discriminant and Linear Discriminant are lower than 50%. ELM, Linear Decision Trees, Fine Gaussian SVM, Fine KNN, Medium KNN, Cosine KNN, Cubic KNN, Weighted KNN, Subspace KNN and Bagged Trees are all accurate more than 80%. In these methods, the accuracy of Bagged Trees and ELM is above 90%, which are 96.50% and 92.23%.

For the future work, two research directions are valuable. On one hand, as the data

obtained from HVDC system is time series data, LSTM-NN is tested in this thesis. Besides the LSTM-NN, there are more algorithm that are designed to handle with time series data, including Hidden Markov Models, Dynamic Bayes Nets and so on. Theses algorithm might be more accurate than the algorithm tested in this thesis. On the other hand, some signal processing methods can be used to extract feature of data. Combining data processing and Machine Learning based methods will increase the differentiation between classifications so that the accuracy is improved.

References

[1] H. Alyami and Y. Mohamed, "Review and Development of MMC Employed in VSC-HVDC Systems," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, 2017, pp. 1-6.

[2] Hualei Wang and M. A. Redfern, "The Advantages and Disadvantages of Using HVDC to Interconnect AC Networks," *45th International Universities Power Engineering Conference UPEC2010*, Cardiff, Wales, 2010, pp. 1-5.

[3] J. V. V. N. Bapiraju and P. Manohar, "Fault Estimation with Analytical Cable Model for MMC-HVDC in Offshore Applications," *2017 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, Bangalore, 2017, pp. 1-6.

[4] A. Kumar, S. Jhampati and R. Suri, "HVDC Converter Stations Design for LCC Based HVDC Transmission System-Key Consideration," *2017 14th IEEE India Council International Conference (INDICON)*, Roorkee, 2017, pp. 1-6.

[5] L. Cao, Y. Xia, J. Wang, G. Zheng, Y. Shen and T. Shan, "Aviation Bearing Fault Diagnosis Method based on CHSMM," *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, Harbin, 2017, pp. 1-5.

[6] J. Huang, G. Chen, L. Shu, S. Wang and Y. Zhang, "An Experimental Study of Clogging Fault Diagnosis in Heat Exchangers Based on Vibration Signals," in *IEEE Access*, vol. 4, pp. 1800-1809, 2016.

[7] G. Sun, Q. Hu, Q. Zhang, A. Qin and L. Shao, "Fault Diagnosis for Rotating Machinery based on Artificial Immune Algorithm and Evidence Theory," *The 27th Chinese Control and Decision Conference (2015 CCDC)*, Qingdao, 2015, pp. 2975-2979.

[8] H. Gu, W. Dong, X. Sun and X. Xu, "Fault Diagnosis for ZPW2000A Jointless Track Circuit Compensation Capacitor based on K-fault Diagnosis," *Proceedings of the 32nd Chinese Control Conference*, Xi'an, 2013, pp. 6305-6312.

[9] A. Qin, Q. Hu, Y. Lv and Q. Zhang, "Concurrent Fault Diagnosis Based on Bayesian Discriminating Analysis and Time Series Analysis with Dimensionless Parameters," in *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2254-2265, 15 March15, 2019.

[10] P. K. Sinha, F. B. Zhou and R. S. Kutiyal, "Fault Detection in Electromagnetic Suspension Systems with State Estimation Methods," in *IEEE Transactions on Magnetics*, vol. 29, no. 6, pp. 2950-2952, Nov. 1993.

[11] W. Xiaomeng and R. Zhang, "A Sensor Fault Diagnosis Method Research Based on Wavelet Transform and Hilbert-Huang Transform," *2013 Fifth International Conference on Measuring Technology and Mechatronics Automation*, Hong Kong, 2013, pp. 81-84.

[12] S. N. Rekha, P. A. Jeyanthy and D. Devaraj, "Wavelet Transform based Open Circuit Fault Diagnosis in the Converter used in Wind Energy Systems," *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, Srivilliputhur, 2017, pp. 1-4.

[13] Y. X. Zhong, H. L. Fan, J. P. Lu, L. Pang and Y. F. Li, "Research on Fault Diagnosis of Rolling Bearing Based on Wavelet Packet Transform and IPSO-SVM," *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Bangkok, 2018, pp. 1682-1686.

[14] Y. Wang and C. Fan, "Study on Fault Diagnosis Method of Emulsifier Based on Empirical Wavelet Transform and SVM," *2018 5th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS)*, Hangzhou, 2018, pp. 404-407.

[15] A. Ben Ayed, M. Ben Halima and A. M. Alimi, "Adaptive Fuzzy Exponent Cluster Ensemble System Based Feature Selection and Spectral Clustering," *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Naples, 2017, pp. 1-6.

[16] K. Kalti and M.A. Mahjoub, “Image Segmentation by Gaussian Mixture Models and Modified FCM Algorithm,” The International Arab Journal of Information Technology, vol. 11, no. 1, Jan 2014.

[17] K. Wu and Jiang Ke, "A Scheme of Real-time Traffic Classification in Secure Access of Power Enterprise Based on Improved Naive Bayesian Classification Algorithm," *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, 2016, pp. 1017-1021.

[18] X. Wang and X. Sun, "An Improved Weighted Naive Bayesian Classification Algorithm Based on Multivariable Linear Regression Model," *2016 9th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, 2016, pp. 219-222.

[19] S. S. Gavankar and S. D. Sawarkar, "Eager decision tree," *2017 2nd International Conference for Convergence in Technology (I2CT)*, Mumbai, 2017, pp. 837-840.

[20] A. Abdelhalim and I. Traore, "A New Method for Learning Decision Trees from Rules," *2009 International Conference on Machine Learning and Applications*, Miami Beach, FL, 2009, pp. 693-698.

