Milad Niaz Azari, Mohsen Sedighi, Mehdi Samami

Intelligent Fault Detection in Power Distribution Systems Using Thermos-grams by Ensemble Classifiers

In today's world, many companies use the thermal imaging (infrared), in order to prevent failures and improve the reliability of the electrical networks. In fact, the technical inspection of the electrical equipment using thermal cameras, is the most effective method for preventive defect detection. This contribution deals with, a systematic method in which, areas suspected of failure, are identified through computer-aided thermal image processing. To this end, the candidate areas are determined, using adaptive threshold and, a number of features are extracted from them. Next, using a genetic algorithm (GA), the irrelevant features are omitted. Finally, by means of a hybrid classifier, the pattern of positive and false positive areas, have been identified. This classifier can also be used as a filter, after extracting the candidate areas. This method is tested on images taken from Tehran northwest substations. As a result, applying the feature selection algorithm leads to a faster intelligent fault detection and higher Reliability, especially in widespread networks, which is known as an effective validation for the proposed method.

Key words: Thermal Imaging, Defect Detection, Pattern Recognition, Zernike Moments

Inteligentna detekcija greške u sustavu distribucije električne energije korištenjem termalnih slika i grupe klasifikatora. U suvremenom svijetu mnoga poduzeća koriste termalne slike (infracrvene) kako bi se spriječili kvarovi i popravila pouzdanost električne mreže. Zapravo, tehnički pregled električke opreme korištenjem termalnih kamera je najučinkovitija metoda za detekciju i prevenciju kvarova. U ovom radu proučava se sistematična metoda u kojoj se sumnjiva područja identificiraju korištenjem računalne obrade termalne slike. Nakon što se odrede moguća područja korištenjem prilagodljivih pragova iz njih se izvode mnoga svojstva. Zatim se korištenjem genetičkih algoritama izostavljaju nevažna svojstva. Konačno, korištenjem hibridnog klasifikatora identificiraju se uzorci pozitivnih i lažno pozitivnih područja. Taj klasifikator može se koristiti i kao filter nakon izdvajanja mogućih područja. Ova metoda testirana je na slikama teheranske sjevernozapadne stanice. Korištenje algoritma za selekciju svojstava dovodi do brže i inteligentne detekcije te veće pouzdanosti, posebno kod rasprostranjenih mreža koje služe za učinkovitu validaciju predloženih metoda.

Ključne riječi: termalne slike, detekcija kvara, raspoznavanje uzorka, Zernike momenti

1 INTRODUCTION

Repair and maintenance is known as an important factor that affects the productivity and increases the efficiency of an industrial system through reduction of equipment failure duration. Various investigations have been conducted by focusing on this issue in power networks and accordingly, different methods are proposed. One of the current common methods is visual inspection of transmission lines and distribution networks. The disadvantage of this method is insignificant precision beside the high cost and the Perpetual need for an expert's opinion. This study proposes thermo-vision technology in order to improve the current methods for repair and maintenance in power networks.

Due to proliferation of various loads during the time,

and also the high cost and complexity of power network expansion, providing proper strategies in order to keep the existing lines in their maximum efficiency seems to be vital. For this reason, continuous inspection and maintenance of transmission lines is needed. It should be stated that, the climatic factors, the loss of electrical properties of materials and poor quality control during construction processes are the leading reasons for this continuous monitoring. During the last years inspection-maintenance systems have developed due to economic incentives and the decrease in human risks associated with this issue. On the other hand, the expansion of the electric power industry and its vital role in the development of countries, caused a new attitude, based on productivity enhancement and higher efficiency, particularly in connection with dis-

tribution networks. So far, various methods have been suggested in order to obtain a better monitoring of the power networks including, repairs after failures, repairs with a productivity-based attitude and preventive repair and maintenance. But nowadays, the attitude based on prevention before failures, is considered as the main attitude in a repair and maintenance system. To this end, many companies around the world use the thermal imaging (infrared). In fact, the technical inspection of electrical equipment using thermal cameras, is the most effective method for preventive diagnostics which causes the reliability improvement of power networks.

