Leif E. Peterson / Int.J.Data.Sci. & Big Data Anal. 4(1) (2024) 34-51
https://doi.org/10.51483/IJDSBDA.4.1.2024.34-51

ISSN: 2710-2599

SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Data Science and Big Data Analytics

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper

Open Access

Improved Dimension Reduction for Pattern Classification Using Noise Eigenspace Projection

Leif E. Peterson[1*]

[1]NXG Logic LLC, Houston, Texas USA 77030. E-mail: peterson.leif.e@gmail.com

Article Info

Volume 4, Issue 1, May 2024
Received : 23 January 2024
Accepted : 18 April 2024
Published : 05 May 2024
doi: 10.51483/IJDSBDA.4.1.2024.34-51

Abstract

A novel subspace method is proposed to improve dimension reduction for pattern classification. As a dimension reduction (DR) technique, information-to-noise estimators (INEs) project feature values on class-specific noise eigenvectors derived from the correlation matrix for each class. Parameter and ablation studies indicate that INEs offer improved stability for classification performance. Class prediction results indicate that use of INEs for 14 datasets resulted in a 15.5% increase in accuracy from 84.5% to 97.6% compared with other DR methods. Use of INEs prior to classification analysis also outperformed typical feature selection prior to classification with an 18.3% improvement in mean accuracy (0.97 vs. 0.82). In addition, the standard deviation of accuracy over all datasets was 0.14-0.16 for other DR methods and 0.078 for INE, suggesting a ~48% decrease. We advocate intraclass noise eigenspace projection for pattern classification as an alternative to input feature selection or DR with other methods prior to classification analysis. In conclusion, use of INEs based on the class-specific noise subspace can greatly improve classification performance and model parsimony without use of input feature selection.

Keywords: *Classification, Correlation, Eigenspace, Feature extraction, Principal components analysis*

1. Introduction

1.1. Basic Model - Interclass PCA

Data analysts often face the "curse of dimensionality," whereby a dataset's number of features greatly exceed the number of objects ($p >> n$). For this reason, unsupervised class discovery of objects commonly involves dimension reduction (DR) to gain insight on patterns or clusters of objects. Interclass principal component analysis (PCA) is one of the most popular classical methods used for reducing the dimensions of a dataset (Pearson, 1901). At the heart of PCA lies eigendecomposition, a linear matrix-based technique for extracting eigenvalues and eigenvectors from a feature-by-feature square non-symmetric or square symmetric matrix,

* *Corresponding author: Leif E. Peterson, NXG Logic LLC, Houston, Texas USA 77030. E-mail: peterson.leif.e@gmail.com*

which is typically based on covariance or correlation (Jolliffe, 1992). A $p \times p$ correlation matrix $\mathbf{R}$ has p eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and accompanying nonzero eigenvectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$. Each λ_j ($j = 1, 2, \ldots, p$) and its corresponding $\mathbf{e}_j$ satisfy

$$\mathbf{R}\mathbf{e}_j = \lambda_j \mathbf{e}_j \quad \ldots(1)$$

By the Principal Axis Theorem, $\mathbf{R}$ can be expressed as

$$\begin{aligned} \mathbf{R} &= \sum_{j=1}^{p} \lambda_j \mathbf{e}_j \mathbf{e}_j^{\top} \\ &= \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^{\top}, \end{aligned} \quad \ldots(2)$$

where $\mathbf{E}$ is the matrix of orthogonal eigenvectors and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues with zero off-diagonals. In expanded form, we get

$$\mathbf{R} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^{\top} = \begin{bmatrix} e_{11} & \cdots & e_{1p} \\ e_{21} & \cdots & e_{2p} \\ \vdots & \ddots & \vdots \\ e_{p1} & \cdots & e_{pp} \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{bmatrix} \begin{bmatrix} e_{11} & \cdots & e_{1p} \\ e_{21} & \cdots & e_{2p} \\ \vdots & \ddots & \vdots \\ e_{p1} & \cdots & e_{pp} \end{bmatrix}^{\top} \quad \ldots(3)$$

Some useful properties of eigenvalues from $\mathbf{R}$ are as follows:

1. The sum of eigenvalues equals the number of dimensions of $\mathbf{R}$: $p = \Sigma\lambda_j$;
2. The cumulative variance of $\mathbf{X}$ explained by the kth eigenvalue of $\mathbf{R}$ when $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq \ldots \geq \lambda_p$ is $Var(\mathbf{X})_k = \left(\sum_j^k \lambda_j\right) \Big/ p$;
3. The determinant $|\mathbf{R}|$ is $\Pi\lambda_j$;
4. The inverse $\mathbf{R}^{-1}$ is $\mathbf{E}\boldsymbol{\Lambda}^{-1}\mathbf{E}^{\top}$, where $\boldsymbol{\Lambda}^{-1} = \text{diag}\{1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_p\}$;
5. The square root matrix $\mathbf{R}^{1/2}$ is $\mathbf{E}\boldsymbol{\Lambda}^{1/2}\mathbf{E}^{\top}$, where $\boldsymbol{\Lambda}^{1/2} = \text{diag}\left\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_p}\right\}$.

Properties 1 and 2 listed above only apply to a correlation matrix, and not the covariance matrix or other positive definite square symmetric matrices, while properties 3-5 apply to all square symmetric matrices.

Using PCA for DR requires estimation of the principal component (PC) scores. The non-standardized PC scores for $\mathbf{R}$ are based on the relationship

$$\begin{aligned} \hat{f}_{ik} &= \left(\frac{x_{i1} - \bar{x}_1}{s_1}\right)\hat{e}_{k1} + \cdots + \left(\frac{x_{ip} - \bar{x}_p}{s_p}\right)\hat{e}_{kp} \\ &= z_{i1}\hat{e}_{k1} + z_{i2}\hat{e}_{k2} + \cdots + z_{ip}\hat{e}_{kp} \\ &= \mathbf{z}_i^{\top}\hat{\mathbf{e}}_k, \end{aligned} \quad \ldots(4)$$

where i represents the ith object ($i = 1, 2, \ldots, n$). In matrix notation, the matrix of non-standardized PC scores is

$$\underset{n\times p}{\hat{\mathbf{F}}} = \underset{n\times pp\times p'}{\mathbf{Z}}\ \underset{}{\hat{\mathbf{E}}} \quad \ldots(5)$$

and the standardized PC scores are

$$\underset{n\times p}{\hat{\mathbf{F}}} = \underset{n\times pp\times pp\times p}{\mathbf{Z}}\ \underset{}{\hat{\mathbf{E}}}\ \underset{}{\hat{\boldsymbol{\Lambda}}}^{-1/2} \quad \ldots(6)$$

Thus, for an input $n \times p$ data matrix $\mathbf{X}$, there will be an $n \times p$ matrix $\widehat{\mathbf{F}}$. For each object there will be p PC scores in each row of $\widehat{\mathbf{F}}$, and in each column of $\widehat{\mathbf{F}}$, there will be n PC scores for all of the objects. Some important properties of PC score column vectors is that they are orthogonal with zero correlation

$$\mathbf{f}_1 \perp \mathbf{f}_2 \perp \cdots \perp \mathbf{f}_p \qquad ...(7)$$

and are distributed standard normal, $\mathcal{N}(0,1)$.

