Data-driven Methods for Sensor-less Condition Monitoring of Electric Drive-train from Phase Current Measurements

Md Jakir Hossain
Electrical Engineering Department, University of South Florida, Tampa, Florida, USA
mdjakir@mail.usf.edu

***Abstract*—Electric drive-train e.g. electric motors are a vital component of electro-mechanical rotary systems. Real-time condition monitoring of these machine is a necessary requirements for most of the smart systems like electric vehicle. But it generally requires additional sensor installation which bring complexity and expanses. The goal of this work is to investigate whether it is possible to do condition monitoring from external parameters like phase currents and phase angle regardless of any additional sensor installation except for regular volt-Ampere measurements. The methods discussed in this work is data-driven and does not require prior knowledge of the machine components. Thus making the condition monitoring autonomous and sensor-less.**

***Index Terms*—Data-driven, Electric Drive-train, Sensor-less, Phase-current, KNN, LDA, QDA, SVM, Decision-tree.**

I. Introduction

The active condition monitoring autonomous systems is always a central research topic. In this work, a method to perform sensor-less diagnosis of an autonomous electric drive train from measured phase angle data will be reviewed. The unit under investigation is composed of a synchronous motor and several attached components, e. g. bearings, axles, and a gear box. It, thus, represents a typical and crucial component within a plant or other machinery. Damage to the drive may cause severe disturbances and increases the risk of encountering breakdown costs. Condition monitoring for such applications usually requires additional sensors. But the goal is to directly use the phase currents of the motor for health state characterization of the entire unit. The measurement does not involve any sensors apart from the built-in current sensors of the motor control. In this respect the diagnosis performs autonomously.

Data Set Characteristics:	Multivariate	**Number of Instances:**	58509	**Area:**	Computer
Attribute Characteristics:	Real	**Number of Attributes:**	49	**Date Donated**	2015-02-24
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	55262

Fig. 1: brief description of the sensor-less drive diagnostic data-set used [1].

Features are extracted from motor phase current; the motor has intact and defective components. The Empirical Mode Decomposition (EMD) was used to generate a new database for the generation of features. The first three intrinsic mode functions (IMF) of the two-phase currents and their residuals (RES) were used and broken down into sub-sequences. For each of this sub-sequence, the statistical features mean, standard deviation, skewness and kurtosis were calculated. This results in 11 different classes with different conditions. Each condition has been measured several times by 12 different operating conditions, this means by different speeds, load moments and load forces. The current signals are measured with a current probe and an oscilloscope on two phases [1].

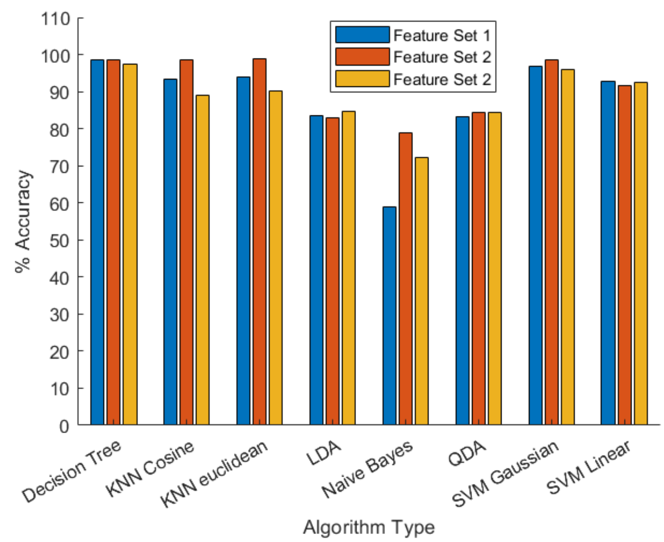


Fig. 2: Comparison of Accuracy of all the algorithm applied.

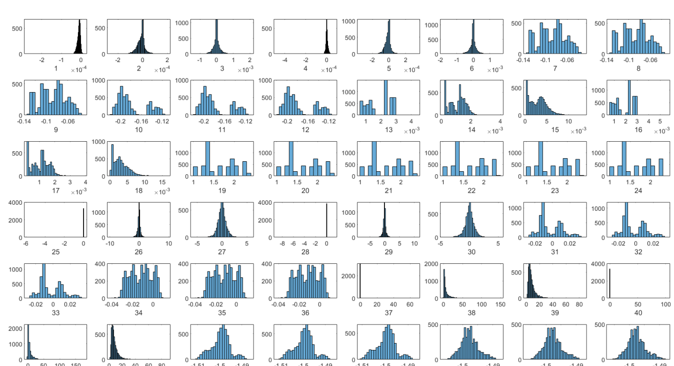


Fig. 3: Histogram plot of all the features of the data-set.

From the problem statement and description of the data-set it is evident that this sensor-less diagnostic process is strictly a multi-class classification problem. There are total 11 operating condition of that machine including normal operation and 10 faulty operating condition which leads to 11 classes in total. Several classification methods such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Discriminant Analysis(LDA, QDA), Decisions Tree (DT) etc. will be tested and also in the data pre-processing steps, missing value replacement, standardization, feature selection, dimensionality reduction will be implemented.

II. Data Pre-processing

The data-set does not contains any missing or NaN values. In general missing values/ NaN values are replaced with column mean/median/mode values. To extract the features that has maximum prediction capability two different method will be followed.