Various researches have been conducted for diagnosis of faults in transmission lines. Ref [1], proposes the use of ultrasonic waves in order to monitor the transmission lines. Ref [2], discusses about a number of sensitive and important technologies, such as: the electric field and infrared acoustic sensors. In [3], monitoring process is done by installation of optical-fiber cables, along the transmission lines and with the help of standard helicopters which detect the faults along the lines by monitoring the changes in the electric field. The other way to perform this type of inspection is, putting trained explorers into the helicopters in order to inspect the lines with the use of cameras and photography [4]. In this way the information can be recorded in a log book. It should be stated that, the problem of this method includes, the feature extraction of the target position, the camera stability, finding and keeping the target in the camera field of view and the data analysis system. In ref [5], a three-dimensional simulation and an optimal design for a robot is presented. It should be noted that, all the above mentioned methods are not intelligent. As a result, it can be a serious motivation in order to lead the researchers toward intelligent methods. In [6], the location of the fault in transmission lines with underground cables is determined with the help of neuro-fuzzy systems. It is noteworthy that, neural networks and fuzzy systems have shown accurate results, distinguishing the type of faults, as well as determination of fault location in many studies [7, 8]. But the main disadvantage of the neural network method is its complexity which results in simulation of a large number of fault cases in order to train the neurons. Such methods are memory intensive and require powerful processors. Several methods have been suggested for fault location in distribution systems based on substation estimated impedance [9, 10]. Generally, the diagnostic methods in electrical systems, are based on two significant methods including, model-based methods and pattern recognition methods. The first procedure requires heavy computations, and consequently few systems can be found which can easily be diagnosed by these methods. Especially, if the system model is complicated and non-linear [11, 12]. But the pattern recognition methods which are
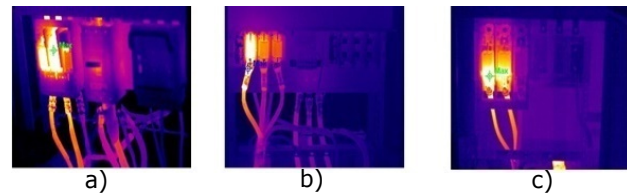

Fig. 1. Images, taken with infrared technique:(a) Fuse failure in Phases R and S with a maximum temperature of 78 degrees; (b) Cable shoe failure in Phase R with a maximum temperature of 91 degrees; (c) Lower jaw failure in Phases R and S with a maximum temperature of 50 degrees

also called knowledge-based methods, do not require the system analytical model, instead they are very sensitive to the training data. For this reason, computations will be easier compared the model-based methods.

As previously mentioned, the technical inspection of electrical equipment using thermal cameras, is the most effective method for preventive diagnostics. On the other hand, analysis and interpretation of thermal images are usually performed traditionally and by means of eyes. By considering the extent, diversity, and dispersion of distribution networks, the need for a proper systematic method in order to provide an intelligent identifications for various types of faults seems to be really significant. This paper presents an intelligent and systematic pattern recognition method, using real images prepared from distribution substations of Tehran northwest network in order to achieve a hybrid model for an intelligent fault detection.

2   AN INTELLIGENT FAULT DIAGNOSTIC SYSTEM DESIGN

The current method, which is used for fault location in all distribution companies, is completely manual, experimental and non-engineering. Therefore, this study is aimed at, providing an intelligent fault detection method, for images which are taken from distribution networks with the help of infrared imaging technique instead of investigating them, case by case and manually. The interpretation and analysis of images and eventually fault detection, are performed quite intelligently and by computers. Obviously, due to extent and dispersion of distribution networks, the need for such a system which is a substitute for the current fault diagnostic methods seems to be inevitable. In this paper, a method is offered, using real images prepared from distribution substations in Tehran northwest network, as shown in Fig.1.

Figure 2 shows the block diagram of the proposed method in this paper. In the following, we examine each of the modules of this block diagram.
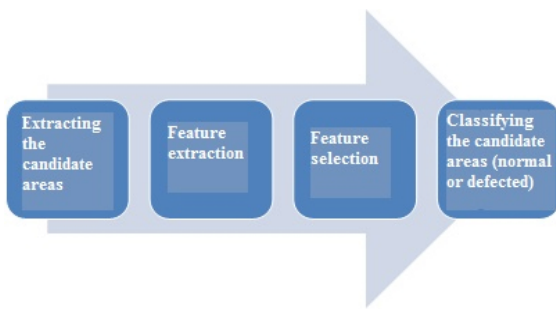
Fig. 2. The designing process block diagram of a visual intelligent fault diagnostic system