Figure 1 shows 3D plots of non-standardized PC scores from (5) for three sets of 6 features with various levels of correlation and the corresponding heat maps representing $\mathbf{R}$. When all 6 features are highly correlated, the major eigenvalue of $\mathbf{R}$ is $\lambda_1 = 6$ and the 1st PC score vector forms a line in 3D space, called content. When there are two sets of 3 highly correlated features, the eigenvalues of $\mathbf{R}$ are $\lambda_1 = 3.05$ and $\lambda_2 = 2.95$, and the two PC vectors form an ellipse or disc in 3D space. For three sets of 2 highly correlated features, the eigenvalues of $\mathbf{R}$ are $\lambda_1 = 2.09$, $\lambda_2 = 2$, and $\lambda_3 = 1.9$, and the 3 PC vectors form a sphere. With near-zero correlation in the case of random noise, all eigenvalues will be unity when $n \to \infty$. Altogether, it is clear that eigenvalues represent the number of sets of features which are highly correlated together.

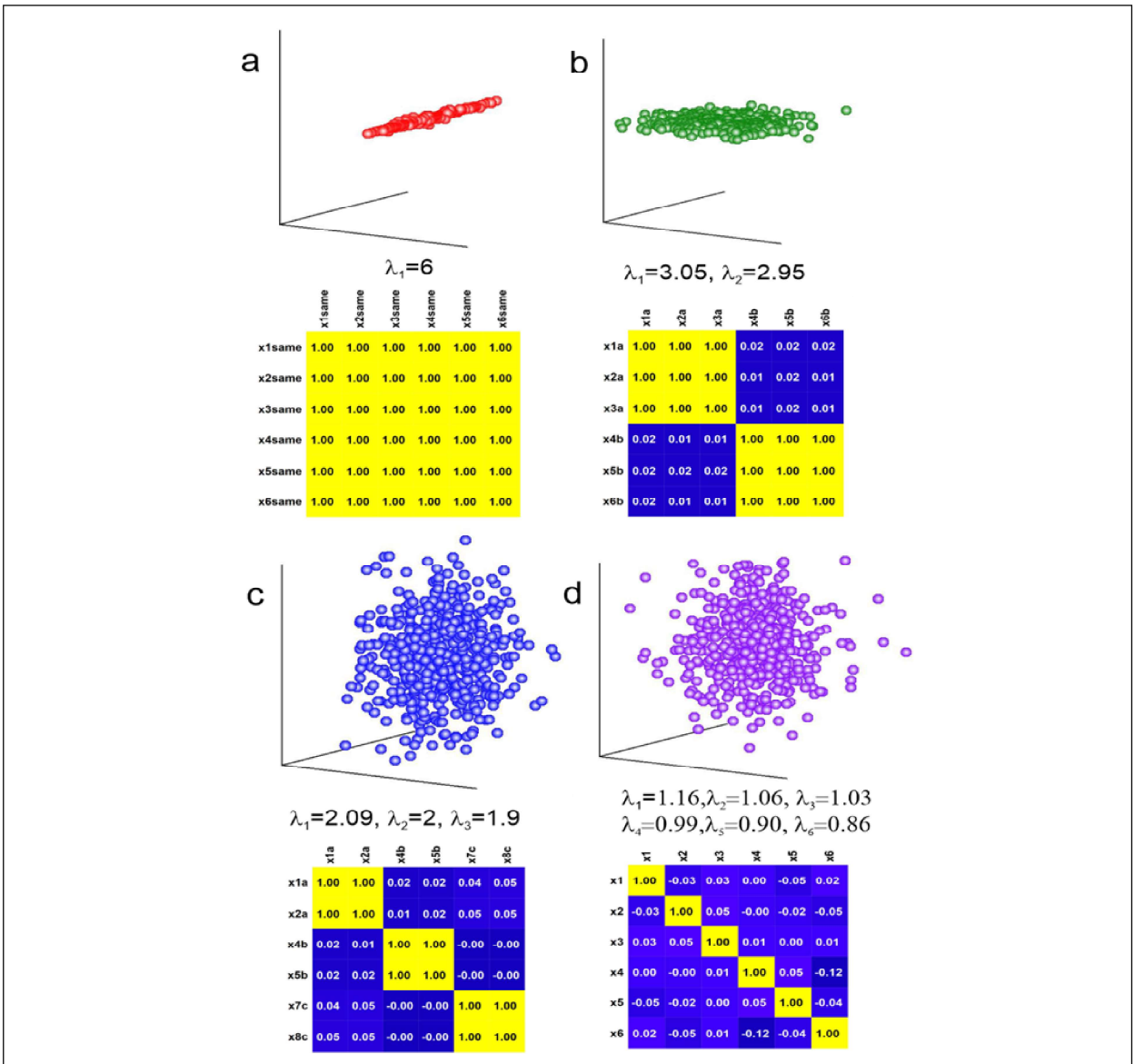

Figure 1: 3D Plots of Non-Standardized PC Scores from (5) and Corresponding Correlation Matrix Heat Maps for Three Sets of 6 Features

Note: (a) One Set of 6 Highly Correlated Features, (b) two sets of 3 highly correlated features, (c) three sets of 2 highly correlated features, and (d) 6 features having random $\mathcal{N}(0,1)$ distributions.

1.2. Advantages of Interclass PCA

Dimension reduction is known to improve model parsimony, which is sought when attempting to attain similar results through the use of fewer dimensions. In addition to DR, PCA also performs decorrelation and noise reduction. In the presence of correlated features, decorrelation can be an important pre-filtering step before clamping features to a feed-forward back-propagation artificial neural network (ANN), because ANNs can "waste time" learning about between-feature correlation (Peterson and Coleman, 2006). Overfitting is another source of bias related to use of an overly complex model that learns too much about the training data, especially the noise (Geman *et al.*, 1992). For this reason, PCA is typically employed as a feature transformation step in workflows involving high-dimensional class prediction.

1.3. Disadvantages of Interclass PCA

Although PCA is useful for DR, decorrelation, and noise removal, there are several disadvantages. Firstly, PCA is considered a linear operation, mostly because of the nature of the matrix operations. Since PC scores are distributed $\mathcal{N}(0,1)$, they remove skewness and kurtosis among feature distributions, however, at the cost of overlaying a normality assumption on the problem being tackled. Many man-made and naturally-occurring feature distributions are not Gaussian or skew-normal, but rather can be triangular, Levy-stable, or multimodal, non-linear, and discontinuous in the limit. Eigendecomposition collapses a larger number of correlated features down to a reasonable set of uncorrelated dimensions that describe a majority of the variance in **X**. The signal space of **R** only represents correlation or multicollinearity, so if multicollinearity is invariant across class labels, the interclass signal will not vary, rendering interclass PCA and PC score generation noninformative for pattern classification.

1.4. Intraclass PCA

When there are class differences between feature correlation, intraclass PCA can be employed to generate PC scores that are specific to each class, as a means of augmenting informativeness of class labels. The choice of using the signal or noise subspace of **R** depends on the transform applied to the eigenvectors. If the signal is the target, then signal eigenvectors are used for generating PC scores, whereas when the noise subspace is of interest, the inverse of noise eigenvector values can be exploited. In the noise space, the length of eigenvectors, $\sqrt{\lambda_j}$, is smaller and therefore their inverse is larger—which can boost the sensitivity for detecting class differences. In terms of angular rotation, the noise subspace often spans a larger volume since there are typically more eigenvalues (eigenvectors) in the noise space when compared to the signal space. Unfortunately, the awareness and popularity of intraclass PCA for noise subspace methods is swamped by the popularity of interclass PCA, mostly because data analysts and statisticians are not formally trained on noise eigenspace methods.

