MACHINE LEARNING APPROACH FOR ANOMALY DETECTION IN WIRELESS SENSOR DATA

Ajay Singh Raghuvanshi[1], Rajeev Tripathi[2], and Sudarshan Tiwari[2]

[1]Department of Electronics and Communication Engineering, Indian Institute of Information Technology, IIITA, Allahabad, India.
[2]Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, India.

ABSTRACT

Wireless sensor nodes can experience faults during deployment either due to its hardware malfunctioning or software failure or even harsh environmental factors and battery failure. This results into presence of anomalies in their time-series collected data. So, these anomalies demand for reliable detection strategies to support in long term and/or in large scale WSN deployments. These data of physical variables are transmitted continuously to a repository for further processing of information as data stream. This paper presents a novel and distributed machine learning approach towards different anomalies detection based on incorporating the combined properties of wavelet and support vector machine (SVM). The time-series filtered data are passed through mother wavelets and several statistical features are extracted. Then features are classified using SVM to detect anomalies as short fault (SF) and noise fault (NF). The results obtained indicate that the proposed approach has excellent performance in fault detection and its classification of WS data.

KEYWORDS

Wireless Sensor Networks, Anomaly Detection, SVM, Wavelet Filters, data fault, fault detection

I. INTRODUCTION

Wireless sensor networks have already emerged as potential source in monitoring and thereby collection of information in remote geographical, industrial, civil infrastructures and even power plants. In fact, a large number of sensor nodes equipped with limited computing and communication abilities are deployed to monitor the variation of physical variables. Due to their uncontrolled use or harsh environment, they are sensible to various faults which may lead to abnormal data patterns in monitoring domain. Literatures [1], [2] and [3] have reported the existence of faulty data monitored by sensors in their deployment in field environment. This is said to be caused either due to defect in hardware design, improper calibration of sensors or low battery levels of sensor nodes. Also any change or uncertainty in the environment being monitored may lead to affect the distribution of data measurements. Anomaly detection in communication network traffic and use of wavelets to identify is proposed in [4] and role of wavelet analysis is studied in [5].

Due to continuous collection of data by wireless sensor network, it becomes cumbersome to aggregate them and difficult in detection of anomalies present. The data collection from wireless sensors can be managed at centralized or distributed level in the network. The centralized approach in study of data pattern/processing posses constraint to prolong life time of network, since limited battery power of nodes gets depleted even in transmission of anomalous signals. On other hand, in case of distributed approach, each node is meant to process the data collected and send the descriptive information to either other neighbouring nodes or base station.

Truly speaking, the research needs to be oriented towards automatic detection and classification of sensor data faults at collection point itself. The investigation on faulty sensor data gains its importance due to the fact that this would help in detection and thereby its elimination at sensor node level itself. This could enhance the battery operating life in sensor node since erroneous data need not be transmitted to the base station thus contributing towards energy efficiency of entire sensor networks. Thus, efficient anomalies detection measures need to be adopted at the node so as to raise the alert in the operating system. They need to have their performance insensitive to any parameter setting in the algorithm or any pattern change in time-series data. Additionally, it is also desired that the technique should involve low computational burden. It is crucial that a centralized network management tool embeds the required expert decision to detect all possible anomaly types, as the network is perceived holistically as an intelligent data delivery system. The design of such efficient and reliable tool demands a comprehensive understanding of all types of wireless sensor data anomalies, their likely causes, and their potential solutions.

This paper considers a study on anomalies detection and classification in wireless sensor data with use of discrete wavelet transform (DWT) and support vector machine (SVM) properties. The proposed approach does not utilize a huge amount of data in processing the information sought and efficiently detects and classifies the different types of fault with little processing time. It is aimed to detect and classify anomalies at node level according to the characteristics of data collected by each individual sensor.

The rest of the paper is organized as follows. In section 2, related work in the fault detection strategy is addressed, followed by methodology of proposed scheme with used techniques in section 3. The performance evaluation and discussion is presented in section 4. Lastly, the conclusion is drawn in section 5.

II. RELATED WORK

In the past, fault detection in WSN has been investigated [6-11]. The authors have presented an approach based on cross-validation of statistical irregularities for on-line detection of faults in sensor measurements [6]. Ruiz et al. [7] have discussed use of external manager for fault detection in event-driven WSN. The fault diagnosis study based on PMC model is presented in [8]. The use of statistical signal processing technique, namely principal component analysis (PCA) in model development to predict the physical measurand phenomenon is presented in [9]. Any deviation in regular physical pattern with respect to model prediction suggests the occurrence of an event. Similarly, rule-based method, estimation method and learning-based method have been discussed for fault detection/classification of real-world sensor data [10-11]. The performance of these three techniques is qualitatively explored to classify the different types of fault in sensor data as short fault (SF), noise fault (NF) and constant fault (CF). The rule-based approach requires predefining the level of threshold based on histogram method to categorize the noise fault, short fault and constant fault as a separate class. The linear least square estimation approach is based on statistical correlation between sensor measurements and a suitable threshold. The value of threshold remain to be determined heuristically either by maximum error or confidence limit. A learning based approach; Hidden Markov model is also discussed to detect and classify the different fault types. The authors in [12] have used change in mean, variance, covariance for detecting distribution changes in sensor data. This detection scheme is based on the fact, probability distribution of sensor data is known a priori, which is unrealistic in field deployments. A distributed fault detection algorithm for detection and isolation of faulty sensors in communication network is presented in [13]. The proposed approach is based on local comparisons of sensed data between neighbours with a suitable threshold decision criteria test.

The problem associated in processing of huge size data is overcome with use of feature extraction by DWT and has been presented for anomaly detection in [14]. The use of DWT for anomaly detection requires predefining a threshold to make a judgment between *normal* and faulty data series.

Recently, combination of self-organizing map (SOM) with wavelet technique is suggested for anomaly detection on synthetic and as well as real world data sets [15]. The comparative study of said approach outperforms over SOM or wavelet as alone. The histogram method is used to select an appropriate value of threshold. Chenglin et al. [16] have demonstrated the use of particle swarm optimization and support vector machine in fault diagnosis of sensor.

Faulty sensors typically report extreme or unrealistic values that are easily distinguishable. Despite the above research effort, still there does not exist well-accepted technique on anomaly detection and

its classification in wireless sensor data. An edge cutting challenge is to develop the capability to carry out fault diagnosis in terms of its identification and classification without requiring any prior knowledge about the data distribution. There is no consensus on the existence of a simple, accurate and efficient approach in this line of research study. Model based event/anomaly detection scheme requires the availability of *normal* data-series in hand. The DWT technique for anomaly detection gets influenced by the value of threshold used, which in turn depends on number of samples N in data series. Thus correct selection of N requires knowledge to be known in advance on variation of non-faulty sensor data. A threshold set too high will result to increased missed detections, while a low value into many false positives rate. Also, a fixed threshold may not perform well under dynamic scenario of environment pattern. The use of SOM in communication applications or WSNs is widely discussed however, suffers due to its limitation in requirement for processing time, which increases with size of input data. The accuracy of SOM algorithm is influenced by size of neurons, thus a compromise must be reached between the processing time and detection/classification accuracy.
The research analysis oriented to above related problem is due to motivation drawn in application of DWT [17] and [18] for fault detection and SVMs [19] and [21] for binary and multi-class automatic classification of power system/power quality disturbances.