2.1 Extracting candidate areas, with the help of image processing

Fortunately, due to high capability of thermo-vision cameras, the quality of thermal images are good enough, thus pre-processing operation is not necessary. An important part for the image processing section, is the separation of defective parts from the rest of the image, so that, the faults can finally be separated from each other, according to the location and characteristics of defective parts. In accordance to the studies conducted on different instruments, investigation on the conditions in which failures occurred, the allowed temperature which was obtained and ultimately observing the samples of these images, it is obvious that the temperature of the defected point in the images, is naturally higher than the other points, and sometimes this difference is extremely high. Therefore, due to the characteristic of these cameras, by which the color of each part of an image is graded based on its temperature. It is clear that, the intensity of the color and brightness of the image in the defective section of components which is one of the four intended faults, is different from the other sections of the equipment. Thus we have done the separation process, according to the role of lighting and thresholding. Thresholding is one of the most important ways for image segmentation [13].

The easiest way for thresholding, is dividing the image histogram, using a threshold T, then by a pixel by pixel survey and labeling each pixel, as an object or background, depending on whether the gray scale of that pixel is larger or smaller than T, the image is segmented. As previously stated, the success of this method, completely depends on the quality of dividing the histogram.

In the gray-scale images which were extracted from the colored thermal images, the pixel values are in the range of 0 to 255. Therefore, determination of an optimal threshold value for the brightness, which is useful for this issue, is very important. In order to have an intelligent design and due to differences in various images (in terms of brightness), offering a method which is applied for all images

identically, along with a suitable threshold value is necessary. Therefore the images are not going to be investigated case by case. On the other hand, it is not necessary to draw a histogram for each one and also to conclude and determine the threshold value after interpreting the curves. The used method includes, examining the histogram of a number of images, and using, the mean brightness as the threshold value:

$$th = \frac{\max(I) + min(I)}{2} \qquad (1)$$

In (1), (I) is the pixels of the image, which represents the brightness. In fact, first, the maximum value of the image brightness is obtained as well as its minimum value, then the summation of the mentioned values will be divided by two.

After examining a number of thermal images, by applying (1), it is concluded that, the following threshold value should be used in order to separate the defective points of electrical equipment from their thermal images:

$$TH = 1.5 * th \qquad (2)$$

Now it is the time for image binarization. At this point, the pixels of gray-scale images are compared with the threshold value, and the brightness of pixels which is greater than the threshold limit, is considered to be equal to 1, and the brightness of pixels which is less than the threshold limit is considered to be equal to zero. In Fig .3, you can see a number of images and their conversion into grayscale, and finally their binarization. After obtaining the binary images, a morphological actuator called "Closing", is used in image processing, to fill small gaps within the binary objects.

2.2 Feature Extraction

In order to design a proper system for fault detection in electrical equipment, feature selection plays a fundamental role. For this reason Statistical methods; such as Zernike moment, are widely used, Zernike moment is a set of indicators, insensitive to rotation. By normalizing the image according to its parameters, and using the geometric moment, Zernike moment becomes insensitive to scale and transmission [14].

The invariant properties of moments, which are known as sensitive features of the samples, are useful in classification and diagnostic applications [15]. In order to extract features from original images, the relevant information around the defected points should be obtained. So that, the pixel values is considered to be equal to zero outside the defective position. In the next stage, the feature vector is obtained through calculating the Zernike moment from the
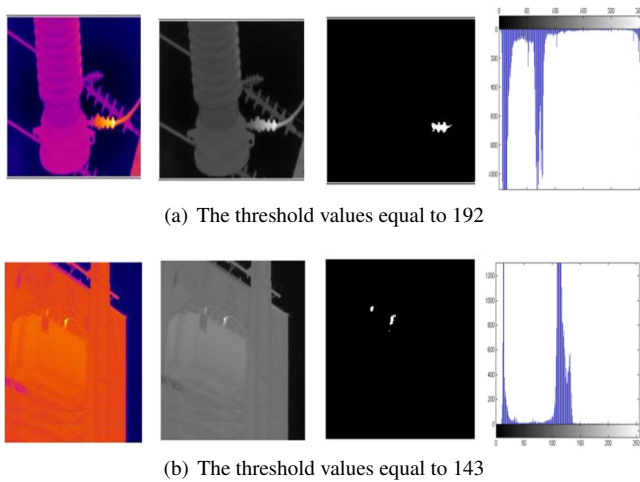
(a) The threshold values equal to 192

(b) The threshold values equal to 143

Fig. 3. Separation process of the defective section of electrical equipment, with different threshold values

sub-image of the previous stage. It should be noted that, in Subsection 2.1, the procedure of binarization and separation of the defective points of equipment, was described, and finally creating a sub-image for feature extraction was shown. In the sub-image, the information of different parts of the electrical equipment, which are not defective, have been omitted.