1.5. Related Work

A review of the state-of-the-art in pattern classification and signal-to-noise ratios (SNRs) indicates that the majority of previous work has been done in the time-frequency domain. For example, Cabrera *et al.* (2018) studied classification of electroencephalograms (EEG) in the high-beta (19-30 Hz) and gamma (30-45 Hz) brainwave bands. Signal demodulation effects on classification have also been investigated using deep learning under varying noise conditions (An and Lee, 2023). Alarabi and Alkishriwo (2021) employed a neural network for classifying several digital modulation schemes such as BPSK, QPSK, 16QAM and 64QAM. Radiofrequency fingerprinting is another area of research where effects of receiver SNR on classification results have been investigated (Rehman *et al.*, 2012). Sahu and Mishra (2011) and Mishra and Sahu (2011) focused on SNRs for informative gene selection and gene expression classification; however, the SNR was merely defined as $(\mu_1 - \mu_2)/(\sigma_1 + \sigma_2)$, which closely resembles a t-statistic but does not involve noise subspace methods based on eigendecomposition.

Signal classification based on noise subspace methods have focused on speech recognition (Tan *et al.*, 2007), hyperspectral images for remote sensing (Tu *et al.*, 1998), and vector classification using signal subspace matching (Wax and Adler, 2023). Various matrix whitening approaches have been developed, including the mutual subspace method (Maeda and Watanabe, 1985) and variants such as the multiple pseudo-whitened mutual subspace method (Yamaguchi and Fukui, 2022). These methods employ the (Jiang *et al.*, 2008) method

for eigendecomposition of the dispersion matrix into a reliable, noise, and null subspace via whitening, and simultaneously extract informative features from high-dimensional data vectors with hundreds to thousands of features—which are typical for image processing.

Subspace methods have also been developed for spectral density estimation in the time-frequency domain. Use of the noise eigenspace of a covariance matrix was first introduced by Pisarenko for frequency estimation during harmonic decomposition of a noisy signal during earthquake identification (Pisarenko, 1973). The peak location in the frequency estimation function (pseudo-spectrum) of the Pisarenko Harmonic Decomposition (PHD) was functionally composed as

$$\hat{P}_{PHD}\left(e^{j\omega}\right) = \frac{1}{\left|\mathbf{e}^T \mathbf{V}_{\min}\right|^2}, \quad ...(8)$$

where $\mathbf{V}_{min}$ is the noise eigenvector and $e = [1 \quad e^{j\omega} \quad e^{j2\omega} \quad \cdots \quad e^{j(M-1)\omega}]^T$. A limitation of PHD is that it is one-dimensional, since only a single noise eigenvector is used. Later, Schmidt (1986) introduced a multisignal classification (M.U.S.I.C) method, which was based on the full matrix of noise eigenvectors when there are p signal eigenvectors, given as

$$\hat{P}_{MU}\left(\theta\right) = \frac{1}{\mathbf{a}\left(\theta\right)^T \mathbf{E}_n \mathbf{E}_n^T \mathbf{a}\left(\theta\right)}, \quad ...(9)$$

where $\mathbf{a}(\theta)$ is a steering vector and $\mathbf{E}_n$ is the matrix of noise eigenvectors from the covariance matrix. The introduction of the full matrix of noise eigenvectors, $\mathbf{E}_n$, for peak detection enabled the use of "root" MUSIC for processing signal data from an array of sensors to determine parameters of multiple wavefronts arriving at an antenna array from measurements made on the signals received at the array elements. Root MUSIC has been shown to increase radar resolution to identify the number of incident wavefronts present, directions of arrival or emitter locations, strength and cross correlations among the incident waveforms, and noise-interference strength. Root MUSIC, otherwise known as "beam forming", has been widely adapted in civilian and military applications involving radar detection of aircraft, and its performance is continually being improved (Kim *et al.*, 2002; Mohammed and Noori, 2014; Ozcetin, 2018). What is clear from most studies cited above is that previous research has eluded systematic investigation of noise eigenspace methods for general pattern recognition involving low-dimensional datasets.

The structure of the remainder of this paper is to introduce the datasets, provide a brief description of DR methods employed prior to classification runs, introduce information-to-noise estimators (INEs), and then describe the classification methods employed. Results are then compared between INEs and the various DR methods considered.

The contributions of this paper are:

- To introduce an information-to-noise estimator (INE) that exploits the noise eigenspace from intraclass eigendecomposition;
- A parameter study to understand the effect of using an arbitrary number of dimensions, as well as various dimension(feature) selection methods on class prediction accuracy;
- Ablation studies to determine the effects of annihilation of dimension values via random shuffling and Gaussian smearing on class prediction accuracy; and
- Compare class prediction results between PCA and other DR techniques with results based on INEs.

2. Methods

2.1. Datasets

The majority of datasets were obtained from the Univ. of California-Irvine-Machine Learning Repository (Kelly *et al.*, 2024). The optchar dataset is based on the first 500 records of the test set for the MNIST optical character recognition data base (Deng, 2012). Table 1 lists the datasets and their number of classes, objects, features, and signal and noise eigenvectors identified using interclass PCA. Features with missing values

Table 1: Datasets Employed for DR and Classification Analysis. #Signal λ and #Noise λ Represent the Number of Signal and Noise Eigenvectors Identified in the Feature Correlation Matrix

Dataset	Classes	Objects	Features	#Signal λ	#Noise λ
	Ω	n	p	$(p - m_{PC})$	m_{PC}
Breastca	2	277	6	1	8
Cancer	2	699	9	1	8
Dermatology	6	366	34	4	30
Glass	6	214	9	2	7
Heart Disease	5	297	13	4	9
Hepatitis	2	80	19	1	18
Ionosphere	2	351	32	5	27
Iris	3	150	4	1	3
Liver	2	345	6	1	5
Optchar	10	500	56	9	47
Pima	2	768	8	2	6
Soybean	15	266	35	4	31
Thyroid Recur	2	383	16	5	11
Wine	3	178	13	2	11

were dropped from the original datasets, since imputation of missing data was not a focus of this investigation. In the MNIST dataset, we also dropped features that were invariant.

