Enhancing Detection Accuracy of Cyber Attacks Through Dimensionality Reduction

Ehsan Hallaji, Roozbeh Razavi-Far and Mehrdad Saif

Department of Electrical and Computer Engineering, University of Windsor, Canada
E-mail: hallaji@uwindsor.ca; roozbeh@uwindsor.ca; msaif@uwindsor.ca

The importance of cyber-security has led to long-standing endeavors dedicated to the design of intrusion detection systems (IDS). Nevertheless, the performance of these data-driven techniques is highly dependent on data quality. Incorporating dimensionality reduction techniques into a hybrid intrusion detection system, we aim to study the effect of dimensionality reduction on the performance of intrusion detection. By this mean, the efficiency of the intrusion detection systems is increased by processing a smaller feature space. Moreover, the reduced feature space also increases the detection accuracy, as redundant and meaningless features are removed in the new feature space. Furthermore, the intrinsic structure of the data is improved, that is different states of the system become more discriminant after dimensionality reduction. For this mean, various state-of-the-art dimensionality reduction techniques are selected. Then, a simulation is performed on a Supervisory Control and Data Acquisition (SCADA) system, which resembles a gas pipeline control system introduced by Morris et al. (2011). A comparative study is then performed to suggest the best dimensionality reduction algorithm in these experiments. The experiments indicate the general improvement of detection accuracy when dimensionality reduction techniques are combined with the IDS in terms of accuracy and standard deviation.

Keywords: : Intrusion detection, cyber-physical systems, gas pipeline, dimensionality reduction, machine learning, SCADA.

1. Introduction

With the advancement of technology and the integration of information technology with the physical systems, security has became an immense concern. Supervisory Control And Data Acquisition Sytems (SCADA) are no exception to this matter, as multiple threats can be posed to the SCADA network through the application layer (Morris et al., 2011).

Various research efforts have been dedicated to the design of Intrusion Detection Systems (IDS), which is used to tackle the security threats in cyber-physical systems (Beaver et al., 2013; Morris et al., 2011). An IDS usually relies on the prior knowledge or training data to construct a detection model. Then, the IDS uses the respective model to monitor the traffic data to monitor the network traffic. Malicious activities are then detected w.r.t. the detection model at hand whether through matching the network activity with known attack patterns or treating it as an anomaly that represents suspicious activity.

On the other hand, IDS frameworks similar to diagnostic systems usually are very dependant to the quality of data due to their data-driven nature (Beaver et al., 2013; Razavi-Far et al., 2017; Razavi-Far et al., 2009; Razavi-Far et al., 2016). While the recorded data is usually acquired through sensors, the raw signals may not lead to a discriminant feature space. This might happen due to a number of reasons Razavi-Far et al. (2019); Razavi-Far et al. (2020). For instance, considering that the recorded data via each sensor represents a feature in the network, similar feature may be observed under unforeseen circumstances. Moreover, the number of sensors in the system is directly related to the dimensionality size of data, that is a large number of sensors results in high-dimensional datasets (Razavi-Far et al., 2019). All of these issues challenge the IDS in a different way that bring about performance deterioration.

The aforementioned challenges can be eliminated through various approaches such as the feature selection and Dimensionality Reduction (DR). The former generally select the beneficial features for constructing the detection model and monitors the traffic based on those features. The latter, which is studied in this paper, transforms the features space into a smaller space, often low-dimensional, via computing a transformation matrix (Tipping and Bishop, 1999; Zhang and Zha, 2004). Reducing the dimensionality size, using either of these methods, usually result in performance enhancement (Hallaji, 2018), as a more discriminative and efficient feature space is used for detection procedure. Furthermore, these methods may devise the class labels of the training data, which is known as supervised learning, to apply additional constraints to the new space in order to improve the classification accuracy (FIsher, 1936; Globerson and Roweis, 2005). Oth-

Proceedings of the 30th European Safety and Reliability Conference and
the 15th Probabilistic Safety Assessment and Management Conference
Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio
Copyright © ESREL2020-PSAM15 Organizers.*Published by* Research Publishing, Singapore.
ISBN: 978-981-14-8593-0; doi:10.3850/978-981-14-8593-0_4593-cd

erwise, the smaller feature space can be obtained via unsupervised learning that do not depend on class labels (Jolliffe, 1986; Belkin and Niyogi, 2003; Coifman and Lafon, 2006). Combination of these methods is also possible via semi-supervised learning, which aims to make use of both labeled and unlabeled data (Razavi-Far et al., 2017, 2019).