III. METHODOLOGY

The reduction in data size can be obtained by extraction of important statistical features with use of wavelet approach from real time-series data sets. These features vector when passed through SVM results into classification of different types of faults. The combined approach of above two has been successfully applied in study of fault detection and classification in electrical power system. The flow chart to explain the steps adopted in series-data anomaly detection and subsequent classification to different class is illustrated in Fig.1. The anomaly detection scheme embedded in the architecture of sensor node is suggested in Fig. 2. Initially, each sensor node senses its action and information is processed. It is necessary to make a distinguish between normal and anomaly data-series. A mother wavelet extraction and feature classification through SVM is embedded in node architecture to ensure that normal data is transmitted to cluster head.

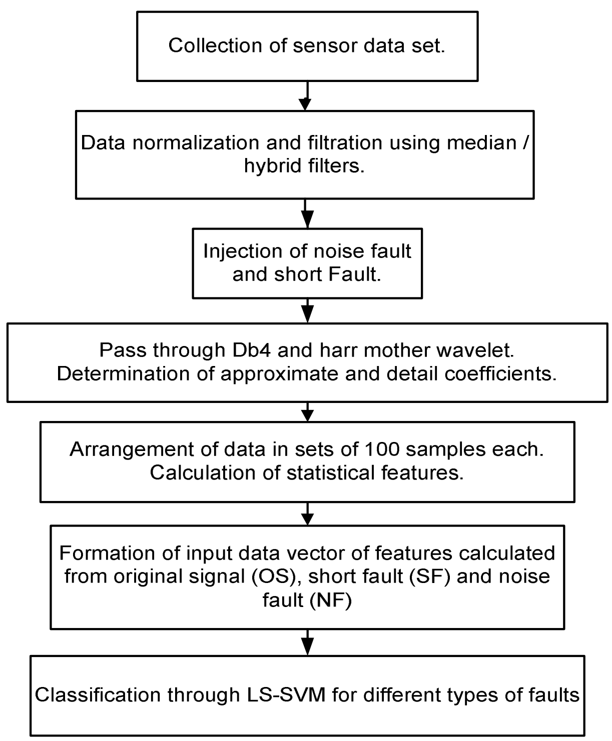


Figure 1. Flow chart of proposed scheme for series-data anomaly detection and classification

3.1 Discrete wavelet transform

The discrete wavelet transform decomposes transients into a series of wavelet components, each of which corresponds to a time-domain signal that covers a specific frequency band containing more detailed information. Wavelets localize the information in the time-frequency plane which is suitable for the analysis of non-stationary signals. DWT divides up data, functions into different frequency components, and then studies each component with a resolution matched to its scale. The separate decomposition of data signal into fine-scale information is referred as detail (D) coefficients, while rough-scale information known as approximate (A) coefficients. The approximation is the high scale, low-frequency component of the signal. The detail is the low-scale, high-frequency components. The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is divided into many lower resolution components which is called the wavelet decomposition tree and is shown in Fig. 3. As decompositions are done on higher levels, lower frequency components are filtered out progressively.

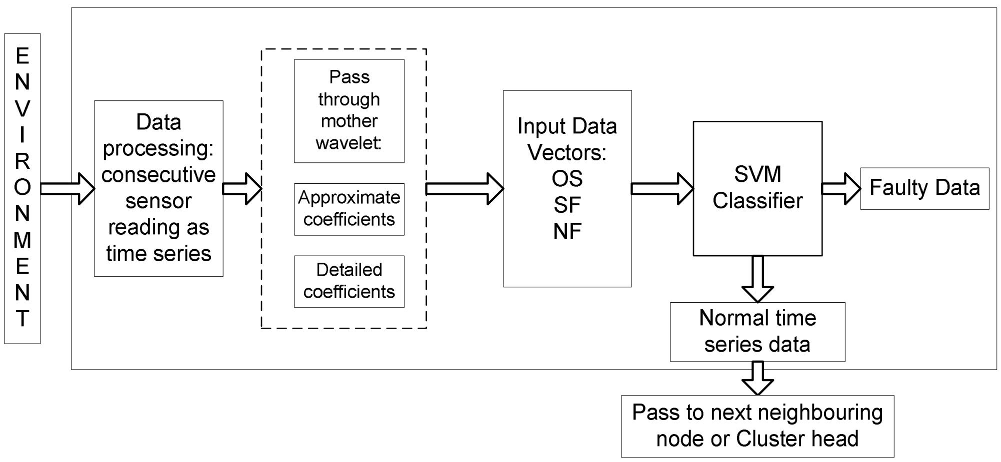


Figure 2. Internal Architecture of anomaly detection scheme

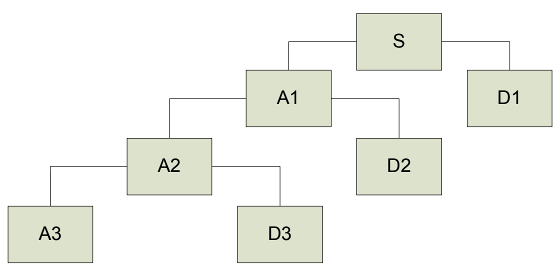


Figure 3. Wavelet decomposition tree

The wavelet transform not only decomposes a signal into frequency bands, but also, unlike the Fourier transform, provides a non uniform division of the frequency domain (i.e., the wavelet transform uses short windows at high frequencies and long windows for low frequency components). Wavelet analysis deals with expansion of functions in terms of a set of basic functions (wavelets) which are generated from a mother wavelet by operations of dilatations and translations.

DWT of sampled data signal can be obtained by implementing the discrete wavelet transform as:

$$DWT(f,x,y)=\frac{1}{\sqrt{x_0^m}}\sum_k f(k)\Psi^*\left(\frac{n-kx_0^m}{x_0^m}\right) \tag{1}$$

Where the parameters x and y in equation (1) are replaced by x_0^m and kx_0^m, k and m being integer variables. In a standard DWT, the coefficients are sampled from the CWT on a dyadic grid. Using the scaling function, the signal can be expressed as:

$$y(t)=\sum_{k=-\infty}^{\infty} c_{jo}(k)2^{jo/2}\phi(2^{jo}t-k)+\sum_{k=-\infty}^{\infty}\sum_{j=jo}^{\infty} d_j(k)2^{j/2}\psi(2^j t-k) \tag{2}$$

Where jo represents the coarsest scale spanned by the scaling function. The scaling and wavelet coefficients of the signal $y(t)$ can be evaluated by using a filter bank of quadrature mirror filters given as:

$$a_j^{AC}(k)=\sum_{m=-\infty}^{\infty} c_{j+1}(m)h(m-2k) \tag{3}$$

$$d_j^{DC}(k)=\sum_{m=-\infty}^{\infty} c_{j+1}(m)h_1(m-2k) \tag{4}$$

Equation (3) and (4) show that the coefficients at coarser level can be attained by passing the coefficients at the finer level to their respective filter followed by a decimation of two. Implementation of DWT involves successive pairs of high pass and low pass filters at each scaling stage of wavelet transform. This can be thought as successive approximations of the same function, each approximation providing the incremental information related to a particular scale (frequency range), and the first scale covering a broad frequency range at the high frequency end of the frequency spectrum, however, with progressively shorter bandwidths. Conversely, the first scale will have the highest time resolution; higher scales will cover increasingly longer time intervals. *Daubechies4* (db4) and *haar* wavelets are used in this work for fault detection in sensor data time-series.

