Cyber-Attack Identification of Synchrophasor Data Via VMD and Multi-fusion SVM

Wei Qiu[1,2], Kunzhi Zhu[1], Zhaosheng Teng[1], Qiu Tang[1]
College of electrical and information engineering
Hunan University[1]
Chang Sha, China
qiuwei@hnu.edu.cn, zhukunzhi@hnu.edu.cn,
tengzs@126.com, tangqiu@hnu.edu.cn

Wenxuan Yao[2, 3], Yuqing Dong[2], Yilu Liu[2, 3]
Department of EECS
University of Tennessee[2], Oak Ridge National Laboratory[3]
Knoxville, TN, U.S.
wyao3@utk.edu, ydong22@utk.edu
liu@utk.edu

2020 IEEE Industry Applications Society Annual Meeting | 978-1-7281-7192-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/IAS44978.2020.9334870

***Abstract*—A large amount of synchrophasor data in the Wide Area Measurement System (WAMS) needs to be collected and transmitted to the phasor data concentrator, thereby increasing the possibility of being attacked by hackers. The attacked data is therefore hidden into the normal data. To remedy this problem, an identification framework is proposed to detect the cyber-attack in WAMS utilizing Variational Mode Decomposition (VMD) and Multi-fusion Support Vector Machine (MSVM). First, VMD is used to transforms the modified data into multiple modal components and then MSVM is employed to classify the deterministic features using the linear combined multi-kernel. This linear combined multi-kernel fuses multiple types of features including time, frequency and statistical variables of the synchrophasor data. Utilizing the actual data from FNET/GridEye, different experiments are conducted under multiple attack strengths and types. The results show that the identification framework has a higher accuracy and robustness compared with other conventional classifiers.**

***Keywords*—Linear combined multi-kernel, Multi-fusion Support Vector Machine, Synchrophasor data, Variational Mode Decomposition**

I. INTRODUCTION

The quality of the synchrophasor data collected in the Wide Area Measurement System (WAMS) is critical for grid situational awareness and disturbance event location. However, the synchrophasor data is vulnerable to the cyber-attack, such as False Data Injection Attack (FDIA) and Denial of Service. Moreover, the FDIA methods can be achieved secretly due to security holes of IEEE C37.118 [1]. Apart from this, a variety of FDIA methods make it difficult to detect the attack behavior. To enhance the synchrophasor data quality and availability, it is necessary to detect the FDIA from the normal data in the WAMS.

Generally, the cyber-attack identification can be categorized into model-driven and data-driven methods [2]. The model-based detection methods are proposed based on the power system parameters and configuration. The Weighted Least Squares (WLS) is one of the most commonly used model-based methods, which can be used to estimate the system states and topology change caused by FDIA [3]. However, the WLS assumes that the system is operating under steady-state conditions.

To reduce dependence on the model and system parameters, different data-driven methods are proposed to learn and distinguish the FDIA. For example, Artificial Neural Networks (ANN) is used to identify the event of the cyber-attack in the compromised meters [4]. The electrical theft, by tampering with billing alterations, can be detected using the Decision Tree (DT) and Support Vector Machine (SVM) [5]. However, the ANN and DT can generate over-complex nodes and trees, resulting in decreased performance in testing data. To overcome this problem, some advanced methods such as deep belief networks [6] and convolutional neural network [7] are used to automatically extract the features of attack signals. Due to the diversity of FDIA methods, the performance of these networks is limited by the single input information.

Combined with the advantages of strong feature extraction capabilities of data-driven methods, a novel framework based on Variational Mode Decomposition (VMD) and Multi-fusion SVM (MSVM) is proposed to identify multiple cyber-attacks in synchrophasor data. Particularly, the MSVM can fuse various attack features through the linear combined multi-kernel, thus avoiding the problem of insufficient information in a single input.

The remainder of this paper is organized as follows: in Section II, the VMD and definition of extracted features are presented. Then, the proposed MSVM method is described to detect the cyber-attack in Section III. Different experiments are conducted in Section IV. Finally, the experimental results are discussed in Section V.

II. FEATURE EXTRACTION BASED ON VMD

A. Principle of VMD

VMD is a new multiresolution technology for adaptive and non-recursive signal decomposition, which is suitable for analyzing non-linear and non-stationary signals [8].