2.2. Dimension Reduction (DR) Methods

2.2.1. DR Based on Interclass PCA

All computational procedures were performed using methods in the Explorer CE package (NXG Logic LLC,

Table 2: Dimension Reduction Methods Employed Prior to Class Prediction

Method	Acronym	Ref.
All PC scores from PCA $(\lambda_1, \lambda_2, \ldots, \lambda_p)$	PCA	Pearson (1901)
Signal PC scores from PCA	PCASIG	Pearson (1901)
Crisp K-means cluster	CKM	Lloyd (1957)
Unsupervised particle swarm optimization	UPSO	Kennedy and Eberhart (1995)
Kernel Gaussian radial basis function PCA	KGPCA	Schölkopf and Smola (1998)
Diffusion maps	DM	Lafon and Lee (2006)
Unsupervised artificial neural networks	UANN	Hornik *et al.* (1989)
Laplacian eigenmaps	LEM	Belkin and Niyogi (2004)
Sammon mapping	SAMM	Sammon (1969)
Non-negative matrix factorization	NMF	Lee and Seung (1999)
Classic multidimensional scaling	CMDS	Torgerson (1952)
Typical selection of normalized input features	NORM	
Information-to-noise Estimators	INE	(this study, Eq. 12)

2024). The number of signal eigenvectors of **R** for an entire dataset with n objects and p features was based on the Marcenko-Pastur Law (Marcenko and Pastur, 1967), which states that eigenvectors in the signal space will be associated with eigenvalues λ_j above a cutoff value of

$$\lambda^+ = \left(1 + \sqrt{\frac{p}{n}}\right)^2, \qquad ...(10)$$

that is, $\lambda_j > \lambda^+$. Analogously, all noise eigenvectors are associated with eigenvalues $\lambda_j \leq \lambda^+$. The number of eigenvalues in the noise space of each dataset was determined with the relationship

$$m_{PC} = \sum_{j=1}^{p} I(\lambda_j \leq \lambda^+), \qquad ...(11)$$

and the number of signal eigenvalues was then equal to $p - m_{PC}$. The far-right column in Table 1, "Noise λ," reveals the number m_{PC} of noise eigenvectors found in each dataset. The result of calculations described above was an $n \times (p - m_{PC})$ **F** matrix of non-standardized signal PC scores from (5) for each dataset. Two approaches were employed when using interclass PCA before classification analysis: "PCA" and "PCASIG." The PCA method employed selection of PCs (dimensions) using the OAA-MW method (see Section 2.6.2), whereas the PCASIG method employed only the $p - m_{PC}$ signal PCs prior to classification.

2.2.2. DR Based on Intraclass Eigenanalysis - Proposed Information-to-Noise Estimators (INEs)

Although beam forming is principally used to find the root of a polynomial in the time-frequency domain, exploiting the noise eigenspace can yield remarkable insights into pattern classification. By analagous borrowing from denominator terms in (8) and (9), we introduce an information-to-noise (INE) estimator for each noise eigenvector in the form

$$\hat{P}_{INE}(\mathbf{x}_i) = \hat{P}_{ik} = \sum_{j=1}^{p} \frac{x_{ij}}{\left|e_{kj}^{\omega}\right|} \quad \mathbf{x}_i \in \omega, \qquad ...(12)$$

where x_{ij} is the normalized feature value for the ith object in class ω ($\omega = 1, 2, \ldots, \Omega$), and $|e_{kj}^{\omega}|$ is the absolute value of the jth element ($j = 1, 2, \ldots, p$) of the kth noise eigenvector ($k = 1, 2, \ldots, m_{\omega}$) when eigenanalysis is run on the feature correlation matrix from objects in class ω. To guard against inflation of the estimator from miniscule eigenvector values, the estimator is rescaled using

$$\hat{P}_{ik} = \log_{10}(\hat{P}_{ik} + 1) \qquad ...(13)$$

The full size matrix **P** of INEs with dimensions $n \times m$ derived from the Ω classes is then used to replace the data matrix **X** used in classification analysis. INEs provide a novel method for DR of the data matrix prior to class prediction.

For each dataset, we first dropped features having invariant within-class values, since this would result in $\sigma_j = 0$, leading to null correlation coefficients with other features and a singular **R** for the given class. Next, we calculated the eigenvalue noise cutoff for class ω by using

$$\lambda_{\omega}^+ = \left(1 + \sqrt{\frac{p}{n_{\omega}}}\right)^2, \qquad ...(14)$$

and then determined the number of noise eigenvalues in class ω using

$$m_{\omega} = \sum_{j=1}^{p} I(\lambda_{j\omega} \leq \lambda_{\omega}^+), \qquad ...(15)$$

where $\lambda_{j\omega}$ is an eigenvalue from eigendecomposition of the correlation matrix based only on the n_{ω} objects in class ω. We then used (12), which resulted in an $n_{\omega} \times m_{\omega}$ matrix $\mathbf{P}_{\omega}$ of class-specific INEs. The minimum number of noise eigenvectors over all Ω classes was then determined using

$$m_{INE} = argmin_{\omega \in \Omega} m_{\omega}. \quad ...(16)$$

The class-specific matrices, $\mathbf{P}_{\omega}$, were then stacked to create the full $n \times m_{INE}$ $\mathbf{P}$ matrix of INEs for the entire dataset.

2.2.3. DR Methods Based on Manifold Learning

DR can also be performed using manifold learning methods. Manifold learning attempts to retrieve the original structure of a high-dimensional dataset and can identify a surprisingly small number of dimensions to describe characteristics of the original data structure. A manifold is defined as the image of a low-dimensional space that is linearly or non-linearly embedded in a high-dimensional domain. PCA is a linear manifold learning method, since it maps a high-dimensional set of features to a low-dimensional space using the minimal number of orthogonal dimensions required to describe the majority of the variance in the high-dimensional space. Non-linear manifold learning (NLML) can tackle problems when the manifold is not linearly embedded in the original domain. Data integration of high-dimensional datasets can be accelerated by using NLML, since patterns in the original data are compacted and preserved so that they can be exploited for association mapping between datasets.

Recent implementations of NLML include kernel PCA (Schölkopf and Smola, 1998), diffusion maps (Lafon and Lee, 2006), and Laplacian eigenmaps (Belkin and Niyogi, 2004). The basic assumption of these algorithms is that the high-dimensional data lies on a low-dimensional manifold embedded in a high-dimensional structure. Table 2 lists the DR methods employed prior to class prediction analysis.

2.3. Classifiers

The following classifiers were employed (Peterson, 2013): K-nearest neighbor (KNN), naive Bayes classifier (NBC), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), linear regression (LREG), polytomous logistic regression (PLOG), learning vector quantization (LVQ1), kernel regression (KREG), supervised neural gas (SNG), supervised artificial neural network (SANN), gradient ascent support vector machines (SVM), and constricted particle swarm optimization (SPSO). For KNN, k was set equal to the number of classes and was padded with 1 if k was even, in order to prevent ties. A Gaussian RBF was used for the kernel trick in KREG and SVM. For the SANN, one hidden layer was used with the number of input and hidden nodes set equal to the number of dimensions used. Batch learning, 500 epochs, and the tanh activation function was used at hidden nodes, and softmax was used on the output-side. Network error was based on MSE and the number of output nodes for the SANN was equal to the number of classes. A grid search was used to determine the best learning rate for the SANN, and momentum was set to 0.7. Weight decay was not used. A grid search was also employed for the SVM classifier, to determine the best estimates for C and γ. The SPSO classifier used 30 particles, and 150 generations. The number of training iterations for LVQ1 and SNG was set to 50. Ensemble classifier fusion was also performed based on majority votes (EMV) and weighted majority votes (EWMV).

2.4. Cross-Validation

10-fold cross-validation was used during class prediction whereby objects were randomly assigned to folds $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_{10}$. During the first fold, training data were based on folds $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_9$, while objects in fold $\mathcal{D}_{10}$ were used for testing. This was repeated until class prediction was performed for objects in each of the test folds. Classification accuracy was based on dividing the total number of objects in the diagonal of the confusion matrix, for which rows represented true class and columns represented predicted class, by the total number of test objects, which was always equal to n. Feature normalization was used during all 10-fold CV classification runs, where min(x) and max(x) from feature values of training objects were used for transforming feature values of test objects. Test feature values below min(x) from training objects were set to zero, while test feature values greater than max(x) were set to unity.