3.2 Support vector machine

A class of machine-learning algorithm that uses kernel function is capable to emulate a mapping of data measurements from the input space vector to a higher dimensional feature space vector. The linear or smooth surfaces in the feature space result into non-linear surfaces in the input space and thereby classify the data as normal or anomalous. Vapnik et al. [22] introduced binary SVM classifier using theory of kernel-based methods and structural risk minimization. In respect of the limitations of other machine learning techniques like, ANNs, local minima convergence, over-learning and difficulty in selection of appropriate network structure does not pose a constraint in use of SVMs. This approach is a computationally powerful algorithm based on statistical learning theory firstly proposed by Salat and Osowski [19]. The input vector space in SVMs is usually mapped into a high dimensional feature space and a hyper-plane in the feature space is used to maximize its classification ability. SVMs can potentially handle large feature spaces as its training is carried out so that the dimension of classified vectors does not affect the performance of SVM. This suits in the application for large classification problem associated in sensor data fault types. The advantage of SVMs are due to better generalization properties as comparison to conventional neural classifiers because training is based on sequentially minimized optimization (SMO) technique [21-22]. For M-dimensional inputs $F_i(i=1,2,............,M), M$ is the number of features sampled at regular interval in time-series data, which belong to class 1 or class 2 with outputs $o_i=1$ for class OS and $o_i=-1$ for class SF/NF, respectively. The hyper-plane for linearly separable feature F is represented as:

$$f(F)=w^T F+b=\sum_{j=1}^{m} w_j F_j+b=0 \tag{5}$$

where w is an m-dimensional vector and b is a constant. The position of the separating hyperplane is decided by the values of w and scalar b. The constraints followed by the hyperplane are
$f(F_i)\geq 1$ if $o_i=1$ and $f(F_i)\geq -1$ if $o_i=-1$ and thus

$$o_i f(F_i)=o_i(w^T F+b)\geq +1 \; for\; i=1,2,............,M \tag{6}$$

The hyperplane that creates the maximum distance between the plane and the nearest data is called the optimal separating hyperplane as shown in Fig. 4. The geometrical distance is found as $\|w\|^{-2}$ [17]. The optimal hyperplane is obtained based on the quadratic optimization problem:
Minimize

$$\frac{1}{2}\|w\|^{-2} + C\sum_{i=1}^{M}\xi_i \text{ subject to } \quad o_i(w^T s + b) \geq 1 - \xi_i \quad for\ i = 1,2,.....,M \tag{7}$$

$$\xi_i \geq 0 \quad for \quad all \quad i$$

where ξ_i is the distance between the margin, parameter C is error penalty factor that takes into account misclassified point in training/testing set and the examples F_i lying on the wrong side of the margin. Based on Kuhn–Tucker conditions, a maximize problem [17] can be formulated and the solution of these optimal problem leads to determination of support vector (SV) which lie on the separating hyper planes. The number of SVMs are less than the number of training samples to make SVMs computationally efficient [19]. The value of the optimal bias b^* can be found from the expression: $b^* = -\frac{1}{2}\sum_{SVs} o_i\alpha_i^*(v_1^T F_i + v_2^T F_i)$ (8)

where v_1 and v_2 are the arbitrary SVMs for class 1 and class 2, respectively.
Then the final decision function is given by

$$f(F) = \sum_{SVs}\alpha_i o_i F_i^T F + b^* \tag{9}$$

Any unknown feature sample F is thus classified as,

$$F \in \begin{Bmatrix} Class-1, \ f(F) \geq 0 \\ Class-2, \ otherwise \end{Bmatrix} \tag{10}$$

The nonlinear classification of sensor data faults can accomplished using SVMs applying a kernel function by mapping the classified data to a high-dimensional feature space where the linear classification is possible [19]. There are different kernel functions used according to the type of classification scenario.

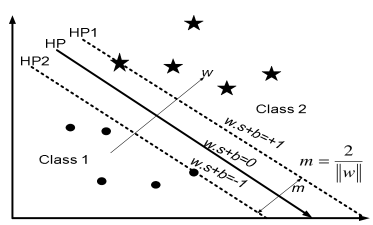


Figure 4. Optimal hyper-plane formed in SVM classification

In this paper, Gaussian radial basis kernel function which gives the best results is selected and the classification accuracy results are compared with other kernel functions, i.e. polynomial kernel. The radial basis kernel function is defined as:

$$K(F,z) = \exp\left(-\frac{|F-z|^2}{2\sigma^2}\right) \tag{11}$$

where σ is the width of the Gaussian function known as Gaussian kernel parameter. The detailed explanation about the SVMs is given in [19]-[21].

3.3 Real-time series data signal processing

The combination of above two techniques is implemented to support the proposed strategy of anomaly detection in a collection of real-time series data obtained from Smart-Its [23]. A Smart-It unit embodies a sensor module consisting of light sensor, microphone thermometer, X-axis and Y-axis accelerometers and pressure sensor along with a communication module. The series time variation of sound, light and pressure signals are shown in Fig. 5. These data sets were obtained over several states of environment. The constant value of pressure sensor over the entire data series is depicted which suggests a “constant” fault type. The real-time wireless sensor data of sound, light and pressure signals is processed after being passed through median filter and median-hybrid filter. Median filter is the nonlinear filter used to preserve abrupt shifts (edges) and remove the impulsive noise from the data-series. The main issue that exists with median filter is due to its high computational cost. While on the other hand, linear median-hybrid filters have been suggested to combine the good properties of linear and median filters by linear and nonlinear operations. They are computationally much less expensive than standard median filters. The series-data in study for anomaly detection is normalized to eliminate the potential outliers as:

$$Normalized\ data = \frac{Raw\ data - Mean(Raw\ data)}{Variance(Raw\ data)} \quad (12)$$