While DR techniques have different assumptions and rely on different criteria, they may not show similar behaviour to all cases. Another issue of concern is the compatibility of DR algorithms with classification methods that are commonly paired to form an IDS.

In this work, we aim to study the effect of DR on the detection accuracy of cyber-attacks on a cyber-physical case. For this mean, ten advanced DR algorithms are employed to improve the data quality. In addition, two classifiers, k Nearest Neighbours (kNN) and Decision Tree (DT) are combined with the DR methods to form an IDS. By this mean, the combination of utilized DR methods with these classifiers is studied to evaluate their compatibility with distance-based and DT-based classifiers. In this process, we simulate the intrusion detection on a gas pipeline cyber-physical system using the dataset provided by Morris et al. (2011).

The remainder of this paper is organized as follows. Section 2 contains the literature review on the dimensionality reduction techniques. Section 3 presents the design of the IDS. Section 4 report and analyze the experimental results. Finally, the paper is concluded in Section 5.

2. Related work

Here, we briefly give an overview of the DR algorithms used throughout our experiments. Unless stated otherwise, the following DR techniques make use of unsupervised learning.

2.1. *Probabilistic Principal Component Analysis*

While the Principal Component Analysis (PCA) (Jolliffe, 1986) is not based on any probabilistic model, the Probabilistic PCA (PPCA) (Tipping and Bishop, 1999) aims to formulate the DR process as a maximum likelihood procedure, using the probability density function. The principal axes in PPCA are then estimated using the Expectation Maximization (Dempster et al., 1977).

2.2. *Linear Discriminant Analysis*

Linear Discriminant Analysis (LDA) (FIsher, 1936) is a supervised linear transformation method that is commonly used for DR. LDA aims to estimate a projection matrix by which between-class covariance is maximized. Satisfying this objective, classification performance is expected to be improved, as samples of opposite classes are well-separated in the new feature space.

2.3. *Laplacian Eigenmaps*

Assuming that the data lies on a low-dimensional manifold in a high-dimensional feature space, Laplacian Eigenmaps (Belkin and Niyogi, 2003) (LE) uses neighborhood weights to construct a graph on the data. Then, obtained graph can be considered as the graph structure. Given that the constructed manifold preserves local properties, transformation can take place using the Laplace–Beltrami operator.

2.4. *Local Tangent Space Alignment*

Local Tangent Space Alignment (LSTA) (Zhang and Zha, 2004) is a non-linear method, which constructs the data manifold using the tangent spaces of data samples. Then, the global coordinates of samples are obtained w.r.t. the underlying manifold, by aligning the estimated tangent spaces.

2.5. *Neighborhood Preserving Embedding*

Aiming to maintain the local manifold structure on the data, Neighborhood Preserving Embedding (NPE) (Xiaofei He et al., 2005) constructs a weighted matrix to capture the relationship between data samples. This matrix is formed w.r.t. the linear correlation of adjacent samples. The transformation is then carried out in a way the captured local structure is preserved in the new feature space.

2.6. *Diffusion Maps*

Diffusion Maps (DM) (Coifman and Lafon, 2006) is a non-linear graph-based DR algorithm that utilizes eigenvalues of Markov chain matrix in order to create coordinates of data samples in a Euclidean space. In the embedded space, the distance between samples is calculated through diffusion distance of probability distributions corresponding to each of those samples.

2.7. *Coordinated Factor Analysis*

Assuming that the data is representable on a low-dimensional manifold in a high-dimensional space, Coordinated Factor Analysis (CFA) (Verbeek, 2006) builds a linear model using the mixture of factor analyzers to model the density w.r.t. the manifold. Moreover, CFA enables global parameter adjustment for the estimated manifold, which enhances the convergence rate.