2.5. Classification Runs - All Classifiers vs. Diverse Classifiers

We ran two types of classification runs: one using only two DR methods (PCA and INE) and all classifiers, and the other using all DR methods and only the most diverse classifiers (NBC, LVQ1, SNG). The first method enables evaluation of INEs for their stability over datasets. The second method enables evaluation of INEs for

their stability over DR methods. In another recently completed investigation based on similar datasets, we observed that the NBC, LVQ1, and SNG classifiers were more diverse (Kuncheva and Whitaker, 2003; Ivascu *et al.*, 2022) than the other classifiers employed, in that they maximized the variance over datasets and minimized the generalization error (results not shown). Therefore, we elected to report only EWMV accuracy from these 3 classifiers for runs based on all DR methods, since the goal of this study was to investigate effects of DR on classification, and not classifier diversity.

2.6. Parameter Studies

2.6.1. Arbitrary Number of Dimensions

A common guideline when using PCA for dimension/noise reduction is to extract PC vectors for dimensions whose eigenvalues $\lambda_j > 1$, which implies use of information from features that are correlated with other features and are not unique. Such dimensions also contribute more to the variance explanation of the data matrix **X**. While these assumptions may be helpful for PCA, they provide no basis for selecting informative dimensions derived from the other DR techniques used in this investigation (Table 2).

2.6.2. Filtering for Significant Dimensions

We also explored the effects of filtering dimensions(features) that were significant predictors of class labels prior to input into classification analysis. Nine dimension selection methods were used. The first four filtering methods were based on one-against-all class comparisions (OAA) using the T-test, Mann-Whitney (MW) test, information gain, and Gini index (Peterson *et al.*, 2006; Peterson, 2013). Significance for the T-test and MW test were based on two-tailed $p < 0.05$. The top N dimensions selected with information bias and Gini index were based on the number of dimensions that were significant when using the MW test. The MW test was used as a basis for information gain and Gini index because it is not susceptible to outliers or skewness. The second set of 4 filtering methods were based on all-possible-pairs (APP) class comparisons for the T-test, MW test, information gain, and Gini index. The last dimension selection method was a greedy forward-reverse plus-take-away 1 method called GreedyPTA. GreedyPTA first invokes Mahalanobis distance to find the feature with the smallest between-class distance, and then iterates using a balance between F-to-enter and F-to-remove criteria to construct a full model with dimensions derived using Wilks' Lambda statistic (Peterson, 2013).

2.7. Ablation Studies

2.7.1. Random Shuffling Within Dimensions

Random shuffling of values within dimension (features) was undertaken to study the effect of annihilation of data informativeness for classification accuracy for the various DR methods and datasets used. Dimensions were filtered by using the OAA-MW method. Random shuffling of values within each dimension first involved setting the random seed to the CPU-clock's current millisecond. Next, within each selected dimension, the ith object's values were shuffled with another randomly selected object if $U(0, 1) < f$, where f was $1/2^k$, and $k = 1, 2, 3, 4, 5, 6$. This represented object random sampling fractions of $f = \{0.5, 0.25, 0.125, 0.0625, 0.03125, 0.015625\}$. When $k = 7$, no shuffling was performed.

2.7.2. Gaussian Smearing

Another ablation study involved adding Gaussian noise in an attempt to smear values within dimensions prior to classification. Dimensions were filtered by using the OAA-MW method. Smearing was then accomplished by adding standard normal $\mathcal{N}(0, \sigma^2)$ variates to each value of the selected dimension, where $\sigma = \{10, 5, 1, 0.5, 0.1, 0.05\}$.

3. Results

3.1. Parameter Studies

3.1.1. Arbitrary Number of Dimensions

Figure 2 shows that, for an arbitrary number of dimensions from four datasets with the greatest number of features, use of INEs as inputs to classification analysis performed better than the other DR techniques. Results also indicate that at approximately 10 dimensions, the majority of DR methods seem to reach the

maximum classification accuracy, hence the rule-of-thumb to use 10 PCs from PCA for input into classifiers does seem to have merit for the datasets considered. An important observation was that the curvilinear change in accuracy as a function of dimensions for INE was more stable when compared with other DR methods.

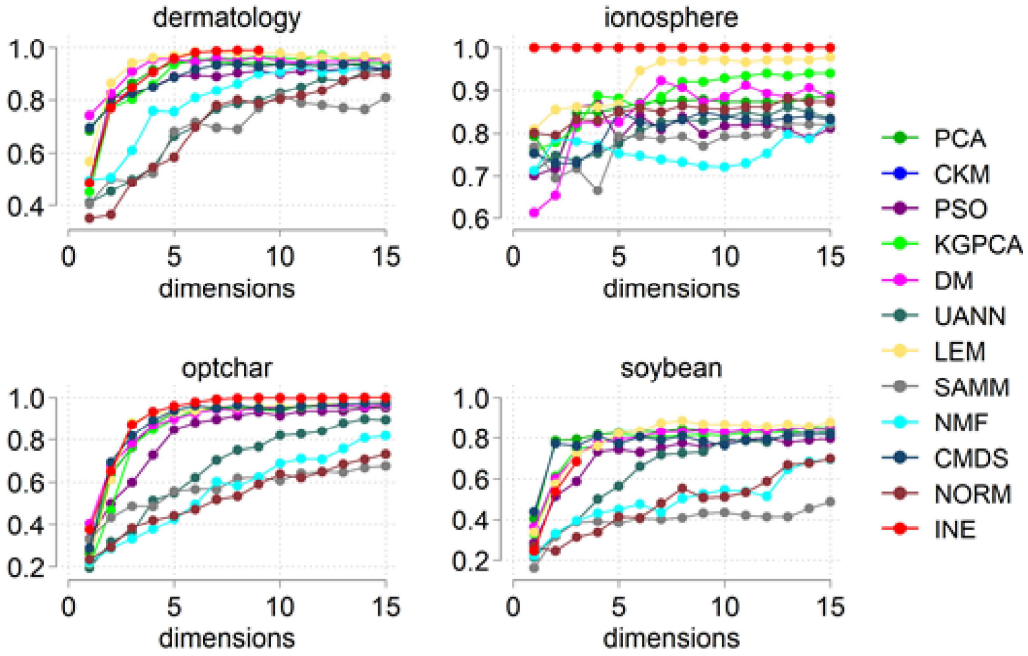

Figure 2: Parameter Study Results: Effect of Arbitrary Number of Dimensions on Classification Accuracy for Various DR Methods. Accuracy is Ensemble Weighted Majority Voting, EWMV, from 3 Classifiers (NBC, LVQ1, SNG)