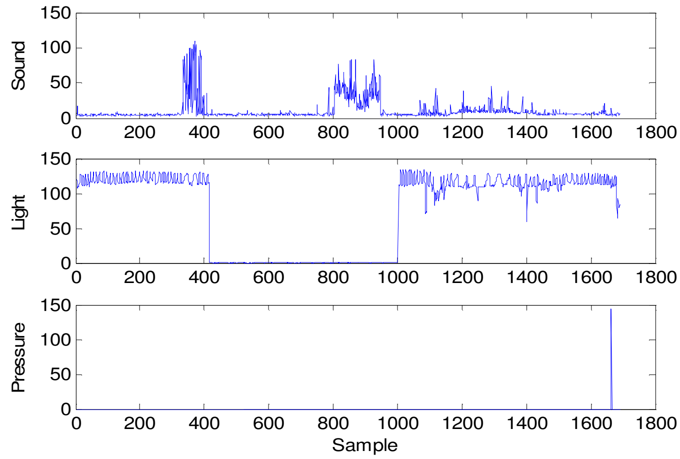


Figure 5. Real-time series variation of raw signals

3.4 Sensor data faults:

The three common types of sensor data faults as according to the definition in [8] are short fault, noise fault and constant fault. The short fault refers to sharp change in monitored quantity at an instant with respect to its previous sample. The noise fault is characterized by an increased variance over a definite period, i.e. successive samples unlike short fault at single sample only. On the other hand, constant fault describes a constant value, may be either higher or lower compared to normal measurements for successive samples. Such fault type results to zero value of standard deviation for monitored samples. In the study reported here, only two types of faults; short fault and noise fault are considered. These faults have been experimentally observed in several environmental monitoring platforms.

A sample of short fault (SF) data is obtained by injecting short fault intensity $f = \{3.5\}$ to a data value as: $d_i^{sf} = d_i \times f$ (13)

at a randomly picked data sample d_i.

Fig. 6 shows the instants at which short fault were injected into the signal obtained through filters for their detection classification. The total percentage of short fault injected into series data is about 1.0%.Similarly, a series of noise fault (NF) is introduced into normalized raw data through random selection of successive samples d_s and superimpose of a random signal with 20dB noise content having signal property of zero mean and unity variance. The variation of sound series data with noise introduced at randomly chosen 200 successive samples over three different intervals is shown in Fig. 7. Thus, total number of noise fault samples in the series data is 35.5%.

3.5 Combination of DWT and SVM:

The approximate and detail coefficients are obtained through *db4* and *haar* wavelet from the normalized data after being passed into median and hybrid filter. These coefficients belong to original signal (OS) without any fault, short fault and noise fault injected in time series data. To reduce the size of input data fed to SVM, four features; namely mean, standard deviation, moment and variance are extracted from each 100 samples in time series data. Thus time-series data is transformed into sets of features $\{f_{mean}, f_{STD}, f_m, f_{var}\}$ and now to be represented as:

$$F_{OS}, F_{SF}, F_{NF} = \begin{bmatrix} f_{mean} & f_{STD} & f_m & f_{\text{var}} & for\ 1-100\ samples \\ : & : & : & : & : \\ f_{mean} & f_{STD} & f_m & f_{\text{var}} & for\ 1501-1600\ samples \end{bmatrix} \tag{14}$$

Thus, feature vector of time-series data consists of 16 rows with 4 columns.

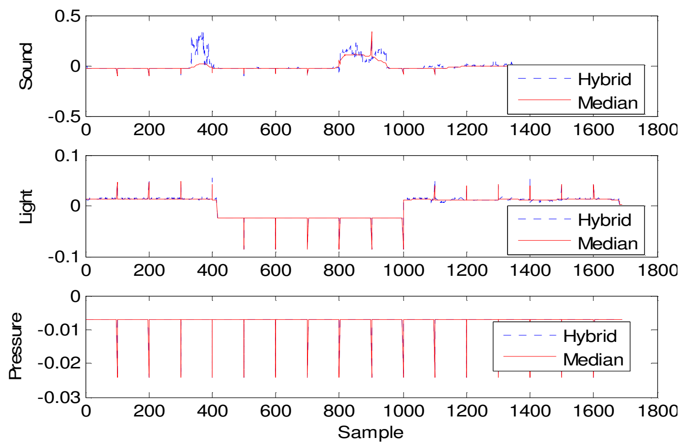


Figure 6. Short fault injected into the raw signal (normalized)

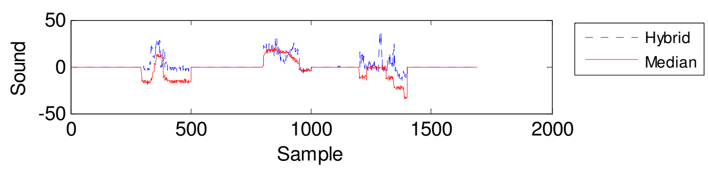


Figure 7.Noise fault introduced into the raw signal (normalized)

The data collection by sensor may have any pattern of anomaly present in the entire length of time-series. A subset of data measurements over some continuous time frame may differ in their pattern from the general trend to warrant being considered as anomalous data series. Hence to take into account such phenomenon occurrence, the input data vector fed to SVM is represented in two different forms; sequential-series (SE) and staggered-series (ST). A sequential-series of features refers to time-series wherein, entire length of data consists of samples corresponding to original signal followed by anomaly signal. On other hand, staggered-series relates to time-series that consists of alternate sampled series of original signal and anomaly signal. An enhanced performance in classification may be achieved with use of more number of data sets in training of SVM. So, use of duplicate data sets corresponding to each pattern is considered in study. Thus, input vector fed to SVM for classification is given as:

$$(Input\ vector)_{SE} = \begin{bmatrix} F_{OS} \\ F_{OS} \\ F_{SF,NF} \\ F_{SF,NF} \end{bmatrix}; \qquad (Input\ vector)_{ST} = \begin{bmatrix} F_{OS} \\ F_{SF,NF} \\ F_{OS} \\ F_{SF,NF} \end{bmatrix} \tag{15}$$

and forms 32 rows with 4 columns.

With the above input vector, the objective remains to partition set of features belonging to each category of type of signal, i.e. $F_{OS} \cap F_{SF} = \Phi$ and $F_{OS} \cap F_{NF} = \Phi$. The output of SVM algorithm for sets of features that belong to OS class is defined as 1, while for fault types, as -1 to differentiate between the two categories. The input vector (15) obtained using time-series data passed through median filter is considered for training, while those from hybrid filter as testing of SVM classifier.