A. *Feature Selection: Visual Inspection*

Figure 2 shows the histogram plot of all the data features. From figure 2 we can see that most of the features has Gaussian or Gaussian-mixture distribution. Few of the features has multimodal distribution as well. By analyzing visually I have found several feature sets that are linear multiple of each other and has same distribution. So, kept one feature form each set. For example feature [7,8,9], [10,11,12], [43,44,45], [46, 47,48] etc. are linear multiple of each other and has same distribution, so kept one feature from each group. Thus reduced the feature from 48 to 29 and leads to a new predictive feature set.

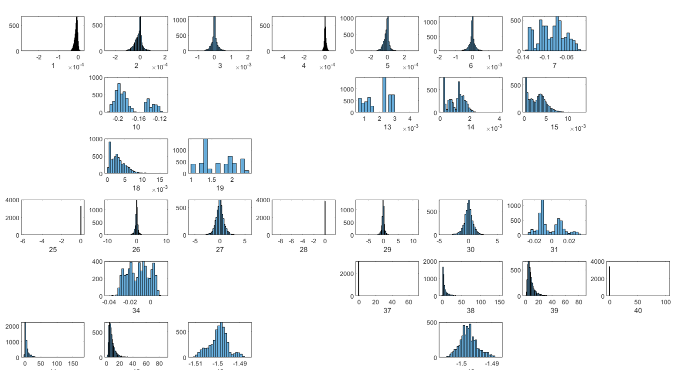


Fig. 4: Histogram plot of visually reduced feature set.

To evaluate the prediction capability of new reduced feature data-set, I have fitted multivariate gaussian models to data for each class and calculated the likelihood of each data instances over each model. The model which gives maximum likelihood indicates that the data instance is belong to that class. This way the classification capability is judged for newly selected feature set.

From figure 4 we can see that new reduced feature set (feature set 2 with 29 feature) has nearly equal miss-classification rate as using all features. Thus without sacrificing much accuracy we can cut the computational complexity to half.

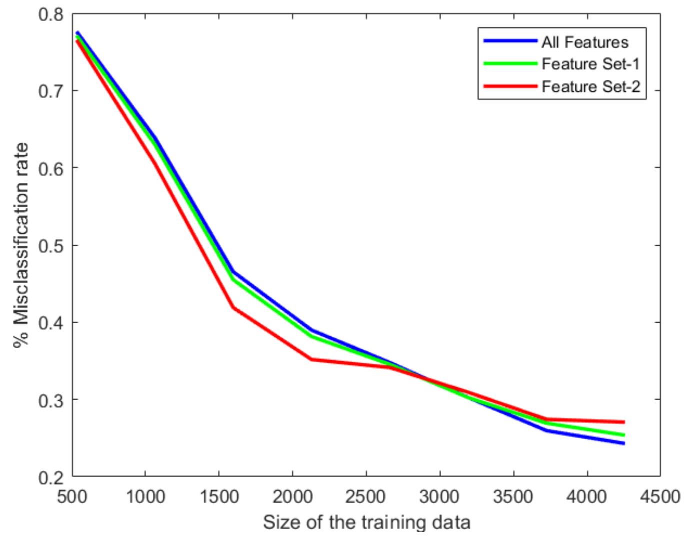


Fig. 5: Miss-classification rate over multivariate gaussian model based maximum likelihood classifier for three different feature set.

B. *Feature Selection: Rank Importance of Predictors Using 'Relieff' Algorithm*

Relief algorithm take a data set with n instances of p features, belonging to two known classes. Within the data set, each feature should be scaled to the interval [0 1] (binary data should remain as 0 and 1). The algorithm will be repeated m times. Start with a p-long weight vector (W) of zeros. At each iteration, take the feature vector (X) belonging to one random instance, and the feature vectors of the instance closest to X (by Euclidean distance) from each class. The closest same-class instance is called 'near-hit', and the closest different-class instance is called 'near-miss'. Update the weight vector such that $W_i = W_i - (x_i - \text{nearHit}_i)^2 + (x_i - \text{nearMiss}_i)^2 W_i = W_i - (x_i - \text{nearHit}_i)^2 + (x_i - \text{nearMiss}_i)^2$ Thus the weight of any given feature decreases if it differs from that feature in nearby instances of the same class more than nearby instances of the other class, and increases in the reverse case. After m iterations, divide each element of the weight vector by m. This becomes the relevance vector. Features are selected if their relevance is greater than a threshold . ReliefF searches for k near misses from each different class and averages their contributions for updating W, weighted with the prior probability of each class.

relieff() will give rank the feature based on their predictive capability from high to low. I will use first 30 top ranked feature among 48. This will lead to new a feature set.

III. Dimentionality Reduction

For dimentionality reduction I will use Principle Component Analysis (PCA). PCA takes orthogonal projection of each data instances into n independent orthogonal axis. Thus leads to n independent principle component.