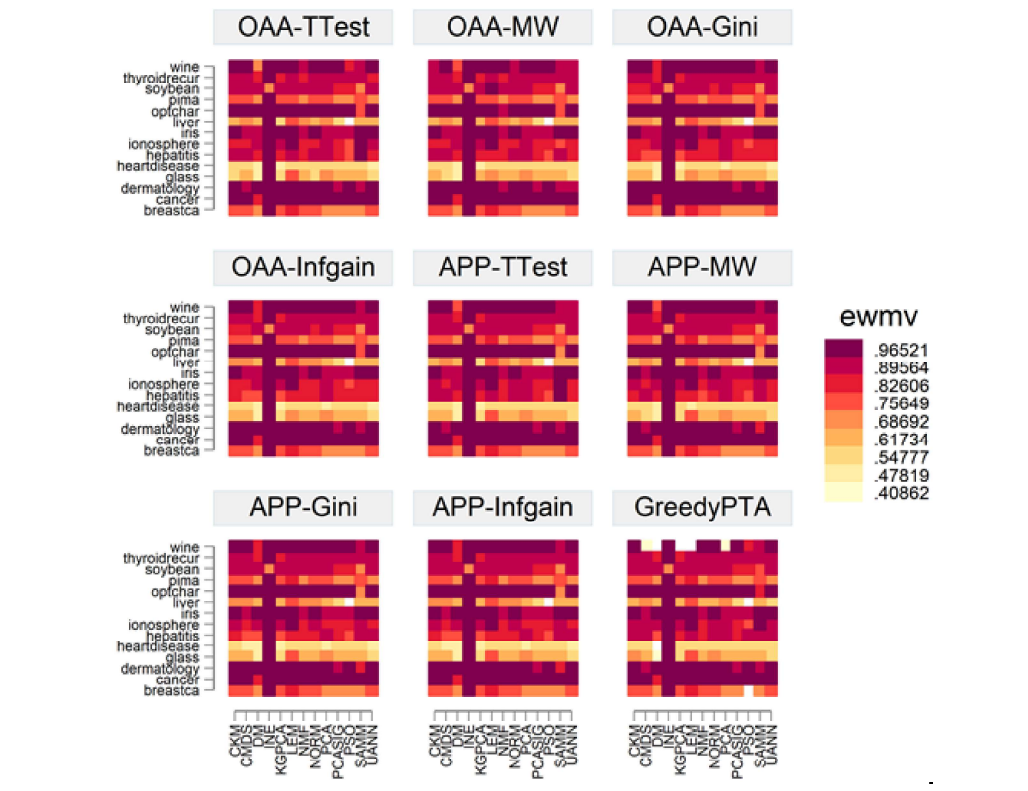

Figure 3: Parameter Study Results: Effect of Filtering Dimensions with Various Dimension Selection Methods on Classification Accuracy for Various DR Methods. Accuracy is Ensemble Weighted Majority Voting, EWMV, from 3 Classifiers (NBC, LVQ1, SNG)

3.1.2. Filtering for Significant Dimensions

Figure 3 indicates that results of using various dimension selection techniques was quite similar, and use of INEs as inputs to classification analysis resulted in better performance. This finding warrants use of OAA-MW for the majority of dimension filtering for later classification runs, since its use would not upwardly bias class prediction performance.

3.2. Ablation Studies

3.2.1. Random Shuffling - Stability of INEs vs. Other DR Methods

Figure 4 illustrates the classification accuracy as a function of random sampling fraction of objects during shuffling of within-dimension values. The curvilinear change in accuracy with increasing random shuffling fractions for INE was more stable when compared with other DR methods.

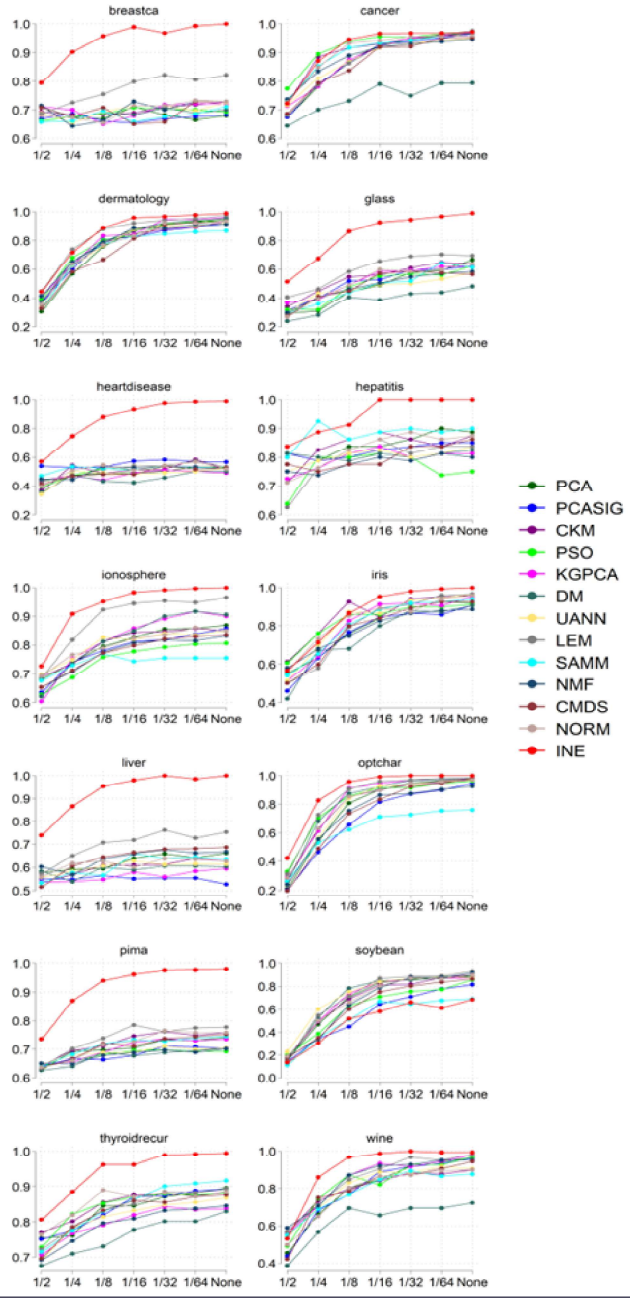

Figure 4: Ablation Study Results: Effect of Shuffling Within-Dimension Values on Classification Accuracy at Various Object Random Sampling Fractions, *f* = {1/2, 1/4, 1/8, 1/16, 1/32, 1/64}. Dimensions Used were Significant Predictors of Class Label via the OAA-MW Test (p<0.05). Accuracy is EWMV from 3 Classifiers (NBC, LVQ1, SNG)

3.2.2. Gaussian Smearing - Similarity of INEs with Other DR Methods

Figure 5 shows classification accuracy as a function of Gaussian smearing of within-dimension values. Results indicate that, irrespective of the dataset, DR method, and value of unsmeared accuracy, the behaviour of increasing annihilation of INEs was similar to and not substantially different from other smeared DR methods.