IV. PERFORMANCE EVALUATION AND DISCUSSION

This section presents the performance evaluation of proposed scheme; integration of DWT and SVM in detection and classification of anomaly in time-series data collected by wireless sensor. The results presented here are produced using real-time series data sets obtained from sensor modules deployed in real environment. The performance indices (16-18) are used to assess the performance of proposed scheme of anomaly detection in real time-series data sets [21]. Consider $\{P,N\}$ be the positive and negative instance classes as assigned and $\{P_c,N_c\}$ be the classifications obtained by the SVM classifier. Also consider, $P(P|I)$ be the posterior probability for an instance I that is positive. Then, True positive rate (TPR) of the classifier is:

$$TPR = P(P_c|P) \approx \frac{positives\ correctly\ classifed}{total\ \ positives\ assigned} \tag{16}$$

False positive rate (FPR) of the classifier is:

$$FPR = P(P_c|N) \approx \frac{negatives\ incorrectly\ classifed}{total\ negatives\ assigned} \tag{17}$$

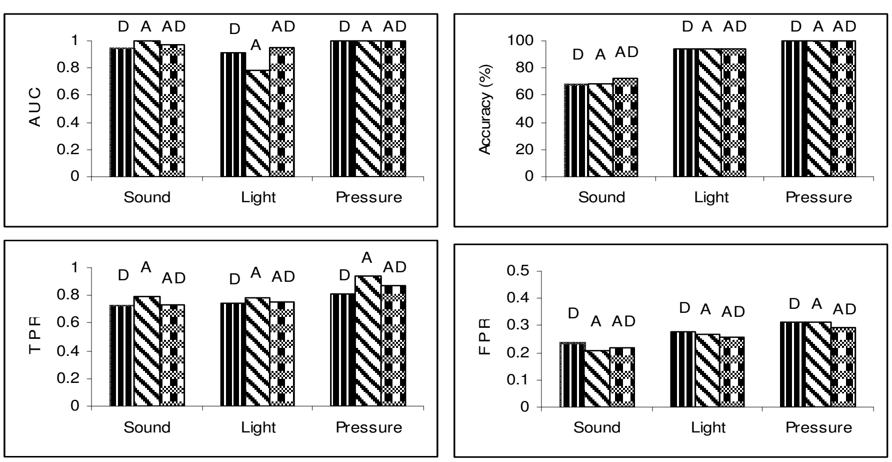
Detection accuracy (DA) of the classifier is:

$$Detection\ accuracy = \frac{TPR}{TPR+FPR} \times 100\,\% \tag{18}$$

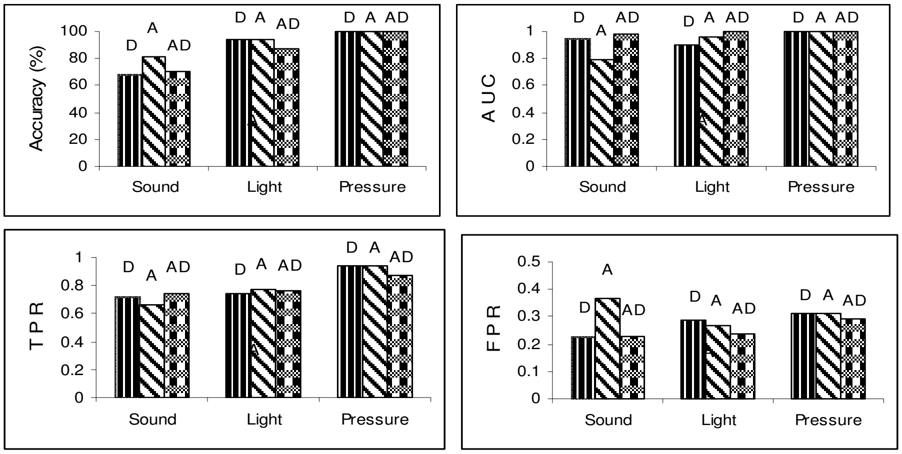
Area under the receiver operating characteristic (ROC) curve (AUC): The area under the ROC curve, or simply AUC, provides a good "summary" for the performance of the ROC curves [22].

4.1 SVM as binary classifier:

The performance indices of classifier scheme are evaluated using features extracted from detail (D), approximate (A) and both approximate and detail (AD) coefficients of wavelet. The analysis of these indices determined for time-series data belonging to original signal and short fault is shown in Fig. 8. The AUC value of classifier is observed to be in the range from 0.90-1.0. A unity value of AUC is indicated for pressure data series. In fact, the original pressure signal exhibits a constant value and a short fault injected within 100 samples, are distinctly represented in form of statistical feature. Thus, such change in data pattern is distinctly classified as a separate class. Fig. 9 shows the classification performance of original signal against noise fault. As observed, AUC gets increased with use of features extracted from both approximate and detail (AD) coefficients of wavelet. The classification pattern generated from SVM classifier for light signal and sound signal is depicted in Fig. 10 and 11 respectively. As observed, the features are distinctly represented through the classifier boundary.