Zernike polynomials are an orthogonal set of complex polynomials [16]:

$$V_{nm}(x,y) = R_{nm}(x,y)\exp(jm\tan^{-1}(\frac{y}{x})) \quad (3)$$

where $|m| < n \ x^2 + y^2 \leq 1$, $n > 0$ and $n - |m|$ must be even.

It should be noticed that, the R_{nm} (polynomial) is defined as below:

$$R_{nm}(x,y) = \sum_{s=0}^{n-|m|} S_{n,|m|,\text{s}}(x^2 + y^2)^{n/2-s} \quad (4)$$

$$S_{n,|m|,\text{s}} = (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2} - s)!(\frac{n-|m|}{2} - s)!} \quad (5)$$

The n-order complex Zernike moments with m iterations, are obtained through the following formula:

$$ZM_{nm} = \frac{n+1}{\pi} \sum_{x} \sum_{y} f(x,y)V_{nm}^*(x,y) \quad (6)$$

In the calculation of Zernike, the center of the image is placed in the center of the circle, and the rest of the image is mapped into the circle. Those parts, locating outside the circle, do not enter the calculation.

After the introduction of Zernike conversion, in the previous subsection, the feature extraction from the images of the database is studied, using Zernike conversion (in fact, the binarized images, in which the position of the defective part, is separated from the rest of the image). The order of the moments which are used, and their combination to form the feature vector for each image, are significant. The method, which is used here is described as follows; first The Zernike moment for different orders (n) from 1 to 15 is calculated, It is noteworthy that, for each order of the moment, all possible (m)s have also been considered. Table 1 shows how to form a feature vector. Then, Zernike general feature vector forms from combining the vectors, obtained from each stage. Of course, it should be noted that, the aim of this section is only forming a feature vector, and this issue that, which one is the desirable feature, and what elements of the feature vector are more effective in designing a classifier, are not discussed here. In the next chapter, the procedure of selecting the desirable features, will be described.

According to Table 1, at each stage, for each (n) and its corresponding (m)s, Zernike moments have been calculated, and they have been located next to each other in the form of vectors. The number of vectors for each stage, which actually represents the feature vector, is also specified in the mentioned table. For example, by assuming, m=1 and n=2; the number of the elements of the feature vector equal to 1, and for m= 0, 2 and n=2; the mentioned amount equal to 2. Finally, Zernike general feature vector forms from combining the vectors of each stage together.

In this paper, four other features are extracted in addition to Zernike features. These features are extracted from the grayscale of the candidate area. The first feature of the mean gray levels, is the pixels of suspicious area. The second feature is the variance of these gray levels. The third feature is kurtosis[1], and the fourth feature is the skewness[2] of these gray levels. By putting these four features together with 71 Zernike features which were obtained from Table.1, 75 features are extracted altogether, which will form the general feature vector.

2.3 Feature Selection

In order to reduce false positive samples, a learning algorithm should be used, so that, it can learn the features of the samples appropriately, and also can be employed as an appropriate filter. First it should be noted that, all extracted features may not be appropriate for this area. Therefore, first, a feature selection algorithm is applied to reduce the

---

[1]kurtosis
[2]skewness

Table 1. The elements of the feature vector based on Zernike moments

The number of the elements of the feature vector	The number of iterations	The order of Zernike moment
1	m=1	n=1
2	m=0,2	n=2
2	m=1,3	n=3
3	m=0, 2, 4	n=4
3	m=1, 3, 5	n=5
4	m=0, 2, 4, 6	n=6
4	m=1, 3, 5, 7	n=7
5	m=0, 2, 4, 6, 8	n=8
5	m=1, 3, 5, 7, 9	n=9
6	m=0, 2, 4, 6, 8, 10	n=10
6	m=1, 3, 5, 7, 9, 11	n=11
7	m=0, 2, 4, 6, 8, 10, 12	n=12
7	m=1, 3, 5, 7, 9, 11, 13	n=13
8	m=0, 2, 4, 6, 8, 10, 12, 14	n=14
8	m=1, 3, 5, 7, 9, 11, 13, 15	n=15

dimension (to reduce the number of features) and also to omit irrelevant and redundant features. In this regard, the genetic algorithm is combined with a criterion called CFS [17] to enhance the performance of the algorithm. Thus, in order to increase the precision of classification, a number of classifiers, using the voting mechanism should be combined, because it has been shown that, combining the classifiers will increase their generalizability, and therefore, they will act better in learning the real-world problems.