2.8. *Maximally Collapsing Metric Learning*

Maximally Collapsing Metric Learning (MCML) (Globerson and Roweis, 2005) is another super-

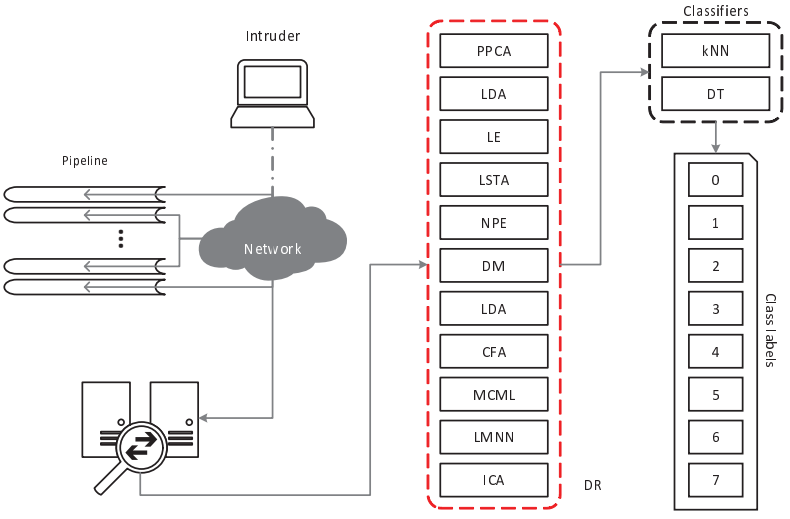

Fig. 1. Illustrative diagram of the intrusion detection system.

vised DR technique that endeavours to reach a geometrically ideal condition, where samples of each class are placed on the same point in the feature space. Accordingly, each class should be pushed far away from other classes. This has been done through a stochastic selection rule, for which a convex optimization problem has been solved.

2.9. *Large Margin Nearest Neighbor*

The goal of Large Margin Nearest Neighbor (LMNN) (Weinberger and Saul, 2009) is to reduce the dimensionality in a way samples in a neighborhood share the same label, and opponent classes become well-separated in the new space. This supervised method tries to find a linear mapping, by which the training samples satisfy the stated goal w.r.t. the Mahalanobis distance.

2.10. *Independent Component Analysis*

Independent component analysis (ICA) (Hyvärinen and Oja, 2000) is a DR technique mainly used for finding root components of data in a high-dimensional feature space. Assuming that the data is formed through independent sources, ICA searches for statistically independent and non-gaussian components within the data. In other words, ICA is a special case of the so called Blind Source Separation.

3. Intrusion Detection System

The ultimate goal of IDS in this case study is to detect and classify seven types of cyber-attacks in the gas pipeline network in addition to the normal state, as shown in Table 1. To reach this goal, the traffic data is monitored, as illustrated in Fig. 1. This traffic data will then go through the DR module. Once the dimensionality of data is reduced, a new detection model is trained based on the given classifiers.

Table 1. Cyber-attacks injected in the SCADA network.

Classes	Types of injected attack
0	N/A (The system is safe).
1	Naive malicious response.
2	Complex malicious response.
3	Malicious state command.
4	Malicious parameter command.
5	Malicious function command.
6	Denial-of-service attack.
7	Reconnaissance attack.

During the testing phase, once the data is transformed into the lower-dimensional feature space in the DR module, the reduced data is classified into 8 different classes, as shown in Fig. 1, which are detailed in Table 1.

The most critical condition that can severely challenge the designed IDS is the presence of a non-stationary environment, in which concept drift exists. However, we do not consider this case as it is already addressed in the literature and can be applied to this case study as well (Razavi-Far et al., 2019).

4. Experimental Results

We initially report the experimental setting used in the simulations. Then, the obtained results are analyzed in terms of accuracy (Acc.) and standard deviation (Std.). Throughout the experiments, we compare the DR results to a baseline condition, where DR is not used prior to the classification.

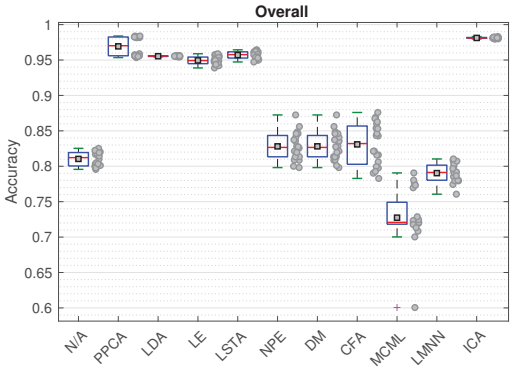

Fig. 2. Overall results of dimensionality reduction in terms of accuracy. Solid circles denote obtained accuracies through the cross-validation, using kNN and DT.