For the attacked synchrophasor data $f(t)$, the VMD can automatically decompose signal $f(t)$ into multiple Intrinsic Mode Functions (IMFs) with sparse characteristics and limited bandwidth. The sum of different IMFs can restore $f(t)$. Specifically, the VMD optimizes the following constraints to generate IMFs, which can be expressed as

$$\begin{cases} \min\limits_{\{u_n\},\{\omega_n\}} \left\{ \sum_{n=1}^{N} \left\| \partial_t \left[\left(\delta(t) + \dfrac{j}{\pi t} \right) * u_n(t) \right] e^{-j\omega_n t} \right\|_2^2 \right\} \\ s.t. \ \sum_{n=1}^{N} u_n(t) = f(t) \end{cases} \quad (1)$$

where N is the number of IMFs, $\delta(t)$ is the Dirac delta function, $u_n(t) = \{u_1(t), u_2(t), \ldots, u_N(t)\}$ are shorthand notations for the set of all IMFs, $\omega_n = \{\omega_1, \omega_2, \ldots, \omega_N\}$ are shorthand notations for the center frequency of $u_n(t)$.

This work is supported by the Engineering Research Center Program of the National Science Foundation and DOE under NSF Award Number EEC-104187 and the CURENT Industry Partnership Program. The first two authors contributed equally to this paper.

To calculate $u_n(t)$, the amplitude-modulation-frequency-modulation signals can be used as

$$u_n(t) = A_n(t)\cos(\Phi_n(t)) \tag{2}$$

where $A_n(t) \geq 0$ is the envelope of $u_n(t)$, $\Phi_n(t)$ is the phase of $u_n(t)$.

To obtain the optimal solution of constrained variational problems in Equation (1), the Lagrange multiplier method and penalty term are introduced. The following expression is obtained as

$$\begin{aligned} &L(\{u_n\},\{\omega_n\},\lambda) = \\ &\alpha\sum\nolimits_N \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_n(t) \right] e^{-j\omega_n t} \right\|_2^2 + \\ &\left\| f(t) - \sum\nolimits_N u_n(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum\nolimits_N u_n(t) \right\rangle \end{aligned} \tag{3}$$

where the α is penalty term, λ is Lagrange multiplier.

Combined with Alternate Direction Method of Multipliers (ADMM), the Parseval/Plancherel Fourier isometry under the L2-norm is used to convert the u_n and ω to frequency domain [8]. The iterative formulas of $\hat{u}_n(\omega)$ and ω_n, are expressed as follows

$$\hat{u}_n^{m+1}(\omega) \leftarrow \frac{\hat{f}(\omega) - \sum_{i<n}\hat{u}_i^{m+1}(\omega) - \sum_{i>n}\hat{u}_i^{m}(\omega) + \frac{\hat{\lambda}^m(\omega)}{2}}{1 + 2\alpha\left(\omega - \omega_n^m\right)^2} \tag{4}$$

and

$$\omega_n^{m+1} \leftarrow \frac{\int_0^\infty \omega \left|\hat{u}_n^{m+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_n^{m+1}(\omega)\right|^2 d\omega} \tag{5}$$

where the $\hat{u}_n(\omega)$ is Fourier transform of $u_n(t)$, ω is the center frequency of $\hat{u}_n(\omega)$, $\hat{f}(\omega)$ is Fourier transform of $f(t)$, $\hat{\lambda}(\omega)$ is Fourier transform of $\lambda(t)$. To simplify the calculation, the gradient descent method is used to solve $\hat{\lambda}(\omega)$, which can be expressed as

$$\hat{\lambda}^{m+1}(\omega) \leftarrow \hat{\lambda}^{m}(\omega) + \beta\left(\hat{f}(\omega) - \sum\nolimits_N \hat{u}_n^{m+1}(\omega)\right) \tag{6}$$

where β is the quadratic penalty term, which can improve the convergence rate of $\hat{\lambda}(\omega)$. Thereafter, the $u_n(t)$ is obtained by the inverse Fourier transform of $\hat{u}_n(\omega)$.

Here, the number of decompositions N should not be too small to avoid incomplete decomposition or too large to avoid false components. In this paper, the N is optimally set to 6, which means 6 IMF_i components are decomposed, where i=1, 2, …, 6.