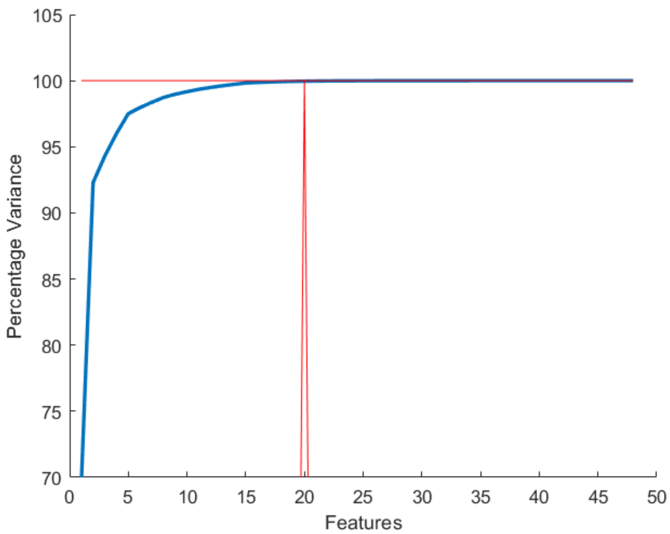


Fig. 6: Sum of percentage variance of features after PCA, we can see that first 20 features carry the 100 variance of the data..

From figure 5 we can see that first 20 principle components carries almost 100% variance of the data. Thus PCA will leads to a new feature set with 20 predictive features. Thus reducing the dimentionality of the data and improving computational speed.

IV. Classification Results

As discussed in section one this sensor-less drive diagnostic is a classification problem. I will evaluate the performance of several popular classification algorithm to classify the type of fault in the machine. To show the performance of the classifier I choose Confusion Matrix. As confusion Matrix are the most compact form of evaluation matrix which shows class-wise accuracy of the classifier. Performance of the classifier will be evaluated on three different feature set as discussed earlier. 1) Feature set 1 - Selected by visual inspection of graphical representation 2) Feature set 2 - First 30 selected from ranked predictive feature given by "relieff()" algorithm 3) Feature set 3 - First 20 principal component which carries almost 100% variance of the data given by "PCA".

A. Classification: Naive Bayes

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. This classification method fits uni-variate gaussian distribution to the individual features of the data and measures.

The naive Bayes classifier combines this model with a decision rule. One common rule is to pick the hypothesis that is most probable; this is known as the maximum a posteriori or MAP decision rule. The corresponding classifier, a Bayes classifier, is the function that assigns a class label $\hat{y} = C_k \hat{y} = C_k$ for some k as follows:

$$\hat{y} = \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k).\hat{y}$$

$$= \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k).$$

B. Classification: Discriminant Analysis (LDA and QDA)

C. Classification: K Nearest Neighbor (Euclidean and Cosine)

D. Classification: Support Vector Machine (Linear and Gaussian)

Since majority of the features are distributed gaussian or mixture of gaussian. Gaussian kernel based SVM performs better than linear kernel based SVM.

E. Classification: Decision Tree

V. Conclusion

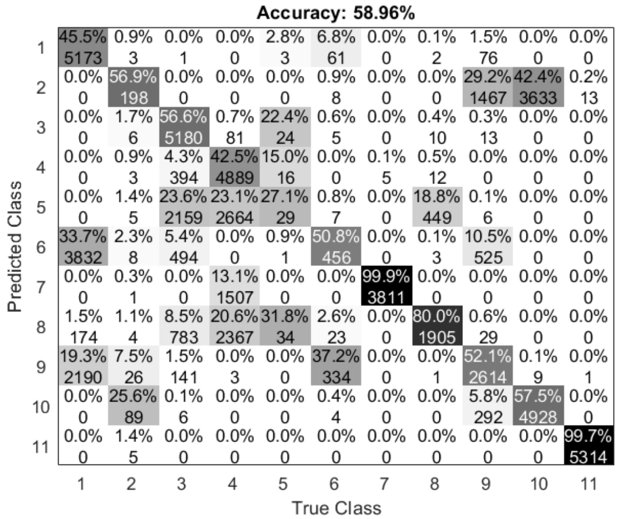
Naive bayes assumes that the features are linearly independent thats why the performance improved after PCA. Over all "RELIEFF" based feature selection improved both the classification performance and computational speed for all algorithms. Gaussian kernel based SVM and Decission Tree algorithm gives the best performance over all in all three feature set.

References

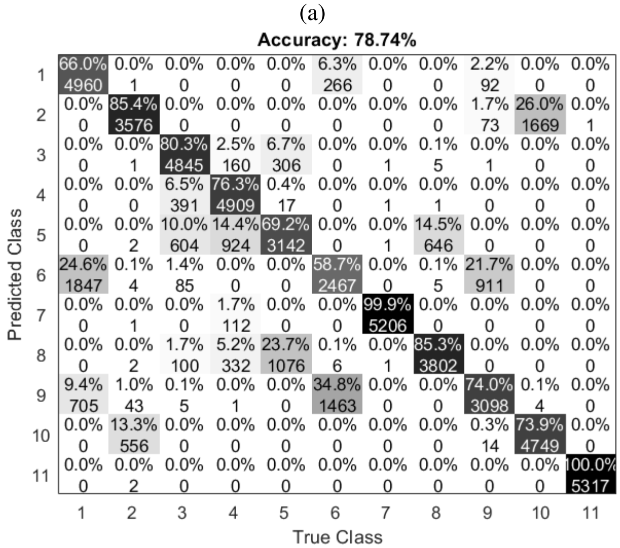
[1] Data Source: Dataset for Sensorless Drive Diagnosis Data Set, UCI Data Repository.

[2] Lohweg, Volker; Paschke, Fabian; Bayer, Christian; Bator, Martyna; Mnks, Uwe; Dicks, Alexander; Enge-Rosenblatt, Olaf, Sensorlose Zustandsberwachung an Synchronmotoren, KIT Scientific Publishing, 2013 (Karlsruher Institut fr Technologie, Institut fr Angewandte Informatik, Automatisierungstechnik. Schriftenreihe 46) ISBN: 978-3-7315-0126-8, S.211-225.