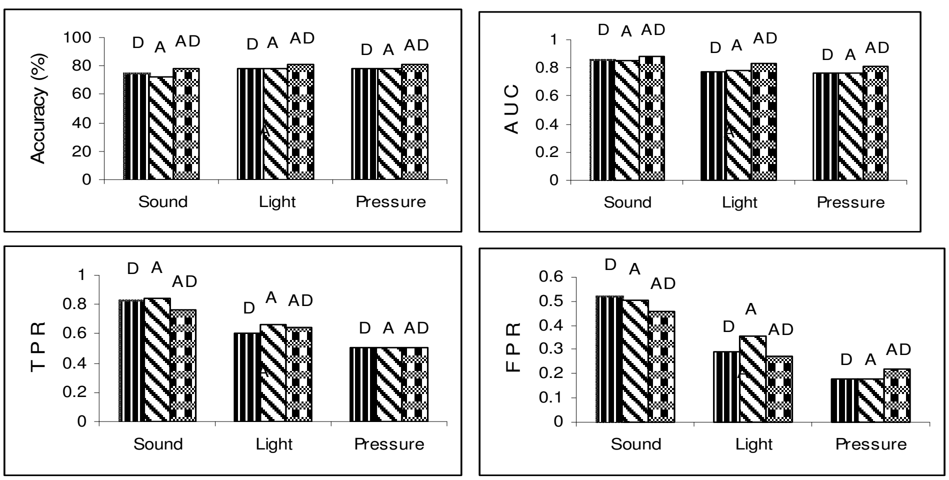


(a) Sequential series

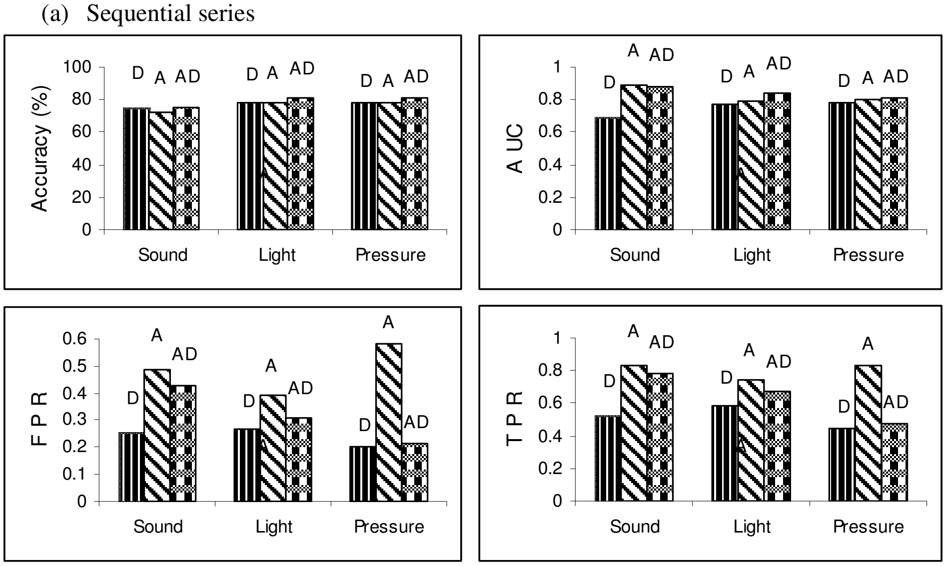


(b) Staggered series

Figure 8. Performance indices of SVM classifier as binary class for OS vs SF

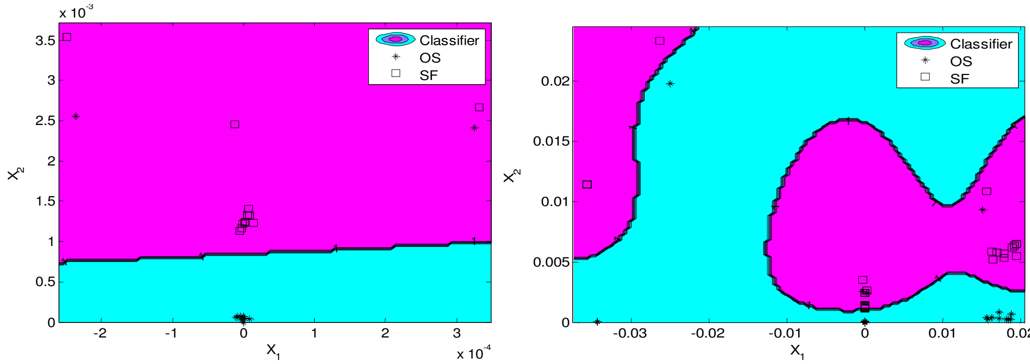


(a) Sequential series



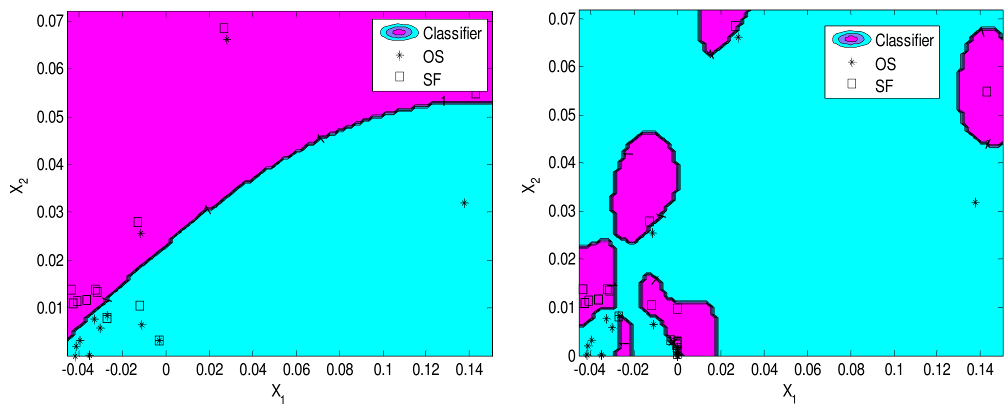
(b) Staggered series

Figure 9.Performance indices of SVM classifier binary class for OS vs NF



(a) Detail coefficient (b) Both approximate and detail coefficient

Fig. 10.Classification pattern of SVM classifier for light signal as sequential series



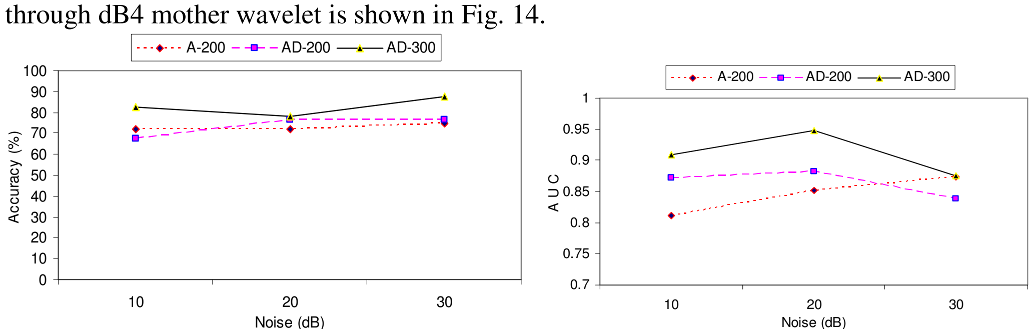
(a) Approximate coefficient (b) Both approximate and detail coefficient

Figure 11. Classification pattern of SVM classifier for sound signal as staggered series

Further, the result is presented for time series data having different magnitude of noise introduced at randomly chosen 200 and 300 successive samples with features fed as sequential series to SVM classifier. The classification performance between original and noise of sound signal by use of approximate and approximate-detail coefficients is presented in Fig. 12. As observed, the classification property has not deteriorated.

Next, classifier performance is tested for time series data having different magnitude of short fault introduced. The results are presented in Fig. 13 for classification between original and short fault light signal with features fed as sequential and staggered series.

The SVM classifier by use of coefficients extracted through haar mother wavelet is also carried out and presented in following paragraph. The results are obtained for short fault, $f = \{3.5\}$ and 20 dB noise introduced in time series data. The comparative performance with AD coefficients extracted through dB4 mother wavelet is shown in Fig. 14.



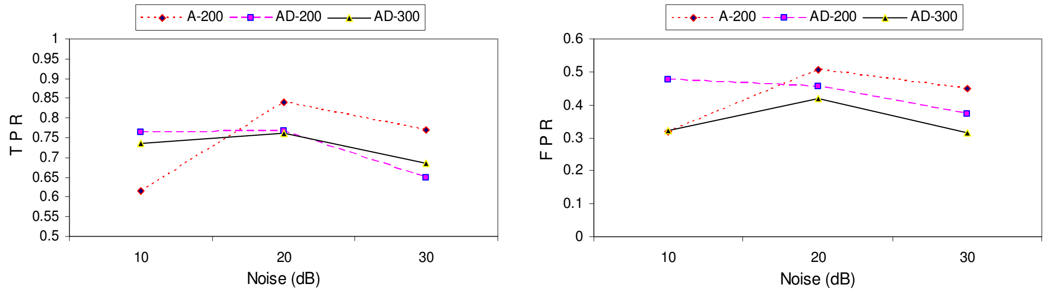


Figure 12.Classification performance for different magnitude of noise introduced at randomly chosen 200 and 300 successive samples