To demonstrate the effect of VMD, an example of two different cyber-attack signals are presented in Fig. 1. It can be seen that the start and end time components of the scale attack are detected in IMF_5 of Fig. (c). Meanwhile, the changing trend of frequency shock attacks is reflected in the residual component (IMF_6).

B. Feature Extraction

After obtaining the IMFs of different attacked synchrophasor data, distinctive attack features are extracted for identification. Specifically, four types of features are used including two statistical features, and another two from the frequency and time domain respectively.

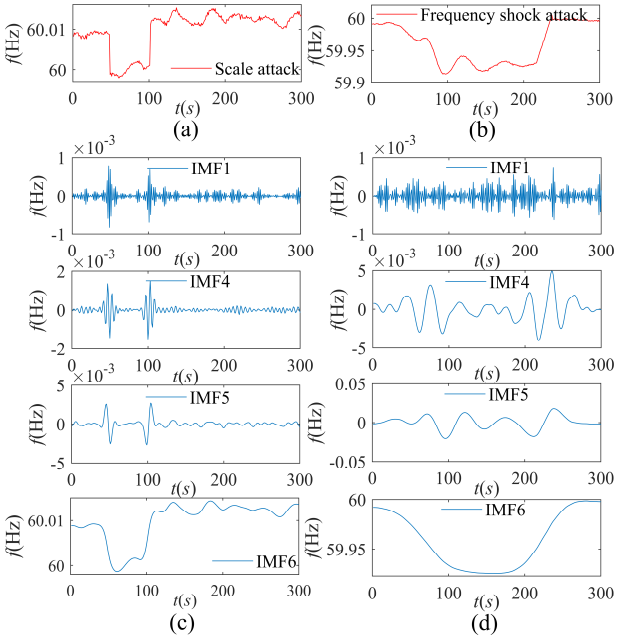

Fig. 1. Example of the signal decomposed by VMD. (a) The scale attack, (b) The frequency shock attack, (c) The VMD result of (a), (d) The VMD result of (b).

These two statistical features are the kurtosis index and envelope entropy to reflect the characteristics of each IMF. The kurtosis index is sensitive to transient signals, thus it can reflect the degree of non-stationarity of each IMF. For IMF_i of length n, its kurtosis can be expressed as

$$Ku_i = \frac{\frac{1}{n}\sum_{j=1}^{n}\left(m_j - \bar{m}\right)^4}{\left(\frac{1}{n}\sum_{j=1}^{n}\left(m_j - \bar{m}\right)^2\right)^2} \tag{7}$$

where m_j is the jth value of IMF_i, and $\bar{m}$ is the average value of IMF_i.

If Ku_i is positive, it means that IMF_i has obvious peak characteristics, which is called super-Gaussian distribution. If Ku_i is negative, it means that IMF_i has no obvious impact or pulse signal, which is called sub-Gaussian distribution.

The second feature is envelope entropy, in which the distribution of source-information can be analyzed. The information entropy of each IMF envelope is a measure of the overall distribution of the signal. For each IMF_i, the envelope entropy can be calculated as

$$\begin{cases} Ee_i = -\sum_{j=1}^{n}\left(p_j \cdot \ln p_j\right) \\ p_j = E_{ij}(t) \Big/ \sum_{j=1}^{n} E_{ij}(t) \\ \sum_{j=1}^{n} p_j = 1 \end{cases} \tag{8}$$

where $E_i(t)$ is the envelope signal of IMF_i obtained by Hilbert transform, Ee_i is the envelope entropy of IMF_i. Generally, the more uniform the distribution of variables in the IMFs, the smaller information entropy value of Ee_i.

The statistical characteristics of attack signals can be extracted by using the Ku_i and Ee_i. The time and intensity and

unique fingerprints are expected to extract to increase the diversity of features. According to the [6], the frequency domain features after removing the main trend term can be used as fingerprints for synchrophasor data. Combined with the result of VMD, the IMF_6 is the residual component that represents the trend of the attacked signal. Therefore, the frequency spectrum fingerprint can be obtained by using the Fast Fourier Transform (FFT), which can be calculated as

$$f_1(\omega)=\sum_{\tau=0}^{N_n-1}\left(f(\tau)-IMF_6(\tau)\right)e^{-j\frac{2\pi}{N_n}\tau k}\quad(k=0,1,...,N_n-1) \tag{9}$$

where the N_n is the length of signal $f(\tau)$-$IMF_6(\tau)$, $\tau\in t$.