2.3.1 Feature Selection Algorithm

After extracting all the aforementioned features, 75 features will be obtained altogether. Now, the number of these features should be reduced, in order to select an appropriate subset of them. Among all the reasons for a reduction in the dimension, the following items can be indicated:

- In the majority of machine learning algorithms, the complexity of the algorithm depends on the number of features (dimension) as well as the number of samples (n). The dimension is reduced in order to reduce the computational complexity and to reduce the required memory. The reduction in the dimension, reduces the complexity of the inference algorithm at the

testing time. For instance, when a rule-based classifier is used, due to reducing the dimension, some rules will be obtained which are more understandable and simpler.

- When a feature is recognized as an irrelevant feature, there will be no need for its extraction, anymore.

- Simpler models are more resistant to smaller data sets, which means that, they will be less sensitive to noise, outlying points and etc.

- By displaying data in fewer dimensions, it can be analyzed, without losing any information.

- As will be shown in the results section, the presence of irrelevant features, will reduces the precision of the classifier.

As mentioned previously, one of the well-known criteria for selecting a proper subset of features, is a criterion called CFS. According to this competence criterion, subsets whose members are highly correlated with the class label and, on the other hand, are not dependent to each other, is more appropriate. According to this principle, the irrelevant features must be omitted, because they have little correlation with the class label of the samples. It should be noted that, redundant features should also be omitted, because they are highly correlated with one or more features, and using one feature is sufficient instead of all of them. The equation corresponding to this competence criterion is as below:

$$M_s = \frac{kT_{cf}}{\sqrt{k + k\,(k-1)\,T_{ff}}} \qquad (7)$$

where M_s is the heuristic criterion of the competence of a set of features "S" including k features. $\overline{r}_{cf}$ is the class-feature mean correlation ($f \epsilon S$), and $\overline{r}_{ff}$ is the feature-feature mean correlation.

The numerator represents the capability of features in predicting the class label, and the denominator represents the redundancy among features. In order to search for the appropriate subset, various search strategies are used, such as: best-first search strategy and the genetic algorithm. Best-first search strategy is a greedy technique for achieving a semi-optimal solution. Best-first search strategy is an artificial intelligence search strategy which is allowed to turn back in the search path.

The genetic algorithm is another search technique which can be used for selecting an appropriate subset. For solving a problem, using a genetic algorithm, there are 2 basic steps: 1. Definition of chromosome 2. Definition of fitness function.

Genetic algorithms use the principles of Darwinian natural selection to find the optimal formula for prediction or pattern matching. Genetic algorithms are often a good option for regression-based predictive techniques. In short, it is said that the genetic algorithm (GA) is a programming technique which uses the genetic evolution as a problem-solving model. The problem which has to be solved consists of the input, the solutions which are encoded according to a model, and the metric function which is also called the fitness function [18]. The genetic algorithm (GA) is a searching technique in the computer science, for finding optimal solutions and searching problems. Genetic algorithms are considered as an important type of evolutionary algorithms which has been inspired by biological sciences, such as: inheritance, mutation, sudden selection, natural selection, and composition.

There are different methods for using genetic algorithms to choose a subset of features [19- 21]. In this paper, the binary vector optimization method is used. In this method, an optimal binary vector is obtained, which is proportional to selecting a feature. If a feature is selected, Bit 1 will be recommended, and if a feature is not selected, Bit 0 will be recommended. The final aim is to achieve a binary vector with the minimum number of (1)s, in order to maximize the CFS criterion.

As mentioned earlier, 75 features have been extracted, thus the chromosomes have 75 genes. If the value of a gene is equal to 0, it means that the feature corresponding that gene has not been selected, and if this value is equal to 1, it means that the feature corresponding that gene has been selected. It should be noted that, the criterion applied in CFS method, can be used in definition of fitness function.

2.3.2 A Compound Classifier Design

There are different voting mechanisms for combining the votes of classifiers such as; unanimous voting, majority voting, and mean weighted voting [22]. In this paper, the weighted voting mechanism is used. In this type of voting the classifiers which have higher precision, have different voting rights depending on their precisions. The classifiers introduced in this section (Neural networks, RBF, SVM, Bagging, Ada-Boost, and rotation forest) are combined, based on a weighted voting mechanism.