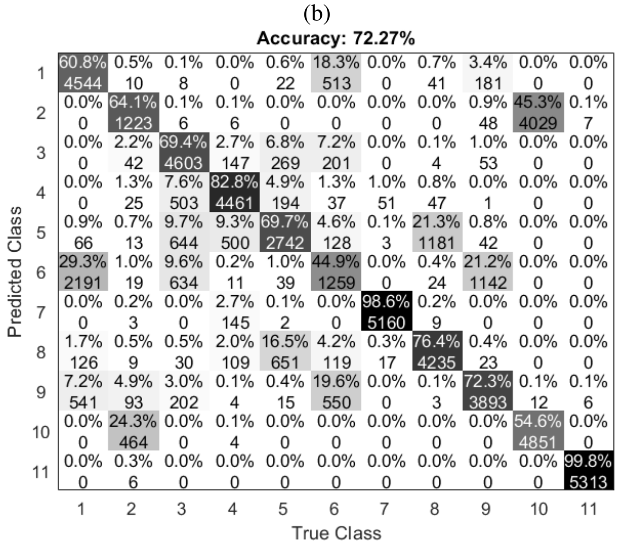
[3] Bayer, Christian Enge-Rosenblatt, Olaf Bator, Martyna Mnks, Uwe & Dicks, Alexander & Lohweg, Volker. (2013). Sensorless drive diagnosis using automated feature extraction, significance ranking and reduction. 1-4. 10.1109/ETFA.2013.6648126.



(a)

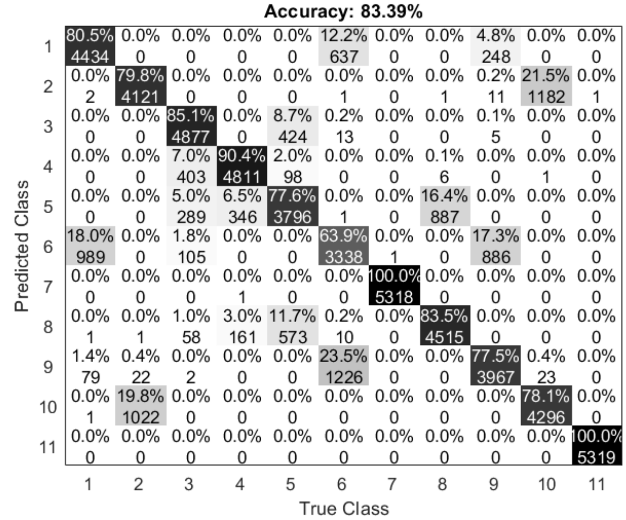


(b)

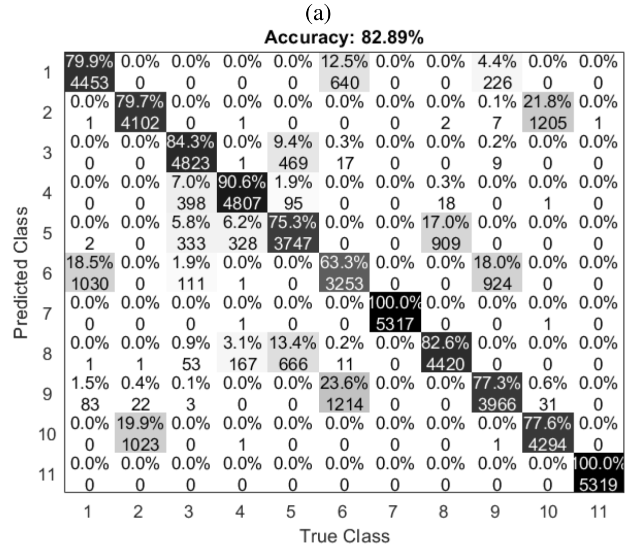


(c)

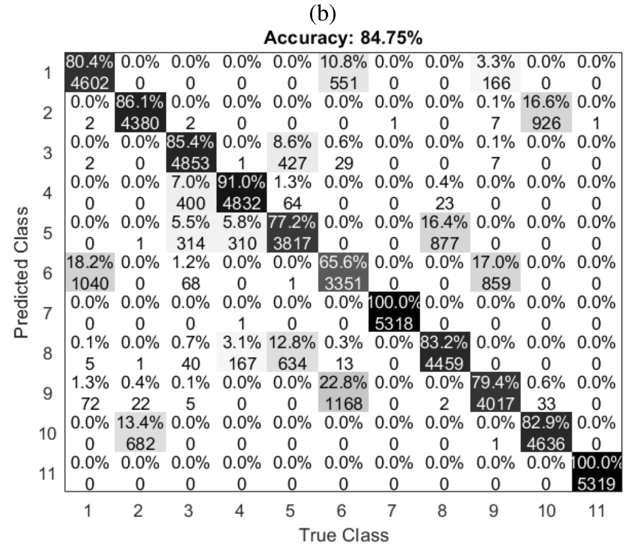
Fig. 7: Confusion Matrix and accuracy of Naive Bayes Classifier on three different feature set a) Graphical b) RELIEFF c) PCA .