Combined with the original time domain signal $f(t)$, all the features can be expressed as follows: $\{S^m\}=\{f(t), f_1(\omega), Ee_i, Ku_i\}$, where m = 1, 2, 3, 4 denotes the order of 4 features.

Fig. 2 is a visual example of the features from scale attacks. In Fig. 2(b), the $f_1(\omega)$ is the spectrum of the attack signal after removing IMF_6, which highlights the features of the attack signal. Theoretically, the functions of kurtosis and envelope entropy are complementary. In Fig. 2 (c) and (d), it also can be found that their changing trends are complementary.

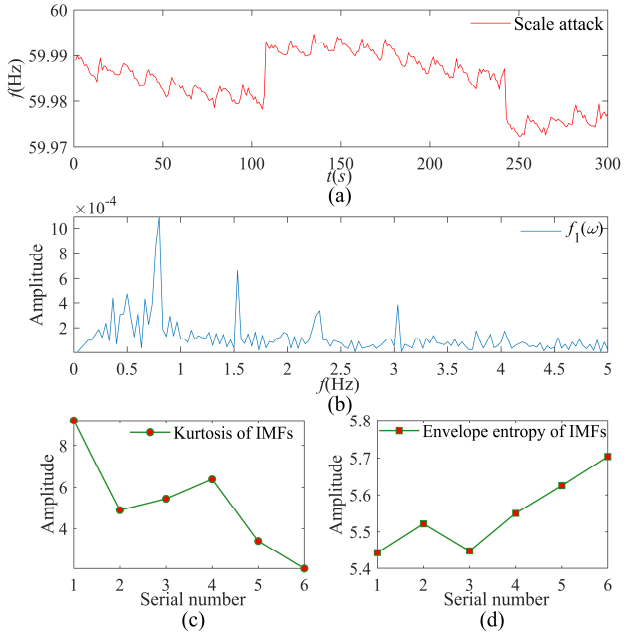

Fig. 2. Visual example of all features, (a) Scale attack signal, (b) FFT of the original signal removing IMF_6, (c) Kurtosis of IMFs, (d) Envelope entropy of IMFs.

III. Cyber-Attack Identification Using Multi-Fusion SVM

A. Principle of SVM and Multi-Fusion SVM

To achieve accurate classifications of different attack signals, an efficient classifier is required. Support Vector Machine (SVM) has a strong learning ability and generalization ability, so it is suitable for solving high dimensional and non-linear classification problems [9].

In SVM, given a set of feature samples $D=\{S^m, y_i\}$, where i = 1, 2, …, n, y_i is the label of S^m. If S^m is linearly separable, the objective of SVM is to find a hyperplane. The definition of optimal classification hyperplane can be obtained as

$$w^T S^m+b=0\quad\left(w\in R^n, b\in R\right) \tag{10}$$

However, if S^m is linearly inseparable, a kernel function is needed to map the sample to a high-dimensional feature space. Here, this high-dimensional space is linearly separable. The most commonly used kernel function is Radial Basis Function (RBF), which can be defined as

$$K_{RBF}\left(s_i^m, s_j^m, \sigma\right)=\exp\left\{-\frac{\left\|s_i^m-s_j^m\right\|^2}{2\sigma^2}\right\} \tag{11}$$

where the σ represents the kernel parameter of RBF, the s_i^m and s_j^m are samples from S^m respectively. The principle of kernel SVM classifier is shown in Fig. 3. In Fig. 3, we can intuitively see the classification process of linearly inseparable samples. Kernel function maps the linearly inseparable samples in low-dimensional space to high-dimensional space to make the samples linearly separable. Particularly, the boundary vector determines the classification performance of SVM. Therefore, it is worth mentioning that the performance of kernel function and the choice of kernel function will directly affect the locations of boundary vectors.

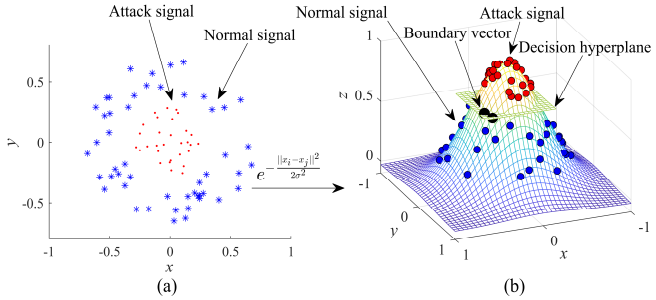

Fig. 3. Principle of kernel SVM classifier.