3 TESTS AND RESULTS

3.1 The Candidate Areas Identification

As mentioned in the previous section, the entire fault detection system can be divided into two sections. In the first section, the candidate fault areas are identified, which is performed through the proposed thresholding. The remarkable point is that, this section determines the maximum sensitivity of the system. This means that, if a fault

is not identified at this stage, it no longer can be identified by the system. The total number of faulty images (and reports had been prepared for them) is about 52. In some of these images, more than one fault have been reported. The total number of the faults, existing in these images, is about 59. After performing the processing step, 178 areas are identified as faulty areas. Obviously, all of these areas are not known as defected areas, and only some of them are really defected. In fact, among these 178 areas, 55 areas have been correctly identified as defected areas, and 123 areas have been misidentified (false positive).

3.2 The False Positive Rate reduction

As it is clear, more than three quarters of the identified areas have been misidentified. Now, using machine learning techniques, mentioned in the previous chapter, an attempt is done in order to design a smart filtering to reduce the false positive error rate. In this regard, the precision of different classifiers on these 178 areas is reported. In these tests, 10-fold cross validation strategy have been used. In this strategy, all the data are divided into 10 parts which are approximately equal. In the first iteration, the first 10% are put aside as a testing set, and the classifier are trained on the remaining 90%. In the second iteration, the second 10% are put aside as a testing set, and the classifier are trained on the remaining 90%. This operation is iterated up to 10 times, and at the end, the general precision represents the precision of the classifier.

In order to evaluate a model, a good concentration on the predictive capability of that model is needed. On the other hand, in order to illustrate the form of classification, a confusion matrix[3] is usually used, as in Table 2.

Table 2. Confusion matrix

The predicted class		
The actual class	*Class = fault*	*Class = no fault*
Class = fault	*a (TP)*	*d (FN)*
Class = no fault	*b (FP)*	*c (TN)*

A set of criteria is extracted from the confusion matrix, which can represent the efficiency of a model. The best known criterion is the accuracy[4] criterion, which is defined as follows:

$$Accuracy = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+FP+TN+FN} \tag{8}$$

The accuracy criterion suffers from a big limitation. Consider a 2-fold class problem, in which the number of

---

[3]Confusion matrix

[4]accuracy

the samples of the "no fault" class, equals 9990, and the number of the samples of the "fault" class, equals 10. Now, assume that, a model assigns the "no fault" label, to any samples which are given to it. In other words, it recognizes all the areas as "no fault". It is clear that, this model has acted poorly, but when the accuracy criterion is calculated for this model, the obtained accuracy will equal to 10000/9990 = 99.9. Therefore, it can be concluded that the accuracy criterion is a misleading criterion, because the aforesaid model has found none of the samples of the "fault" class, achieving a significant accuracy. In unbalanced problems, such as fault detection, in which the number of samples of the negative class is really higher than the number of samples of the positive class, the accuracy criterion is not considered as an appropriate criterion. The other criteria include; precision[5], recall[6] and F criterion:

$$Precision\ (p) = \frac{a}{a+c} \qquad (9)$$

$$recall\ (r) = \frac{a}{a+b} \qquad (10)$$

$$F - measure = \frac{2pr}{p+r} = \frac{2a}{2a+b+c} \qquad (11)$$

The accuracy criterion represents this fact that, what percentage of the areas which have been detected by the relevant model, as the faulty area, are really faulty. But recall means that, how many percent of the areas have been found correctly, by the relevant model. But what would happen if a model has a higher precision compared with another model, but still has a higher recall? The solution for this problem, is F criterion. F criterion merges precision and recall criteria. In fact, F criterion has more concentration on the correct classification of positive samples (fault). In the other word, it pays more attention to the definition of precision and recall criteria. For this reason, in this section, the F criterion are applied as the main criterion for model assessment. As can be seen in Table 3, after combining the classifiers through voting, a higher precision is achieved compared with each of them individually.

3.3 Assessment of the Feature Selection Algorithm

As explained in the previous section, a genetic algorithm is selected, as a search strategy, which uses a criterion known as CFS, as a fitness function. In addition to the fitness function, the genetic algorithm also has other parameters, whose values are mentioned in Table 4. It is worth mentioning that, the mutation and the cross-over rate are considered exactly as our default "weka", and on the other hand, the number of iterations and the initial population are completely arbitrarily.