(a)

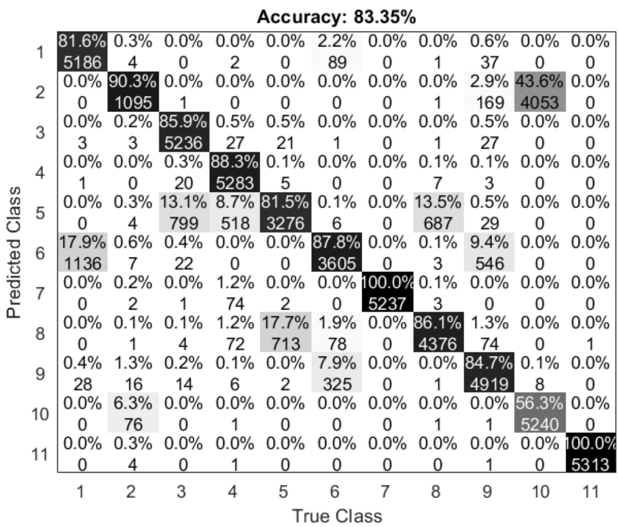


(b)

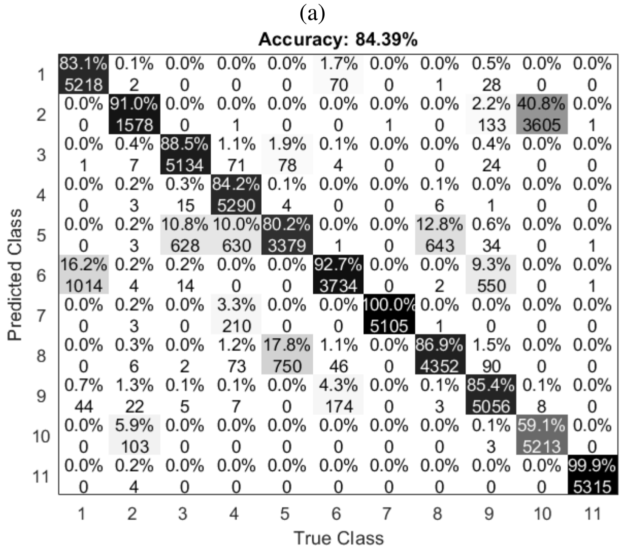


(c)

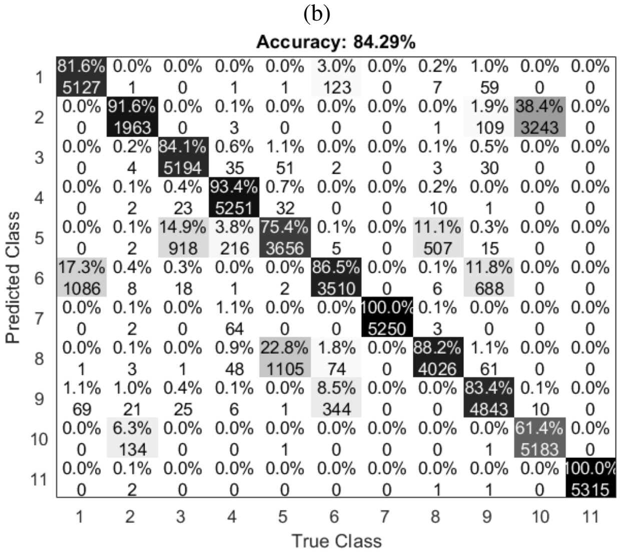
Fig. 8: Confusion Matrix and accuracy of LDA Classifier on three different feature set a) Graphical b) RELIEFF c) PCA .



(a)

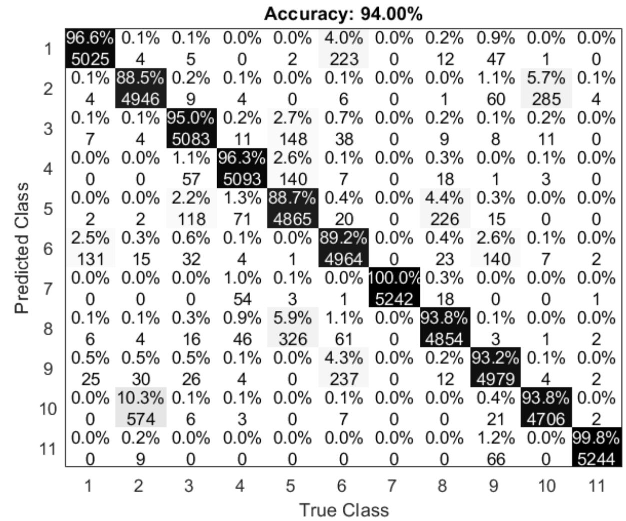


(b)

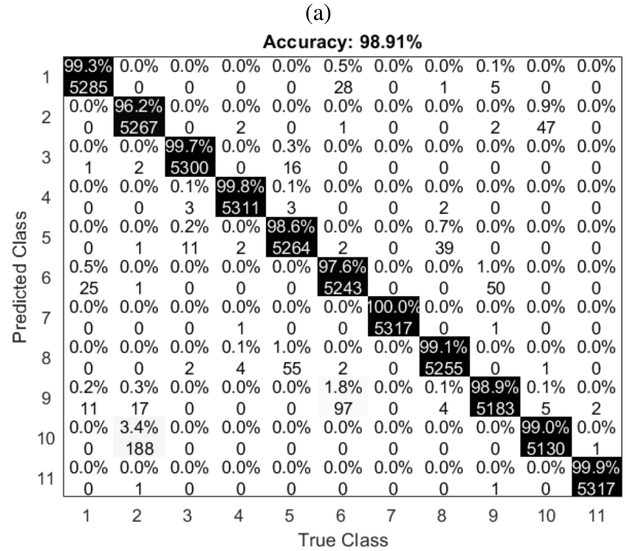


(c)