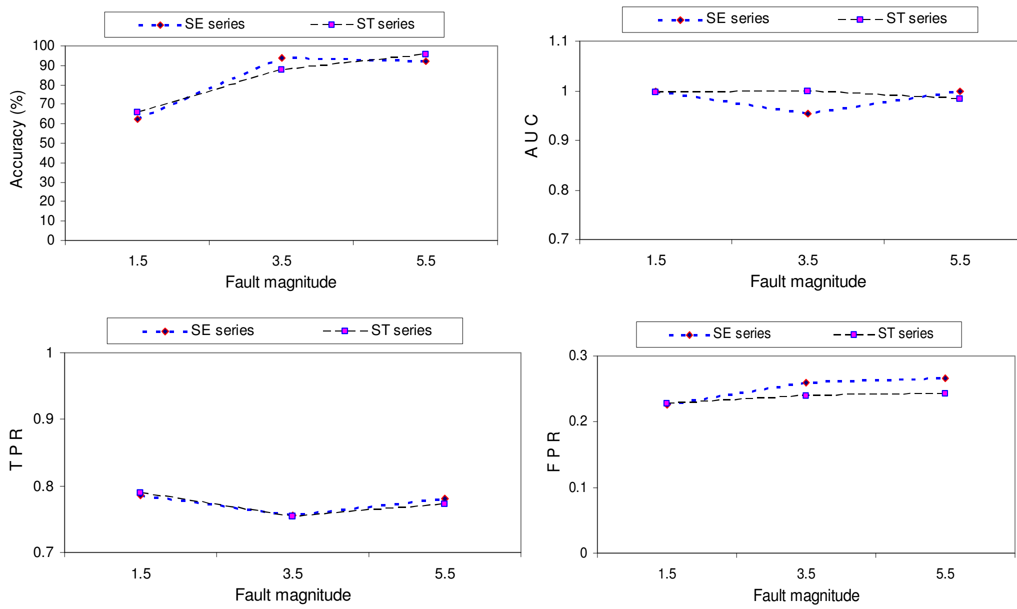


Figure 13. Classification performance for different magnitude of short fault introduced

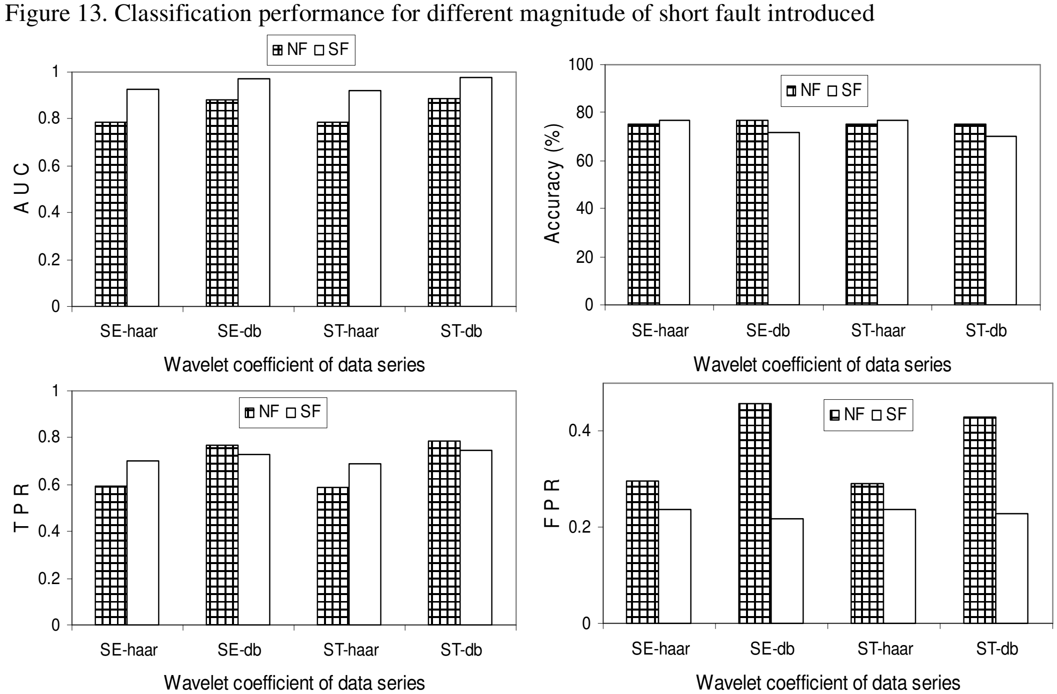
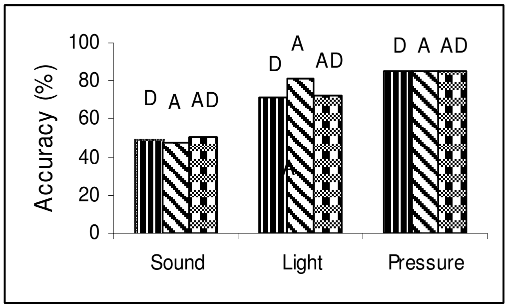


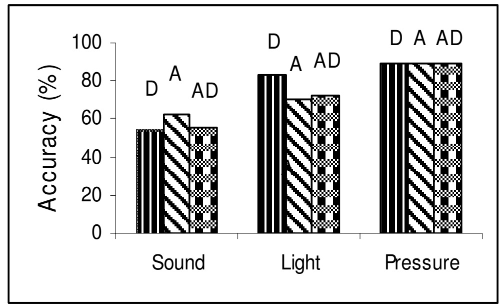
Figure 14. Comparative performance between mother wavelets for OS-SF and OS-NF by use of features as sequential and staggered series

4.2 SVM as multi-class classifier:

The classification of original signal against short fault and noise fault as a multi-class problem is discussed in this sub-section. Since performance in terms of detection accuracy can be considered for multi-class, thus other indices are not evaluated. Fig. 15 presents the detection accuracy with use of features extracted from different coefficients of wavelet



(a) Sequential series



(b) Staggered series

Figure 15. Performance indices of SVM classifier as multi-class for OS vs SF vs NF

V. CONCLUSION

The integration of DWT and SVM for anomaly detection and classification problem was presented in this paper using real-time series data of wireless sensor deployed in field environment. The signal processing property of DWT was utilized in fine-scale and approximate-scale extraction of information from data. The use of statistical features instead of series data in form of wavelet coefficients resulted in reduce size of input vector fed to SVM. The value of AUC as binary class was determined in the range of 0.9-1.0 for OS against SF, while for OS against NF, it lies between 0.75-0.86. The robustness of SVM classifier was demonstrated for fault magnitude change and different noise level introduced in time series data. The detection accuracy as multi-class was also found to be high. The suggested approach in anomaly detection and classification is independent from heuristic adjustment of any parameter and does not require any domain knowledge of non-faulty data series in obtaining high accuracy.

REFERENCES

[1] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, W. Hong,(2005) "A macroscope in the Redwoods," Proc. of 2nd International Conference on Embedded Networked Sensor Systems, New York, USA, pp. 51-63.

[2] N. Ramanathan, L. Balzano, M. Burt, D. Estrin, E. Kohler, T. Harmon, C. Harvey, J. Jay, S. Rothenberg, M. Srivastava,(2006), "Rapid deployment with confidence: calibration and fault detection in environmental sensor networks," CENS, Tech. Report 62.

[3] G. Werner-Allen, K. Lorincz, J. Johnson, J. Lees, M. Welsh,(2006), " Fidelity and yield in a volcano monitoring sensor network," Proc. of 7th USENIX Symposium on Operating Systems Design and Implementation.