4.1. *Experimental Setting*

The utilized dataset has originally 26 features and 97020 samples. The optimal dimensionality of the new feature space is obtained via naive search, where the search ranges are found empirically.

All experiments are run through a nested 10-fold cross-validation procedure Razavi-Far et al. (2017). This nested structure, enable the parameter tuning of classifiers and DR techniques. For this mean, the grid search algorithm is used to ensure the optimal classification performance resulted by DR methods.

4.2. *Results Analysis*

The overall results of the dimensionality reduction are shown in Fig. 2 in terms of accuracy. Solid circles in this figure indicate recorded accuracies through the cross-validation, using kNN and DT; and solid squares show the averaged accuracy. Looking through Fig. 2, one can realize that ICA is the most stable algorithm compared to others. Furthermore, ICA results in the highest overall accuracy. PPCA is ranked second, as it has the closest overall accuracy to ICA. Nevertheless, its overall stability is significantly lower than ICA, albeit still acceptable. LSTA, LDA and LE are more similar to ICA in terms of stability, compared to PPCA. However, they are ranked from third to fifth considering their averaged accuracies. This is while CFA, DM, NPE, baseline, LMNN and MCML are ranked from sixth to eleventh. In other words, LMNN and MCML have failed to improve the classification performance compared to that of baseline. This might be due to the sensitivity of these algorithms to the choice of classifier, that is they require certain classifiers to result in satisfactory results.

More specifically, Fig. 3 shows the results of the cross-validation for each of the ten iterations. Separating the results of kNN and DT, Fig. 3

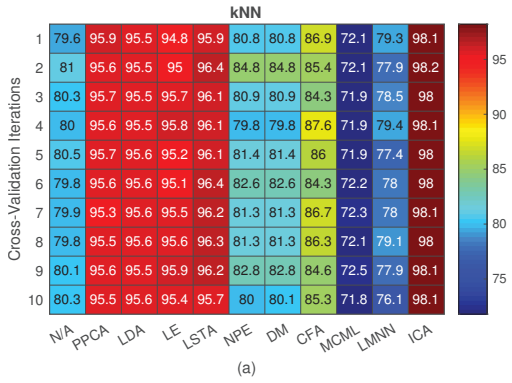

Fig. 3. Cross-validation results in terms of accuracy (%) using kNN and DT. The spectrum is normalized w.r.t. the maximum and minimum for each classifier.

imply that the obtained accuracies are dependant to the choice of classier, as the performance is different in Fig. 3(a–b). For instance, in Fig. 3(a), ICA outperforms other DR methods when it is combined with kNN. However, when DT is used for classification, ICA and PPCA result in almost similar accuracies, as illustrated in Fig. 3(b).

The overall standard deviation of DR methods, which can be observed in Fig. 2, is mainly due to the sensitivity of DR algorithms to the choice of classifier. This can be seen by comparing panels of Fig. 3(a–b). For instance, PPCA CFA, MCML and baseline condition are more stable considering kNN and DT separately. Conversely, due to the different accuracies of kNN and DT, the standard deviation increases as their results are combined. On the other hand, ICA, LDA, LE and LSTA are not sensitive to the choice of classifier, and, thus, they result in lower standard deviations.

The averaged accuracies and standard deviations obtained through the cross-validation are listed in Table 2, in which the best performances are marked in bold. Additionally, the reduced dimensionalities d, by which the DR algorithms reached an optimal performance, are reported in

Table 2. Averaged intrusion detection performance in terms of accuracy. Best accuracies are highlighted in bold.

DR	Accuracy(%)			d	Rank
	kNN	Decision Tree	Overall		
Baseline	80.1286 ± 0.4223	81.9458 ± 0.3267	81.0372 ± 1.0020	26	9
PPCA	95.5970 ± 0.0615	**98.2580** ± 0.0757	96.9275 ± 1.3700	2	2
LDA	95.5370 ± 0.1501	95.5370 ± 0.0435	95.5370 ± 0.0423	15	4
LE	95.3977 ± 0.0435	94.4988 ± 0.3413	94.9483 ± 0.5753	15	5
LSTA	96.1500 ± 0.3652	95.2511± 0.2942	95.7006 ± 0.5245	15	3
NPE	81.5876 ± 0.2127	84.0270± 1.6108	82.8073 ± 1.9619	14	8
DM	81.5976 ± 1.4917	84.0270 ± 1.6108	82.8123 ± 1.9545	12	7
CFA	85.7444 ± 1.4802	80.4311 ± 1.3790	83.0878 ± 2.9920	22	6
MCML	72.0717 ± 1.1462	73.4093 ± 5.7726	72.7405 ± 4.0345	8	11
LMNN	78.1706 ± 0.2140	79.8925 ± 1.0012	79.0316 ± 1.3109	8	10
ICA	**98.0607** ± 0.9890	98.1766 ± 0.0684	**98.1187** ± 0.0869	6	1