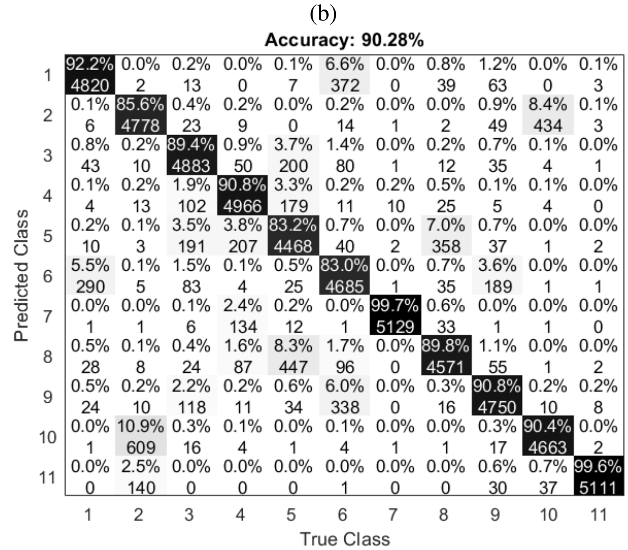
Fig. 9: Confusion Matrix and accuracy of QDA Classifier on three different feature set a) Graphical b) RELIEFF c) PCA.



(a)

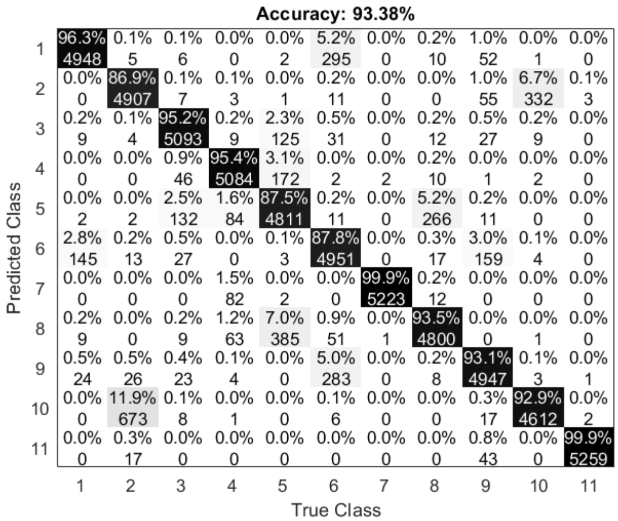


(b)

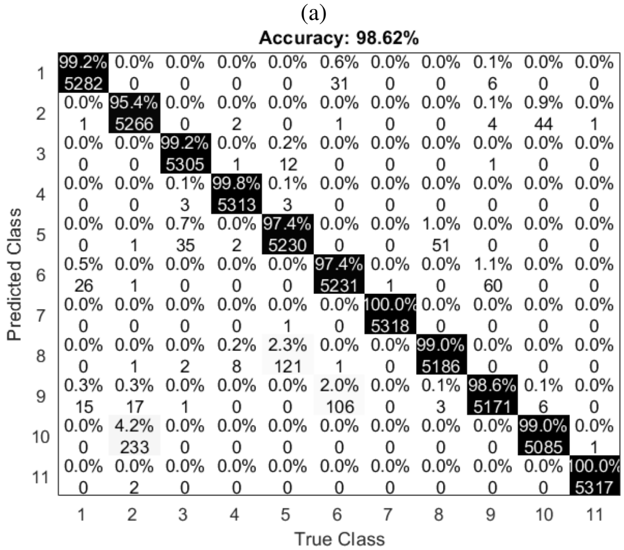


(c)

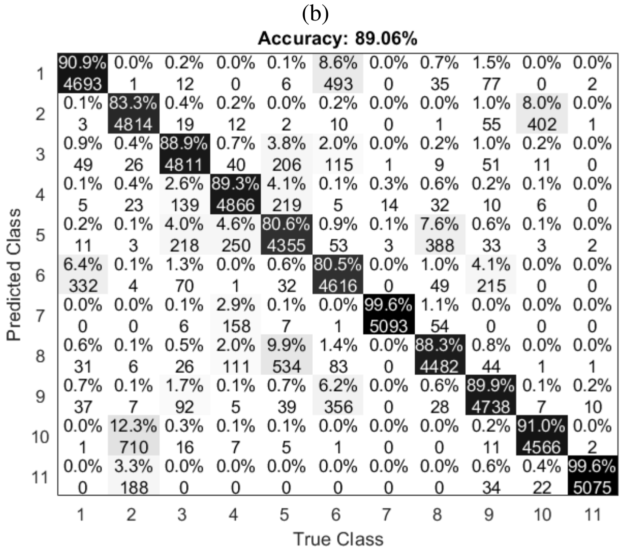
Fig. 10: Confusion Matrix and accuracy of KNN(Euclidean) Classifier on three different feature set a) Graphical b) RELIEFF c) PCA.