[21] Zadeh, L. A. (1965) "Fuzzy sets", Information and Control, 8, 338–353.

[22] Z. Peng, M. Xiaodong, Y. Zongrun and Y. Zhaoxiang, "An Approach of Fault Diagnosis for System Based on Fuzzy Fault Tree," *2008 International Conference on MultiMedia and Information Technology*, Three Gorges, 2008, pp. 697-700.

[23] H. Fang and C. Xia, "A Fuzzy Neural Network Based Fault Detection Scheme for Synchronous Generator with Internal Fault," *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, Tianjin, 2009, pp. 433-437.

[24] B. Paily, S. Kumaravel, M. Basu and M. Conlon, "Fault Analysis of VSC HVDC Systems Using Fuzzy Logic," 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Kozhikode, 2015, pp. 1-5.

[25] A. I. Moustapha and R. R. Selmic, "Wireless Sensor Network Modeling Using Modified Recurrent Neural Networks: Application to Fault Detection," in *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 5, pp. 981-988, May 2008.

[26] C. Lin and E. Boldbaatar, "Fault Accommodation Control for a Biped Robot Using a Recurrent Wavelet Elman Neural Network," in *IEEE Systems Journal*, vol. 11, no. 4, pp. 2882-2893, Dec. 2017.

[27] T. de Bruin, K. Verbert and R. Babuška, "Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 523-533, March 2017.

[28] M. Aminian and F. Aminian, "Neural-network Based Analog-Circuit Fault Diagnosis Using Wavelet Transform as Preprocessor," in *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 47, no. 2, pp. 151-156, Feb. 2000.

[29] F. Lin, I. Sun, K. Yang and J. Chang, "Recurrent Fuzzy Neural Cerebellar Model Articulation Network Fault-Tolerant Control of Six-Phase Permanent Magnet Synchronous Motor Position Servo Drive," in *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 1, pp. 153-167, Feb. 2016.

[30] D. F. Akhmetov, Y. Dote and S. J. Ovaska, "Fuzzy Neural Network with General Parameter Adaptation for Modeling of Nonlinear Time Seriess," in *IEEE Transactions on Neural Networks*, vol. 12, no. 1, pp. 148-152, Jan. 2001.

[31] L. Qu, H. Zhou, C. Liu and Z. Lu, "Study on Multi-RBF-SVM for Transformer Fault Diagnosis," *2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, Wuxi, 2018, pp. 188-191.

[32] Y. X. Zhong, H. L. Fan, J. P. Lu, L. Pang and Y. F. Li, "Research on Fault Diagnosis of Rolling Bearing Based on Wavelet Packet Transform and IPSO-SVM," *2018 IEEE International Conference on Industrial Engineering and Engineering Management*

(IEEM), Bangkok, 2018, pp. 1682-1686.

[33] L. Qu and H. Zhou, "The Multi-class SVM Is Applied in Transformer Fault Diagnosis," *2015 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, Guiyang, 2015, pp. 477-480.

[34] Ramesh, M. , & Laxmi, A. J. . (2012). Fault Identification in HVDC Using Artificial Intelligence — Recent Trends and Perspective. *International Conference on Power*. IEEE.

[35] Moshtagh, J. , Jannati, M. , Baghaee, H. R. , & Nasr, E. . (2008). A novel approach for online fault detection in HVDC converters. *Power System Conference*. IEEE.

[36] Sanjeevikumar, P. , Paily, B. , Basu, M. , & Conlon, M. . (2014). Classification of Fault Analysis of HVDC Systems Using Artificial Neural Network. *Power Engineering Conference*. IEEE.

[37] Buigues, G. , Valverde, V. , Zamora, I. , Larruskain, D. M. , & Iturregi, A. . (2015). DC Fault Detection in VSC-based HVDC Grids Wsed for The Integration of Renewable Energies. *2015 International Conference on Clean Electrical Power (ICCEP)*. IEEE.

[38] Mitra, B. , Chowdhury, B. , & Manjrekar, M. . (2016). Fault Analysis and Hybrid Protection Scheme for Multi-terminal HVDC Using Wavelet Transform. *North American Power Symposium*. IEEE.

[39] Y. Melo, W. Neves and D. Fernandes, "Fault Detection and Localization for HVDC Transmission Lines," *2018 Simposio Brasileiro de Sistemas Eletricos (SBSE)*, Niteroi, 2018, pp. 1-5.

[40] Yeap, Y. M. , & Ukil, A. . (2016). Fault Detection in HVDC System Using Short Time Fourier Transform. *Power & Energy Society General Meeting*. IEEE.

[41] Irnawan, R. , Srivastava, K. , & Reza, M. . (2015). Fault Detection in HVDC-Connected Wind Farm with Full Converter Generator. *International Journal of Electrical Power & Energy Systems, 64*, 833-838.

[42] Murthy, P. K. , Amarnath, J. , Kamakshiah, S. , & Singh, B. P. . (2009). Wavelet Transform Approach for Detection and Location of Faults in HVDC System. *Industrial and Information Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference on*. IEEE.

[43] M. Chen, S. Lan and D. Chen, "Machine Learning Based One-Terminal Fault Areas Detection in HVDC Transmission System," *2018 8th International Conference on Power and Energy Systems (ICPES)*, Colombo, Sri Lanka, 2018, pp. 278-282.

[44] Lipton, Z. C. . (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning. *Computer Science*.

[45] Huang, G., Huang, G. B., Song, S., & You, K. (2015). Trends in Extreme Learning Machines: A Review. Neural Networks, 61, 32-48.

[46] Huang, G. B. (2014). An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. Cognitive Computation, 6(3), 376-390.