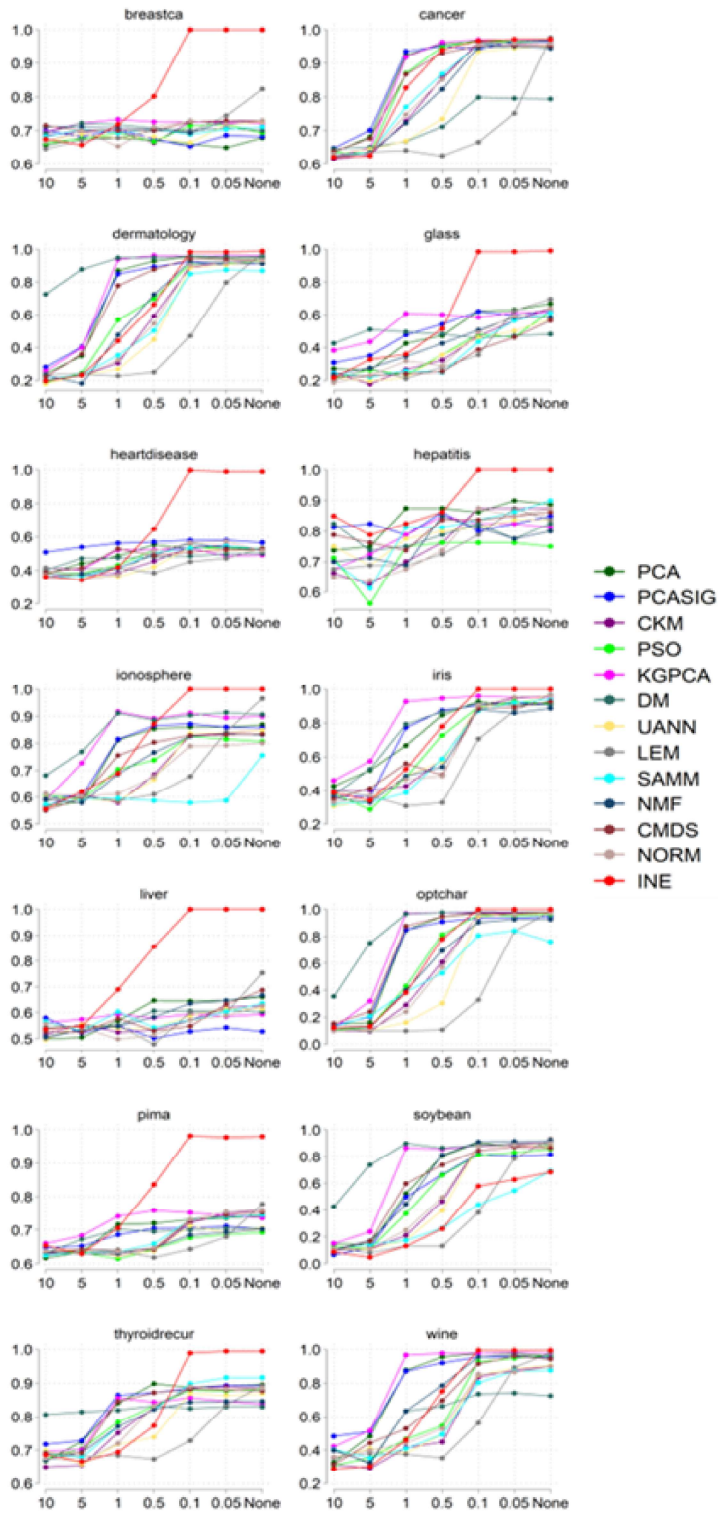

Figure 5: Ablation Study Results: Effect of Gaussian Smearing of Dimension (Feature) Values on Classification Accuracy After Adding Standard Normal Variates, $\mathcal{N}(0, \sigma^2)$, to Feature Values with σ = {10, 5, 1, 0.5, 0.1, 0.05}. Dimensions Used were Significant Predictors of Class Label via the OAA-MW Test (p<0.05). Accuracy is EWMV from 3 Classifiers (NBC, LVQ1, SNG)

3.3. All Classifiers - PCA vs. INE

Table 3 lists 10-fold CV accuracy results for all classifiers when PCA and INE were employed for DR, and dimensions were normalized prior to DR, filtered using the OAA-MW method before class prediction, and normalized during 10-fold CV. As one can note, use of the INEs resulted in improved classification accuracy for all the classifiers. Mean accuracy for ensemble classifier fusion (EWMV) for all datasets when using PCs was 84.5%, while mean accuracy when using INEs was 97.6%, indicating a 15.5% increase in performance. In addition, the standard deviation (s.d.) of EWMV accuracy over all datasets was 0.134 for PCA and 0.078 for INE, suggesting a 42% decrease. Regarding PCs, the SVM and SPSO classifiers had the worst performance, while LDA and LVQ1 had the best. For INEs, PLOG and SPSO had the worst performance, and LVQ1 and KNN classifiers performed the best. The decreased s.d. for INEs suggest greater stability over datasets.

3.4. Diverse Classifiers - All DR Methods vs. INE

Table 4 lists 10-fold CV accuracy results (EWMV only) for diverse classifiers (NBC, LVQ1, SNG) when all DR methods were used prior to class prediction. In these runs, dimensions were normalized prior to DR, filtered using the OAA-MW method before class prediction, and normalized during 10-fold CV. When averaged over datasets, the INE DR method resulted in a mean EWMV accuracy of 0.97, while the remaining DR methods resulted in mean accuracy ranged from 0.77 to 0.86. The s.d. of accuracy from use of INE was 0.08, while s.d. for the remaining DR methods was nearly twice as great (0.14-0.16). The PCA DR method, which involved selection of PCs that were significant predictors of class using filtering with OAA-MW prior to classification, resulted in a mean accuracy of 0.83. This was only slightly greater than when using all signal PCs (0.8). Typical class prediction results from selection of normalized input features (NORM DR method) resulted in

Table 3: Class Prediction Accuracya for All Classifiers and PCA and INE DR Methods

DR Method	Dataset	KNN	NBC	LDA	QDA	LREG	PLOG	LVQ1	KREG	SNG	SANN	SVM	SPSO	EMV	EWMV
PCs	Breastca	0.635	0.697	0.650	0.657	0.679	0.704	0.668	0.718	0.614	0.682	0.697	0.650	0.679	0.693
	Cancer	0.943	0.910	0.960	0.950	0.957	0.967	0.964	0.923	0.970	0.967	0.914	0.959	0.966	0.967
	Dermatology	0.951	0.937	0.956	0.937	0.959	0.888	0.940	0.951	0.954	0.948	0.863	0.839	0.967	0.962
	Glass	0.621	0.547	0.589	0.551	0.542	0.561	0.682	0.607	0.579	0.593	0.136	0.477	0.664	0.673
	Heartdisease	0.566	0.502	0.589	0.539	0.589	0.609	0.535	0.556	0.431	0.603	0.475	0.559	0.603	0.603
	Hepatitis	0.900	0.863	0.888	0.900	0.900	0.900	0.913	0.825	0.863	0.875	0.463	0.888	0.900	0.900
	Ionosphere	0.863	0.895	0.849	0.689	0.852	0.852	0.858	0.556	0.840	0.858	0.903	0.775	0.872	0.875
	Iris	0.940	0.833	0.973	0.960	0.800	0.967	0.920	0.633	0.920	0.967	0.633	0.900	0.940	0.953
	Liver	0.574	0.681	0.681	0.678	0.667	0.672	0.652	0.580	0.609	0.591	0.626	0.661	0.672	0.675
	Optchar	0.952	0.902	0.964	0.970	0.948	0.948	0.974	0.964	0.974	0.816	0.874	0.268	0.980	0.982
	Pima	0.710	0.741	0.760	0.686	0.770	0.771	0.742	0.642	0.660	0.753	0.767	0.738	0.767	0.772
	Soybean	0.654	0.673	0.872	0.008	0.887	0.045	0.872	0.891	0.880	0.808	0.767	0.207	0.906	0.902
	Thyroidrecur	0.856	0.843	0.896	0.783	0.903	0.898	0.898	0.893	0.898	0.898	0.619	0.906	0.896	0.898
	Wine	0.972	0.927	0.978	0.983	0.978	0.966	0.966	0.933	0.961	0.966	0.860	0.978	0.978	0.978
	Average	**0.796**	**0.782**	**0.829**	**0.735**	**0.817**	**0.768**	**0.827**	**0.762**	**0.797**	**0.809**	**0.686**	**0.7**	**0.842**	**0.845**
	s.d.	**0.159**	**0.143**	**0.146**	**0.263**	**0.145**	**0.25**	**0.143**	**0.163**	**0.18**	**0.143**	**0.218**	**0.247**	**0.136**	**0.134**
INEs	Breastca	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000
	Cancer	1.000	0.974	0.999	1.000	0.999	1.000	0.973	0.987	0.967	1.000	1.000	0.970	1.000	1.000
	Dermatology	0.978	0.978	0.973	0.992	0.964	0.989	0.989	0.997	0.981	0.992	0.975	0.973	0.986	0.986
	Glass	0.981	0.986	0.986	0.977	0.939	0.995	0.995	0.949	0.986	1.000	0.991	0.991	0.991	0.991
	Heartdisease	0.997	0.973	0.993	0.983	0.993	0.997	0.990	0.997	0.987	0.993	0.990	0.970	0.997	0.997
	Hepatitis	1.000	0.963	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.963	1.000	1.000	1.000
	Ionosphere	1.000	0.952	1.000	1.000	1.000	1.000	1.000	0.997	1.000	1.000	1.000	1.000	1.000	1.000
	Iris	1.000	0.907	1.000	0.967	0.993	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Liver	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Optchar	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.684	1.000	1.000
	Pima	1.000	0.992	0.987	1.000	0.988	1.000	0.980	0.991	0.975	0.984	0.999	0.987	0.995	0.995
	Soybean	0.602	0.523	0.598	0.410	0.474	0.049	0.707	0.586	0.662	0.560	0.496	0.466	0.684	0.707
	Thyroidrecur	0.997	0.992	0.997	1.000	0.997	1.000	0.997	0.997	0.992	0.992	1.000	0.984	0.997	0.997
	Wine	0.994	0.994	0.994	1.000	0.994	0.994	0.994	1.000	0.994	1.000	1.000	1.000	0.994	0.994
	Average	**0.968**	**0.945**	**0.966**	**0.952**	**0.953**	**0.93**	**0.973**	**0.964**	**0.967**	**0.966**	**0.958**	**0.93**	**0.975**	**0.976**
	s.d.	**0.106**	**0.124**	**0.106**	**0.156**	**0.139**	**0.254**	**0.077**	**0.11**	**0.089**	**0.117**	**0.133**	**0.157**	**0.084**	**0.078**