this table. Table 2 indicates that kNN results in the optimal accuracy when combined with ICA. Furthermore, the best accuracy of DT is attained when PPCA is used for dimensionality reduction. However, as mentioned before, ICA is ranked first when the overall performance is taken into account. This is while ICA and PPCA are more stable with DT and kNN, respectively. Thus, when kNN is used for classification, ICA, LSTA, PPCA, LDA, LE, CFA, DM, NPE, baseline, LMNN and MCML are ranked from first to 11-th. On the other hand, considering DT for the classification, PPCA, ICA, LDA, LSTA, LE, NPE, DM, baseline, CFA, LMNN and MCML are ranked from first to 11-th, respectively.

Fig. 4 illustrates the achieved performance improvement for each DR algorithm. Fig. 4(a) shows that the accuracy always increases when PPCA, LDA, LE, LSTA, NPE and ICA are used for dimensionality reduction. CFA, on the other hand, can enhance the classification accuracy only when it is combined with kNN. As for the choice of classifier, PPCA, NPE, DM, MCML and LMNN seem to be more compatible with DT, i.e., see Fig. 4(a). The rest of the DR methods are suggested to be used along with kNN. Overall, ICA outperforms all other techniques and results in the maximum accuracy when coupled with kNN.

Taking standard deviation into account, Fig. 4(b) indicates to what extent it is decreased by each DR technique. Glancing at this figure, one can realize that PPCA, LDA, LE, LSTA and ICA always improve the standard deviation. Nevertheless, PPCA resulted in an increase of standard deviation when the overall results are considered, that is it shows sensitivity to the choice of classifier. This is while NPE, DM and CFA always lead to higher instability. Similarly, MCML generally worsen the standard deviation; however, stability is improved when it is used along with kNN. Finally, the highest stability improvement is brought about by combining LDA, with a slight difference with ICA.

Hence, the detection performance of cyber-attacks is desirably improved in terms of accuracy and standard deviation through dimensionality reduction. The highest detection accuracy is achieved through combination of ICA and kNN.

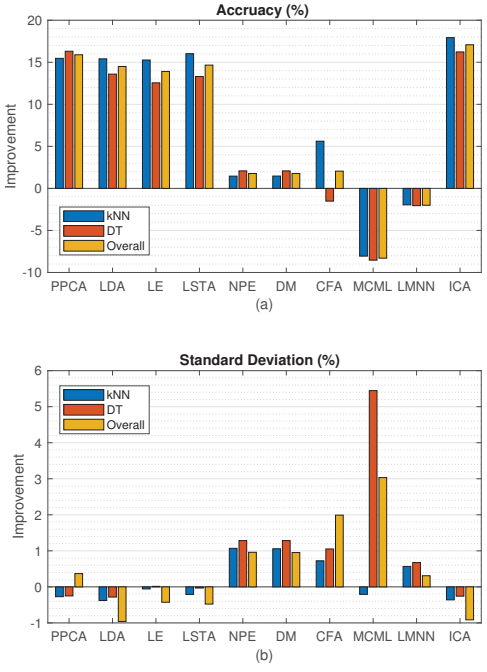

Fig. 4. Performance improvement by each DR method. The criterion is the difference of post-DR performance with the baseline, in terms of accuracy and standard deviation. The improvements are illustrated based on the results of both kNN and DT.

5. Conclusion

In this paper, an intrusion detection system is designed to study the effect of DR on the intrusion detection accuracy on a gas pipeline cyber-physical system. In this process, ten advanced DR methods are compared and studied. Furthermore,

the compatibility of these methods with distance-based and decision tree-based classifiers is assessed within the simulated IDS. Finally, the best combination is introduced to address intrusion detection on the selected case study. The results indicate satisfactory performance improvement of the intrusion detection in a gas pipeline system when DR is utilized.