---
[5]precision
[6]Recall

Table 3. The performance of the base classifiers, used in the proposed classifier, individually, and together

F criterion	Recall	Precision	Accuracy	Classifier
92	88.6	95.6	89.3	SVM
85.4	90.2	81	78.6	RBF Neural Network
90	87.8	92.3	86.5	Ada-Boost
92.2	91.9	92.6	89.3	Bagging
91.9	91	92.9	89	Rotation Forest
93.4	92.7	94.2	91	A combination of the above classifiers

Table 4. The parameters of the genetic algorithm

Population size	# of iterations	Cross over Probability	Mutation probability
500	2000	60%	3.3%

After applying the feature selection algorithm, only 3 features are selected among the 75 extracted features, which speeds up the learning and testing, so that creating the proposed classifier (the compound classifier), considering 75 features, takes 0.89 seconds, while by using these 3 features, it only takes 0.06 seconds. Now, if the number of images increases, this difference will become more obvious. On the other hand, omitting the irrelevant and redundant features, increases the precision of the classifier. As can be seen in Table 5, by using only the 3 selected features rather than 75 features, the performances of all the classifiers have been improved. Thus, it can be concluded that, the proposed feature selection algorithm is a resistant algorithm, which is not dependent on the type of the classifier which is in use.

Table 5. The effect of the feature selection algorithm on classifiers

F criterion	Recall	Precision	Accuracy	Classifier
92.9	91.1	94.9	4.90	SVM
92.1	90.2	94.1	89.3	RBF Neural Network
90.2	89.4	90.9	86.5	Ada-Boost
93.2	93.5	93	91.5	Bagging
91.3	89.4	93.2	88.2	Rotation Forest
93.8	92.7	95	91.57	A combination of the above classifiers

4   CONCLUSION

This contribution, has offered an intelligent method for fault diagnosis in distribution network substations. The thermal images which are taken from electrical equipment with the help of infrared imaging technique, are used as input data, and the features of the mentioned images are extracted, using Zernike moments and the features of gray levels. Subsequently, a feature selection algorithm is applied to reduce the number of features and also to omit the irrelevant features. In this regard, the genetic algorithm is combined with a criterion called CFS to enhance the performance of the algorithm. Ultimately, the faults existing in electrical equipment, have been detected intelligently, using a combination of the classifiers such as; SVM, RBF neural network, forest rotation, Ada-Boost and Bagging, based on the weighted voting strategy. The combination process is done in order to increase the precision of classification which causes the enhancement of the classifiers generalizability. As a result, applying the feature selection algorithm leads to a faster fault detection. The usefulness of the method can be more obvious, considering this fact that, the diversity and dispersion of the distribution networks are very high. Therefore, this investigation, can be implemented in the maintenance and management of the regional power networks, in order to increase the efficiency through reducing the blackouts which leads to decrement of costs.

REFERENCES

[1] Mike S, Wilson and Stefan Hurlebaus, "Power Line Monitoring",Proceedings of the 2007 American Control Conference.

[2] Bing Jiang, "Robotic Monitoring of Power Systems," IEEE Transaction on Power Delivery, Vol. 19, No. 3, July 2004

[3] José Rocha Jo aoSequeira, "The Development of a Robotic System for Maintenance and Inspection of Power Lines".

[4] Lew Rubin, "The future of power line inspection", Electrical World, http://www.Platts.com/engineering/issues/ Electrical World /0009/0009 ew-line inspection.shtml, October 2000.

[5] Serge Montambault, Nicolas Pouliot, "Development of an Inspection Robot Capable of Clearing Obstacles While Operating on a Live Line", IEEE 2006.

[6] C.K. Jung-K.H. Kim-J.B. Lee-and Bernd Klöckl, "Wavelet and neuro-fuzzy based fault location for combined transmission systems", International Journal of Electrical Power & Energy Systems, Volume 29, Issue 6, July 2007, Pages 445-454

[7] H. Wang and W.W.L. Keerthipala, "Fuzzy–Neuro approach to fault classification for transmission line protection", IEEE Trans Power Deliver 13 (1998) (4), pp. 1093–1104.

[8] D.V. Coury and D.C. Jorge, "Artificial neural network approach to distance protection of transmission lines", IEEE Trans Power Deliver 13 (1998) (1), pp. 102–108.

[9] R. Das, "Determining the locations of faults in distribution systems", Ph.D.Dissertation, University of Saskatchewan, Saskatoon, Canada, 1998.

[10] A. Warrington and C. Van, "Protective Relays, Their Theory and Practice", Vol. 1, Chapman and Hall, Ltd, London 1968.