The local and global characteristics of different kernel functions are different. To improve the performance of SVM, different kernel functions can be combined according to Mercer's theorem [10]. Specifically, the Linear Combined Multi-kernel (LCM) method is first proposed to fuse multiple features [11], which can be expressed as

$$\begin{cases}K_{LM}\left(s_i^m, s_j^m\right)=\sum_{m=1}^{4}\gamma_m K_{RBF}\left(s_i^m, s_j^m, \sigma_i\right)\\ s.t.\ \sum_{m=1}^{4}\gamma_m=1\end{cases} \tag{12}$$

where K_{RBF} denotes RBF kernel, s_i^m represents the ith element of mth feature in S^m, σ_i represents the kernel parameter, the γ_m denotes the weights of K_{RBF}. As can be seen, different features are mapped using different kernel functions in the proposed LCM-SVM.

For different features, the proposed LCM method uses different kernel functions and parameters. Meanwhile, the importance of different features is achieved by weights γ_m. This means that each feature can match the most suitable kernel functions, thus a distinguishable hyperplane can be constructed.

It is found that the classification effect of different kernel functions is often complementary for a certain feature. For example, the RBF kernel has better local characteristics while the Polynomial Kernel (PK) function has better global characteristics. Additionally, the factors that affect the SVM

classification is the eigenvectors between hyperplane boundaries.

Based on the above consideration, a new Multi-fusion SVM is further proposed. In MSVM, two kernel functions with different parameters and weights are used for each feature. One of the kernel functions is used for mapping and the other is expected to adjust the boundary vector. In this way, the combination of kernel functions for each feature is optimal. Based on the LCM, the novel combined multi-kernel functions are redefined as

$$\begin{cases} K_{MF}\left(s_i^m, s_j^m\right) = \sum_{m=1}^{4}\left\{\mu_m K_1^m\left(s_i^m, s_j^m, p_{m,1}\right) + \right. \\ \qquad \left. \varepsilon_m K_2^m\left(s_i^m, s_j^m, p_{m,2}\right)\right\} \\ s.t.\ \sum_{m=1}^{4}\left(\mu_m + \varepsilon_m\right) = 1 \text{ and } \mu_m > \varepsilon_m \end{cases} \tag{13}$$

where K_1^m and K_2^m are two different LCM functions using for the mth feature, $p_{m,1}$ and $p_{m,2}$ are the kernel parameters of K_1^m and K_2^m respectively, μ_m and ε_m are their weights. In Equation (13), the K_1^m is considered as the primary kernel. The second kernel K_2^m is called calibration kernel function, which is used to repair the boundary vector. A smaller weight ε_m is assigned to K_2^m to limit its impact on K_1^m.

To learn the optimal classification hyperplane, the optimization framework of the MSVM is introduced as follows

$$\begin{aligned} &\min_{\omega,b,\xi_i}\left\{\frac{\|\omega\|^2}{2} + C\sum_{i=1}^{n}\xi_i^2\right\} \quad \left(\xi_i \geq 0, i = 1,...,n\right) \\ &s.t.\ y_i\left(w^T K_{MF}(s_i^m, s_j^m) + b\right) \geq 1 - \xi_i \end{aligned} \tag{14}$$

where w denotes the weight vector and b denotes the bias term of the decision plane respectively, ξ_i is the slack variable, $C \geq 0$ is the penalty coefficient. The optimal classification hyper-plane can be obtained via partial derivatives from dual Lagrange function.

By solving the dual optimization problem, for the new synchrophasor data z, the obtained decision function in the high-dimensional feature space is

$$\hat{y} = sign\left(w^T K_{MF}\left(s_i^m, z\right) + b\right) \tag{15}$$

B. *The parameters selection of MSVM*

SVM is a parameter sensitive method. This means that the model parameter selection is critically important for the accuracy of MSVM. Based on the structure of MSVM, the process of parameter optimization can be divided into two steps.

Step 1: Finding the optimal combination of K_1^m and K_2^m.