References

Beaver, J. M., R. C. Borges-Hink, and M. A. Buckner (2013, Dec). An evaluation of machine learning methods to detect malicious scada communications. In *12th International Conference on Machine Learning and Applications*, Volume 2, pp. 54–59.

Belkin, M. and P. Niyogi (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation 15*(6), 1373–1396.

Coifman, R. R. and S. Lafon (2006). Diffusion maps. *Applied and Computational Harmonic Analysis 21*(1), 5 – 30. Special Issue: Diffusion Maps and Wavelets.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological) 39*(1), 1–22.

FIsher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics 7*(2), 179–188.

Globerson, A. and S. Roweis (2005). Metric learning by collapsing classes. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, NIPS'05, Cambridge, MA, USA, pp. 451–458. MIT Press.

Hallaji, E. (2018). Semi-supervised learning for diagnosing faults in electromechanical systems. Electronic theses and dissertations. 7470, Univresity of Windsor.

Hyvärinen, A. and E. Oja (2000). Independent component analysis: algorithms and applications. *Neural Networks 13*(4), 411 – 430.

Jolliffe, I. (1986). *Principal Component Analysis* (1 ed.). Springer Series in Statistics. Springer-Verlag New York.

Morris, T., A. Srivastava, B. Reaves, W. Gao, K. Pavurapu, and R. Reddi (2011). A control system testbed to validate critical infrastructure protection concepts. *International Journal of Critical Infrastructure Protection 4*(2), 88 – 103.

Razavi-Far, R., B. Cheng, M. Saif, and M. Ahmadi (2020). Similarity-learning information-fusion schemes for missing data imputation. *Knowledge-Based Systems 187*, 104805.

Razavi-Far, R., H. Davilu, V. Palade, and C. Lucas (2009). Model-based fault detection and isolation of a steam generator using neuro-fuzzy networks. *Neurocomputing 72*(13), 2939 – 2951.

Razavi-Far, R., M. Farajzadeh-Zanjani, S. Chakrabarti, and M. Saif (2016). Data-driven prognostic techniques for estimation of the remaining useful life of lithium-ion batteries. In *IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–8.

Razavi-Far, R., M. Farajzadeh-Zanjani, and M. Saif (2017). An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors. *IEEE Transactions on Industrial Informatics 13*(6), 2758–2769.

Razavi-Far, R., M. Farajzadeh-Zanjani, B. Wang, M. Saif, and S. Chakrabarti (2019). Imputation-based ensemble techniques for class imbalance learning. *IEEE Transactions on Knowledge and Data Engineering*, 1–1.

Razavi-Far, R., E. Hallaji, M. Farajzadeh-Zanjani, and M. Saif (2019). A semi-supervised diagnostic framework based on the surface estimation of faulty distributions. *IEEE Transactions on Industrial Informatics 15*(3), 1277–1286.

Razavi-Far, R., E. Hallaji, M. Farajzadeh-Zanjani, M. Saif, S. H. Kia, H. Henao, and G. Capolino (2019, Aug). Information fusion and semi-supervised deep learning scheme for diagnosing gear faults in induction machine systems. *IEEE Transactions on Industrial Electronics 66*(8), 6331–6342.

Razavi-Far, R., E. Hallaji, M. Saif, and G. Ditzler (2019, Jan). A novelty detector and extreme verification latency model for nonstationary environments. *IEEE Transactions on Industrial Electronics 66*(1), 561–570.

Razavi-Far, R., E. Hallaji, M. Saif, and L. Rueda (2017, Dec). A hybrid scheme for fault diagnosis with partially labeled sets of observations. In *16th IEEE International Conference on Machine Learning and Applications*, pp. 61–67.

Tipping, M. E. and C. M. Bishop (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61*(3), 611–622.

Verbeek, J. (2006, Aug). Learning nonlinear image manifolds by global alignment of local linear models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*(8), 1236–1250.

Weinberger, K. Q. and L. K. Saul (2009, June). Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research 10*, 207–244.

Xiaofei He, Deng Cai, Shuicheng Yan, and Hong-Jiang Zhang (2005, Oct). Neighborhood preserving embedding. In *Tenth IEEE International Conference on Computer Vision*, Volume 1, pp. 1208–1213.

Zhang, Z. and H. Zha (2004). Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal on Scientific Computing 26*(1), 313–338.