[11] T.Sorsa, H.N.Koiro, "Neural Network in Process Fault Diagnostics", IEEE Trans on Sys, Man @ Cyberneties, Vol. 21, No. 4 Jul/Aug.1991.

[12] Y.Sekine, Y.Akimoto, "Fault Diagnosis of Power System", Proceeding of the IEEE, Vol.80, No.5, May.1992.

[13] R. C. Gonzalez and R. E. Woods, "Image Segmentation," in Digital Image Processing, Second Edition, Upper Saddle River,New Jersey: Prentice Hall, 2002, pp. 665-675.

[14] A.Khotanzad and Y.H.Hong, "Invariant Image Recognition by Zernike Moments", IEEE Trans, Analysis and Machine Intelligence, Vol 12, No. 5 May 1990.

[15] S.O.Belkasim, M.Shridhar and M. Ahmadi, "Pattern Recognition with Moment Invariants: A Comparative Study and New Results," IEEE Trans.On Pattern Recognition, 24 (12), 1117-1138, 1991.

[16] C.H.Theand, R.T.Chin,"On Image Analysis by the Methods of Moments", "IEEE Trans, on Pattern Analysis and Machine Intelligence, 10 (4), 496-513, 1988.

[17] M. A. Hall, "Correlation-based Feature Selection for Machine Learning," Ph.D. dessertation, Department of Computer Science, university of Waikato 1999, pp. 51-73

[18] M.Srinivas, L.M.Patnik, "Genetic algorithms: asurvey", IEEE, 1994.

[19] A.Khotanzad and Y.H.Hong, "Invariant Image Recognition by Zernike Moments", IEEE Trans, Analysis and Machine Intelligence, Vol 12, No. 5 May 1990.

[20] S.O.Belkasim, M.Shridhar and M. Ahmadi, "Pattern Recognition with Moment Invariants: A Comparative Study and New Results," IEEE Trans.On Pattern Recognition, 24 (12), 1117-1138, 1991.

[21] C.H.Theand, R.T.Chin,"On Image Analysis by the Methods of Moments", "IEEE Trans, on Pattern Analysis and Machine Intelligence, 10 (4), 496-513, 1988.

[22] I. L. Kuncheva, Combining Pattern Classifiers, First Edition, Hoboken, New Jersey: John Wiley & Sons Inc., 2004, pp. 203-215.

Milad Niaz Azari was born in Babol, Iran, in 1984. He received his B.S degree in electrical engineering from Noshirvani University of Technology, Babol, Iran in 2007. He Graduated in M.S degree in the department of Electrical Engineering, Amir Kabir University of Technology, Tehran, Iran in 2009. Also he graduated in PhD degree in Amirkabir University of Technology, Tehran, Iran in 2013. Since 2015, he has been at University of Science and Technology of Mazandaran , Behshahr, Iran, as an assistant Professor in the department of Electrical Engineering. His areas of interest are electrical machines design and power electronics.

Mohsen Sedighi was born in Tehran, Iran, in 1985. He received his B.S degree in electrical engineering from Sarab Branch, Azad University, Sarab, Iran. and M.Sc. degree in the department of Electrical Engineering, Sari Branch, Azad University, Sari, Iran in 2014. . He is currently a PhD student in the department of Electrical Engineering, Damghan Branch, Islamic Azad University, Damghan, Iran. His areas of interests are operation and maintenance of power systems and distribution network protection.

Mehdi Samami was born in Nowshahr, Iran, in 1984. He received his B.S degree in electrical engineering from South Tehran Branch, Islamic Azad University, Tehran, Iran, in 2007. He Graduated in M.S degree in the department of Electrical Engineering, Mazandaran Science and Research Branch, Islamic Azad University, Sari, Iran in 2015. He is currently a PhD student in the department of Electrical Engineering, Sari Branch, Islamic Azad University, Sari, Iran. His research interests include analysis, design and protection of electrical machines.

AUTHORS' ADDRESSES

Asst. Prof. Milad Niaz Azari, Ph.D.
Department of Electrical Engineering,
University of Science and Technology of Mazandaran,
P.O. Box: 48518-78195, Behshahr, Iran,
Email: miladniazazari@mazust.ac.ir
Mohsen Sedighi, M.Sc.
Mehdi Samami, M.Sc.
Department of Electrical Engineering,
Sari Branch, Islamic Azad University,
Sari, Iran,
Email: mohsensadighi1364@gmail.com,
mehdisamami63@gmail.com