To verify the effect of multiple combinations of kernel functions, different kernel functions are tested in K_1^m and K_2^m respectively. To simplify the calculations, three commonly used kernel functions are used in this test including RBF, PK and Sigmoid Kernel (SK) functions. Results under these three kernels and the corresponding kernel parameters are listed in Table I. Here, we first select kernels for K_1^m. After a satisfactory accuracy is obtained, then the K_2^m is further debugged based on the selected K_1^m. It should be notable that the weight coefficient can be further optimized.

Table I shows that the kernel functions have a greater impact on the performance of MSVM. Particularly, the MSVM obtains 90.95 % when all the kernels are set to RBF in K_1^m. Meanwhile, the accuracy improved by nearly 3%, indicating that the SK and PK help improve the performance of the MSVM. After some tests, the final selected kernel functions are listed in Table II.

TABLE I. PERFORMANCE UNDER DIFFERENT KERNEL FUNCTIONS

Features	Optimal kernel combination ($S^1 + S^2 + S^3 + S^4$)	Accuracy (%)
K_1^m	0.1PK + 0.35SK+ 0.4RBF+0.15SK	58.46
	0.3 SK +0.2PK+ 0.4 PK +0.1SK	84.17
	0.25RBF+ 0.25RBF+ 0.25RBF+0.25SK	86.36
	0.25RBF+ 0.25RBF+ 0.25RBF+ 0.25RBF	90.95
K_2^m	0.035RBF+0.015SK+0.025PK+ 0.025RBF	92.06
	0.02SK+0.03PK+0.025RBF+ 0.025RBF	92.79
	0.03RBF+ 0.04RBF+ 0.02RBF+ 0.01RBF	93.56
	0.01SK+0.03PK+0.02SK+0.04PK	93.87

TABLE II. PERFORMANCE UNDER DIFFERENT KERNEL FUNCTIONS

Features	Optimal kernel combination $K_1^m + K_2^m$
S^1	RBF+ SK
S^2	RBF + PK
S^3	RBF + SK
S^4	RBF + PK

Step 2: Finding the weight combination of different kernel functions and parameters.

In the proposed method, eight weights are assigned to different kernel functions. To avoid getting stuck in the local minimum, the Particle Swarm Optimization (PSO) algorithm is utilized to find the optimal kernel weights and kernel parameters. Specifically, the parameters to be optimized are set to the position of the particle. The classification error of the VMD-MSVM is recorded as fitness function in PSO. Then the parameters can be optimized automatically.

The overall structure of the framework is shown in Fig.4. It shows that the cyber-attack identification of synchrophasor data can be divided into three steps. The IMFs of synchrophasor data are extracted using VMD. Then two features are designed from the IMFs. Combining two features from time and frequency domain, all these four features are fused and mapped using the proposed MSVM.

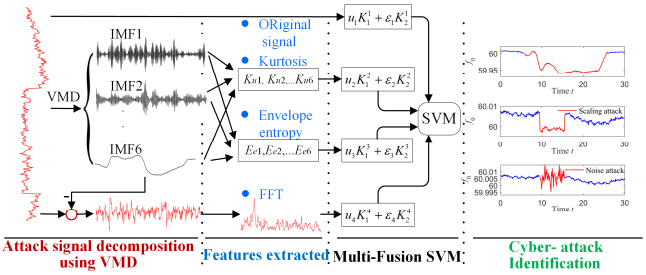

Fig. 4. The overall structure of proposed VMD-MSVM.

IV. Experiment

To verify the actual detection effect of the proposed VMD-MSVM, the synchrophasor frequency data in FNET/Grideye sever from five locations in Eastern Interconnection (EI) are used. Here, six different types of cyber-attacks are selected according to [7][12], including noise, scaling, data loss, replay, false frequency shock and transient oscillation. Using the actual data, 3000 samples are generated for each type of attack by simulation. Under sampling rate of 10 Hz, each sample is truncated with a 30s window length, which corresponds to a length of 300.

Additionally, the samples are divided into three categories including training, verification and testing data set during the model validation. Using the PSO method, the optimized parameters are selected as follows: u_m = [0.174, 0.181, 0.129, 0.275], ε_m = [0.035, 0.072, 0.082, 0.052]. The penalty coefficient C is set to 2000. The kernel parameters of K_1^m and K_2^m are set to: K_1^m = {RBF(6.1), RBF(0.86), RBF(7.92), RBF(0.82)}, and K_2^m = {SK(1.51,1.51), PK(1,1.06), SK(2.34,2.34), PK(1,1.18)} respectively.