(a)

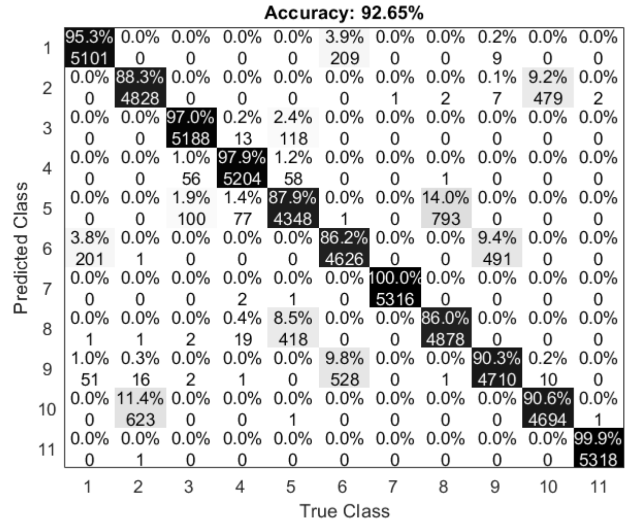


(b)

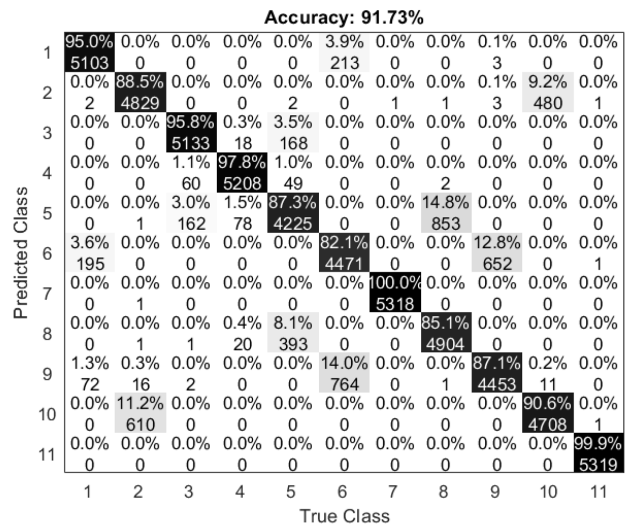


(c)

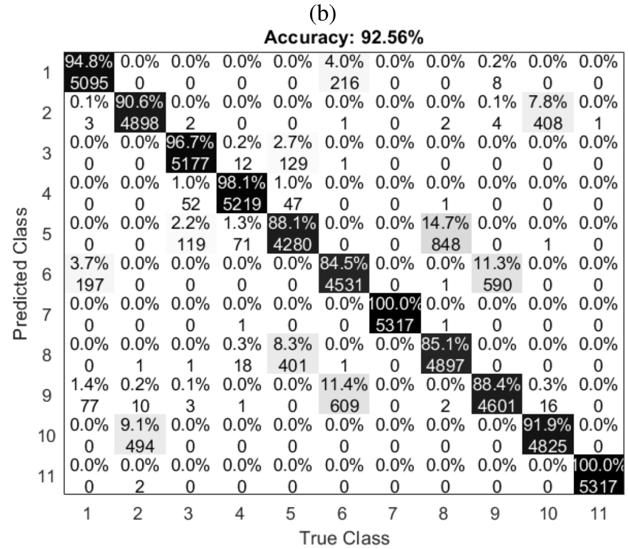
Fig. 11: Confusion Matrix and accuracy of KNN(Cosine) Classifier on three different feature set a) Graphical b) RELIEFF c) PCA.



(a)

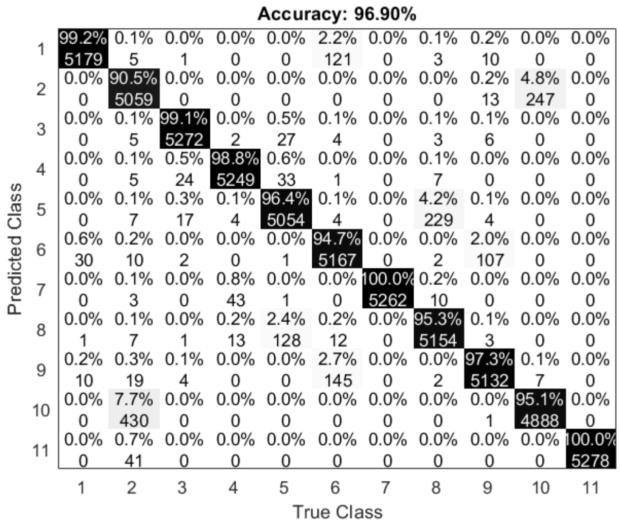


(b)

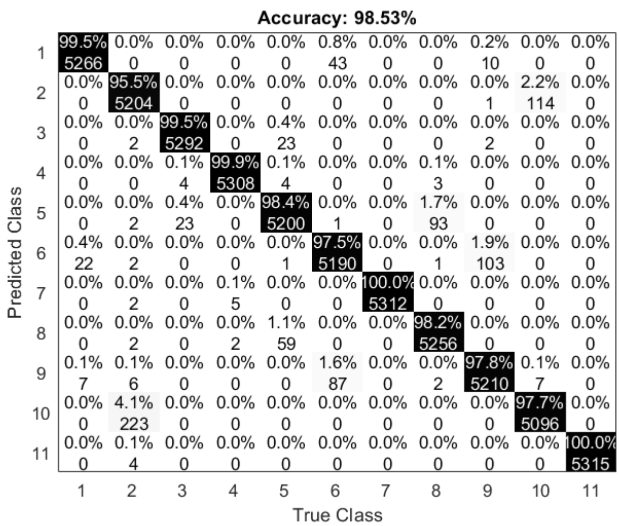


(c)