[4] V.Alarcon-Aquino and J.A. Barria, (2001), "Anomaly detection in communication networks using wavelets," IET Journal of Communication, Vol.148, No. 6 , pp.355-362.

[5] G. Kaur, V. Saxena, and J.B. Gupta, (2010), "Anomaly Detection in Network traffic and Role of Wavelets," IEEE Transactions on Instrumentation and measurement, Vol.7, No.5, pp.46-51.

[6] F. Koushanfar, M. Potkonjak, A Sangiovanni-Vincentelli,(2003), " On-line fault detection of sensor measurements," IEEE Sensors, No.2, pp. 974-980.

[7] L. B. Ruiz, I. G. Siqueira, L. B. Oliveira, H. C. Wong, J. M. S. Nogueira, A. A. F. Liureiro, (2004) , "Fault management in event-driven wireless sensor networks," Proc. of MSWIM'04.

[8] S. Chessa, P. Santi,(2001) Comparison-based system-level fault diagnosis in ad hoc networks, Proc of 20th Symposium on Reliable Distributed System, pp. 257-266.

[9] J. Gupchup, R. Burns, A. Terzis, A. Szalay,(2007), " Model-based event detection in wireless sensor network," Data Sharing and Interoperability on the World-Wide Sensor Web, Boston, 2007.

[10] A. Sharma,L. Golubchik, R. Govindan, (2010), " Sensor faults: detection methods and prevalence in real-world datasets," Transactions on Sensor Networks, Vol. 5, pp. 1-34.

[11] Y. Yao, A. Sharma, L. Golubchik, R. Govindan, (2010), "Online anomaly detection for sensor systems: a simple and efficient approach," Performance Evaluation, Vol. 67, pp. 1059-1075.

[12] A. Tartakovsky, V. Veeravalli, (2008), "Asymptotically optimal quickest change detection in distributed sensor systems," Sequential Analysis, Vol. 27, pp. 441-475.

[13] M.-H. Lee, Y.-H. Choi, (2008)Fault detection of wireless sensor networks, Computer communication, Vol. 31, pp. 3469-3475.

[14] V. A. Aquino, J. A. Barria,(2007), " Anomaly detection in communication networks using wavelets," IEEE Proc. in Communications, Vol. 148, pp. 1113-1118.

[15] S. Siripanadorn, W. Hattagam, N. Teaumroog, (2010), "Anomaly detection in wireless sensor networks using self-organizing map and wavelets," International Journal of Communication, Issue 3, Vol. 4, pp. 74-83.

[16] Z. Chenglin, S. Xuebin, S. Songlin, J. Ting,(2011), " Fault diagnosis of sensor by choas particle swarm optimization algorithm and support vector machine," Article in Press. 2011.

[17] S. J. Huang, C. T. Hsieh, (2002) , "Coiflet wavelet transform applied to inspect power system disturbances-generated signals," IEEE Transactions on Aero. Electronics System, Vol.38, No.1 pp204–210.

[18] Prakash K Ray, Soumya R. Mohanty, Nand Kishor, (2011), "Disturbance detection in grid-connected distributed generation system using wavelet and s-transform," Electric Power System Research, Vol. 81, pp. 805-819.

[19] R. Salat and S. Osowski, (2004), "Accurate fault location in the power transmission line using support vector machine approach," IEEE Trans. on Power Systems, vol. 19, pp. 879–886.

[20] P. K. Dash, S. R. Samantaray and P. Ganapati, (2007) "Fault classification and section identification of an advanced series-compensated transmission line using support vector machine", IEEE Trans. on Power Delivery, vol. 22, pp. 67–73.

[21] Sami Ekici, (2009), "Classification of power system disturbances using support vector machines," Expert Systems with Applications, vol. 36, pp. 9859–9868.

[22] V. N. Vapnik, (1998), "Statistical Learning Theory," Hoboken, NJ: Wiley, 1998.

[23] Smart-Its Project Home Page: http://smart-its.teco.edu/

[24] J. Huang, C. X. Ling,(2005), " Using AUC and Accuracy in Evaluating Learning Algorithms," IEEE Transactions on Knowledge and Data Engineering, Vol. 17, pp.299-310.

Authors Biographies

Ajay Singh Raghuvanshi received his B.Tech. degree in Electronics and Communication Engineering from the Department of Electronics and Communication Engineering, North Eastern Regional Institute of Science and Technology, Northeastern Hill University, India in 1993. He is currently working towards the Ph.D. degree at the Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology, Allahabad, India. He taught in College of Science and Technology, Royal University of Bhutan, from 1993 till 2007. He is presently teaching at Indian Institute of Information technology, Allahabad, India. His research interests are in the area of wireless Sensor networks, with emphasis on Energy Efficient sensor networks.

Rajeev Tripathi received his B.Tech, M.Tech., and Ph.D. degrees in Electronics and Communication Engineering form Allahabad University, India. At present, he is a Professor in the Department of Electronics and Communication Engineering, at Motilal Nehru National Institute of Technology, Allahabad, India. He worked as a faculty member at the University of The West Indies, St. Augustine, Trinidad, WI, during September 2002-June 2004. He was a visiting faculty at School of Engineering, Liverpool John Moorse University, U.K., during May-June 1998 and Nov-Dec 1999. He carried out joint research project under Indo-UK science and technology research fund and other funding agencies. He worked as reviewer of IEEE Communication Letters and West Indian Journal of Engineering. He served as program co-chair of the First International Conference on Computational Intelligence, Communication Systems, and Networks, held in Indore, India, in July 2009. He is on the program committee of several international conferences in the area of wireless communication and networking. His research interests are high speed communication networks, performance of next generation networks: switching aspects, MAC protocols, mobile ad hoc networking, and IP level mobility management.

Sudarshan Tiwari received his B.Tech. degree in Electronics Engineering from I.T.BHU, Varanasi, India in 1976, the M.Tech. degree in Communication Engineering from the same institution in 1978 and PhD degree in Electronics and Computer Engineering from IIT Roorkee, India in 1993. Presently, he is Professor and Head of Department of Electronics and Communication Engineering. Motilal Nehru National Institute of Technology (MNNIT), Allahabad, India. He has also worked as Dean Research and Consultancy of the institute from June 2006 till June 2008. He has more than 28 years of teaching and research experience in the area of communication engineering and networking. He has supervised a number of M.Tech and PhD thesis. He has served on the program committee of several seminars, workshops and conferences. He has worked as a reviewer for several conferences and journals both nationally and internationally. He has published over 78 research papers in different journals and conferences. He has served as a visiting professor at Liverpool John Moorse University, Liverpool, UK. He has completed several research projects sponsored by government of India. He is a life member of Institution of Engineers (India) and Indian society of Technical Education (India), he is a member of Institution of Electrical and Electronics Engineers (USA). His current research interest include, in the area of WDM optical networks, wireless ad hoc & sensor networks and next generation networks.