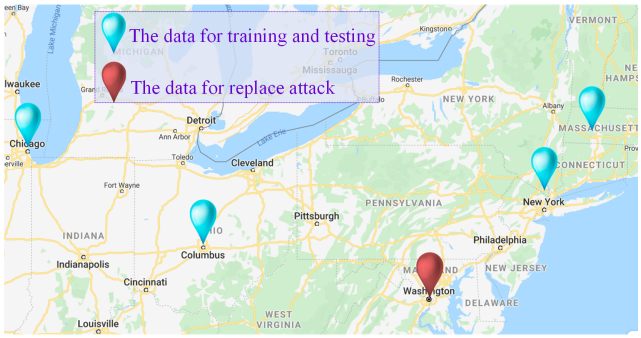

Fig. 5. The locations of synchrophasor data.

A. Performance comparison with different SVMs

To compare the effects of the proposed MSVM framework, the original SVM and LCM-SVM are tested. Additionally, different cyber-attack strengths are used to test the sensitivity. Considering that the error of frequency measurement equipment is generally lower than 5 mHz, thus the minimum attack strength is set to 5 mHz. The strengths of 10 mHz and 20 mHz are also tested. In this test, 500 samples are randomly selected as training data. To make a fair comparison, all the SVM methods are optimized by using PSO. The optimized kernel parameters of LCM-SVM are: γ_m = {0.36, 0.31, 0.19, 0.14}; K_{LM} = {RBF(6.58), RBF(0.72), RBF(7.85), RBF(1.72)}. The accuracy comparison between different frameworks are listed in Table III.

TABLE III. Classification Accuracy by Different Frameworks

SVM methods	Accuracy(%)			Test time per sample (ms)
	5 mHz	10 mHz	20 mHz	
Original DVM	90.71	94.08	93.97	**0.038**
LCM-SVM	94.90	95.67	96.22	0.039
MSVM	**95.64**	**96.96**	**96.51**	0.067

It can be seen from Table III that the original SVM has the lowest accuracy at different attack strengths. The LCM-SVM has 94.90% accuracy, which is nearly 4.2% higher than the original SVM under 5 mHz attack strength. Moreover, the MSVM has the highest classification accuracy among different SVM frameworks, indicating the effectiveness of the calibration kernel K_2^m. The test time of MSVM is higher due to multi-kernel computing, the real-time can still be satisfied.

B. Comparison of MSVM, DT and ANN

To compare the performance of MSVM with some common classification algorithms. Three different classification frameworks are selected including the DT [4], ANN and k-Nearest Neighbors (kNN) [13]. A three-layer ANN is used, and the hidden nodes are optimally set to 300. The number of neighbors of kNN is optimally selected as 6 using grid search. To match feature dimensions, the input features S^m are stitched together for ANN, DT and kNN. In this case, 500 training samples are used. The results are listed in Table IV.

TABLE IV. Comparison of Classification Accuracy

Identification framework	Accuracy (%)	Test time per sample (ms)
DT	80.27	0.049
ANN	85.93	0.051
kNN	75.31	**0.046**
VMD-MSVM	**95.64**	0.067

As can be seen from Table IV, the ANN reaches 85.93 %, which performs better than DT and kNN. However, this accuracy is still 9.71 % lower than the MSVM. The reason is that MSVM has the ability to better integrate multiple input information.

The generalization ability of the model, namely the learning ability under different training samples, reflects the recognition results for unknown attack signals. If the model can get higher recognition accuracy with fewer training samples, it means that the practicability of the model is better. To verify the generalization ability of the proposed method, we randomly select 1% to 20% of the sampling data from each attack category as the training data. The remaining samples are the test data set. The accuracy under different ratios samples are recorded as shown in Fig. 6.

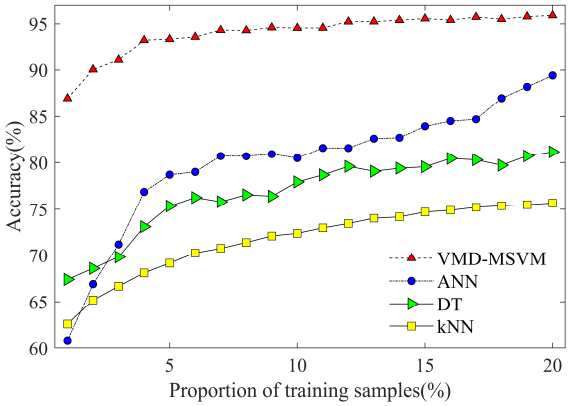

Fig. 6. Comparison of generalization ability under different number of training samples.