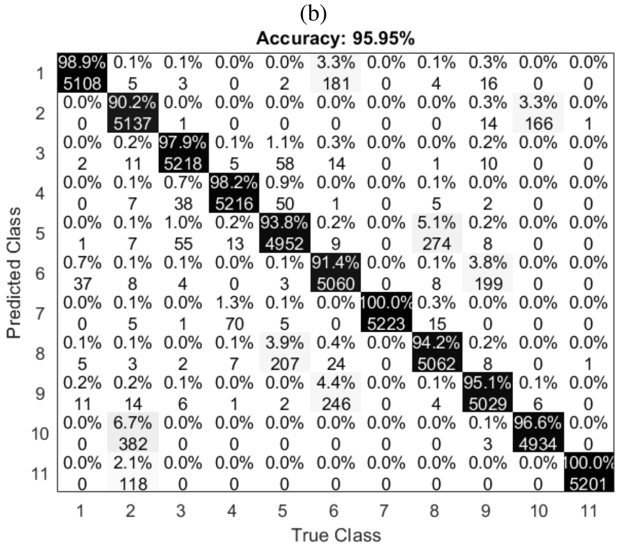
Fig. 12: Confusion Matrix and accuracy of SVM(Linear) Classifier on three different feature set a) Graphical b) RELIEFF c) PCA.



(a)

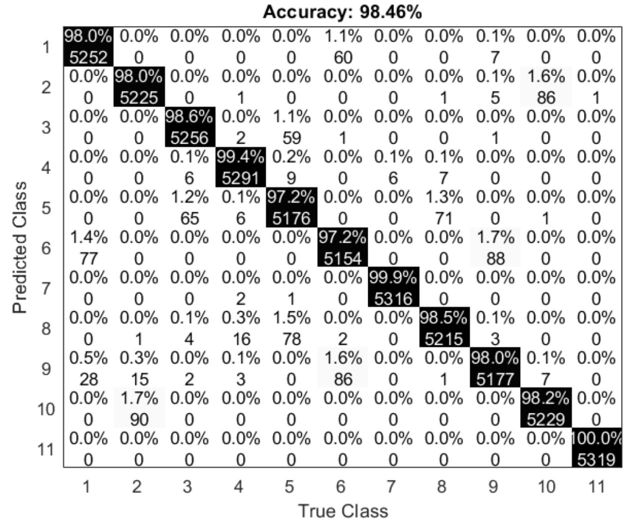


(b)

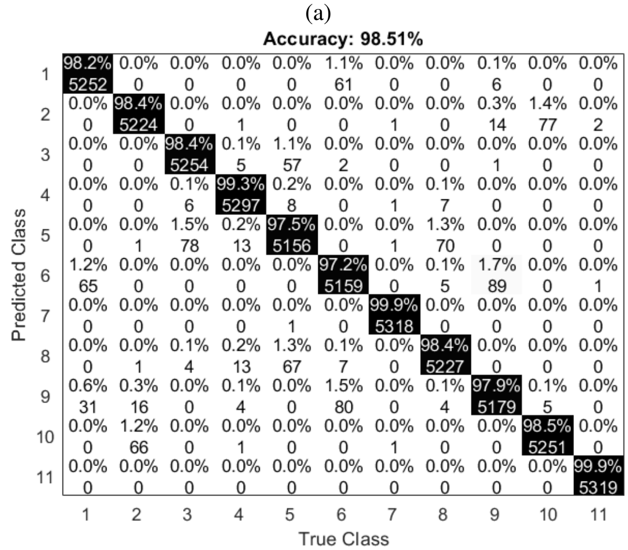


(c)

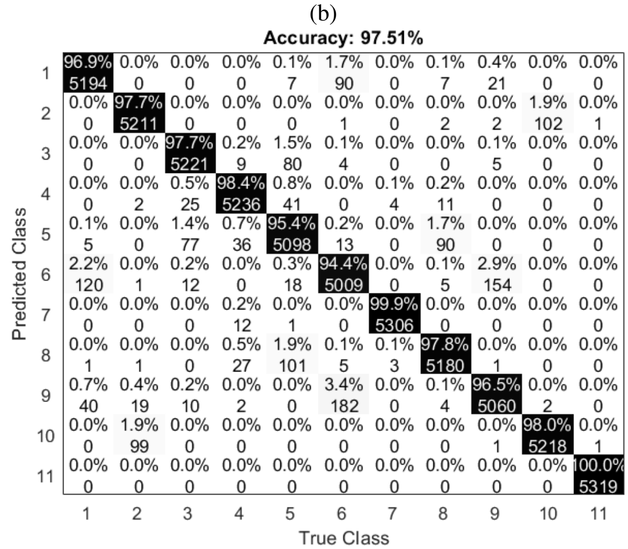
Fig. 13: Confusion Matrix and accuracy of SVM(Gaussian) Classifier on three different feature set a) Graphical b) RELIEFF c) PCA.



(a)



(b)



(c)

Fig. 14: Confusion Matrix and accuracy of Decision Tree Classifier on three different feature set a) Graphical b) RELIEFF c) PCA.