The results show that as the amount of training samples increases, the model accuracy gradually increases. When the training sample size is less than 5%, the accuracy of the model changes quickly. When the number of training samples

is higher than 10%, the test accuracy tends to be stable. This is because as the sample features increase, the learning space of MSVM is also expanded. Compared with different methods, the MSVM obtains better accuracy under multiple sample spaces.

V. Conclusion

To detect cyber-attacks on power systems, a multi-scale feature fusion based variational mode decomposition and multi-fusion SVM are proposed. Utilizing the decomposition modal functions, three distinctive features are extracted from VMD. The recognition results under different features indicate that the modal component IMFs contains unique attack components. Thereafter, four different scales features are fused and automatically learned based on the proposed MSVM. Using the actual synchrophasor data, the results of different single and multiple kernel functions show that combined kernel function has a better learning ability. Moreover, these multiple kernels further optimize classification capabilities. Experiments with different attack strengths and training samples are conducted to verify the proposed VMD-MSVM. Compared with commonly used classifiers, the result shows that the VMD-MSVM has strong adaptability and robustness.

References

[1] R. Khan, K. McLaughlin, D. Laverty and S. Sezer, "Analysis of IEEE C37.118 and IEC 61850-90-5 synchrophasor communication frameworks," *2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA*, pp. 1-5, 2016.

[2] A. S. Musleh, G. Chen and Z. Y. Dong, "A Survey on the Detection Algorithms for False Data Injection Attacks in Smart Grids," *IEEE Transactions on Smart Grid*, pp. 1-1, 2019.

[3] H.-M. Chung, W.-T. Li, C. Yuen, W.-H. Chung and C.-K. Wen, "Local cyber-physical attack with leveraging detection in smart grid," *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2017.

[4] K. Khanna, B. K. Panigrahi and A. Joshi, "AI-based approach to identify compromised meters in data integrity attacks on smart grid," *IET Generation, Transmission & Distribution*, vol. 12, no. 5, pp. 1052-1066, 13 Mar. 2018.

[5] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar and S. Mishra, "Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005-1016, June 2016.

[6] L. Wei, D. Gao and C. Luo, "False Data Injection Attacks Detection with Deep Belief Networks in Smart Grid," *2018 Chinese Automation Congress (CAC), Xi'an, China*, pp. 2621-2625, 2018.

[7] W. Qiu, Q. Tang, Y. Wang, L. Zhan, Y. Liu and W. Yao, "Multi-view Convolutional Neural Network for Data Spoofing Cyber-Attack Detection in Distribution Synchrophasors," *IEEE Transactions on Smart Grid*, pp. 1-12, 2020. doi: 10.1109/TSG.2020.2971148

[8] Dragomiretskiy, K., and D. Zosso. "Variational Mode Decomposition," *IEEE Transactions on Signal Processing* 62, no. 3 (2014): 531-44.

[9] Vapnik V N , "The Nature of Statistical Learning Theory," *Springer*, 2000.

[10] MERCER J, " Functions of positive and negative type and their connection with the theory of integral equations," *Philosophical Transactions of the Royal Society London*, 1909, 209: 415-446.

[11] Q. Tang, W. Qiu and Y. Zhou, "Classification of Complex Power Quality Disturbances Using Optimized S-Transform and Kernel SVM," *IEEE Transactions on Industrial Electronics*, pp. 1-9, 2019.

[12] H. M. Khalid and J. C. -. Peng, "A Bayesian Algorithm to Enhance the Resilience of WAMS Applications Against Cyber Attacks," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2026-2037, July 2016.

[13] E. M. d. L. Pinto, R. Lachowski, M. E. Pellenz, M. C. Penna and R. D. Souza, "A Machine Learning Approach for Detecting Spoofing Attacks in Wireless Sensor Networks," *2018 IEEE 32nd International Conference on Advanced Information Networking and Applications (AINA), Krakow*, pp. 752-758, 2018.