Note: [a] – Accuracy based on 10-fold CV. – Dimension (feature) selection based on OAA-MW ($p<0.05$). All dataset features normalized before DR and selected dimensions normalized during 10-fold CV.

a mean accuracy of 0.82. All together, these results suggest that input feature selection prior to class prediction provides no substantial benefit when compared with the use of signal PCs for class prediction—for the 14 datasets considered. The decreased s.d. for INEs suggest greater stability over DR methods.

Table 4: Class Prediction Accuracy Based on EWMV from 3 Diverse Classifiers (NBC, LVQ1, SNG) for All DR Methods. Numbers in Parentheses Indicate Dimensions Used

DR Method	Breastca	Cancer	Dermatology	Glass	Heartdisease	Hepatitis	Ionosphere	Iris	Liver	Optchar	Pima	Soybean	Thyroidrecur	Wine	Avg	s.d.
PCA	0.68(3)	0.97(4)	0.95(15)	0.67(8)	0.52(8)	0.89(4)	0.87(7)	0.92(3)	0.66(4)	0.98(25)	0.74(6)	0.88(34)	0.90(6)	0.96(5)	0.83	0.15
PCASIG[a]	0.68(1)	0.97(1)	0.92(4)	0.62(2)	0.57(2)	0.85(1)	0.86(5)	0.91(1)	0.53(1)	0.95(7)	0.70(2)	0.82(4)	0.89(2)	0.96(2)	0.8	0.15
CKM	0.73(2)	0.96(6)	0.93(16)	0.64(8)	0.52(11)	0.88(7)	0.85(8)	0.96(4)	0.63(5)	0.97(26)	0.76(6)	0.90(31)	0.89(7)	0.90(10)	0.82	0.14
UPSO	0.69(3)	0.97(6)	0.93(15)	0.62(8)	0.53(12)	0.75(8)	0.81(8)	0.91(4)		0.96(26)	0.69(6)	0.86(35)	0.88(8)	0.97(12)	0.81	0.14
KGPCA	0.72(1)	0.97(3)	0.96(13)	0.62(7)	0.49(8)	0.81(5)	0.90(7)	0.95(4)	0.59(2)	0.98(27)	0.74(6)	0.88(30)	0.84(7)	0.98(6)	0.82	0.16
DM	0.70(1)	0.79(1)	0.96(12)	0.48(7)	0.50(7)	0.83(2)	0.91(7)	0.91(3)	0.60(4)	0.98(26)	0.70(6)	0.89(30)	0.83(6)	0.72(4)	0.77	0.16
UANN	0.73(3)	0.95(6)	0.92(17)	0.57(8)	0.52(11)	0.84(7)	0.85(8)	0.97(4)	0.61(2)	0.97(26)	0.70(5)	0.88(35)	0.87(7)	0.90(9)	0.81	0.15
LEM	0.82(4)	0.98(6)	0.98(13)	0.70(7)	0.53(6)	0.84(6)	0.97(6)	0.97(4)	0.75(5)	0.98(24)	0.78(6)	0.93(30)	0.89(7)	0.98(6)	0.86	0.14
SAMM	0.71(4)	0.96(6)	0.87(17)	0.61(8)	0.53(9)	0.90(7)	0.75(8)	0.94(4)	0.63(5)	0.76(26)	0.74(6)	0.69(26)	0.92(8)	0.88(12)	0.78	0.14
NMF	0.73(1)	0.95(6)	0.91(16)	0.58(8)	0.53(10)	0.80(7)	0.83(8)	0.89(4)	0.67(4)	0.93(26)	0.70(6)	0.92(33)	0.85(8)	0.95(11)	0.8	0.14
CMDS	0.73(1)	0.95(3)	0.93(12)	0.57(7)	0.53(8)	0.86(4)	0.83(7)	0.93(3)	0.69(4)	0.98(19)	0.75(5)	0.86(28)	0.88(7)	0.94(4)	0.82	0.14
NORM[b]	0.73(2)	0.96(6)	0.93(16)	0.64(8)	0.52(11)	0.88(7)	0.80(1)	0.96(4)	0.63(5)	0.97(26)	0.76(6)	0.90(31)	0.89(7)	0.90(10)	0.82	0.14
INE	**1.00(4)**	**0.97(6)**	**0.99(9)**	**0.99(4)**	**0.99(11)**	**1.00(7)**	**1.00(8)**	**1.00(3)**	**1.00(5)**	**1.00(26)**	**0.98(6)**	**0.68(3)**	**0.99(8)**	**0.99(10)**	**0.97**	**0.08**

Note: [a] – PCASIG method based on all $p - m_{PC}$ signal PCs, i.e., no dimension selection. [b] - NORM method based on significant normalized input features (OAA-MW feature selection), not reduced dimensions. Dimension (feature) selection based on OAA-MW ($p<0.05$). All dataset features normalized before DR and selected dimensions normalized during 10-fold CV.

3.5. Dimension Plots for Various DR Methods

Line plots for a variety of DR methods for various datasets are shown in Figure 6. The first 3 plots for the 2-class cancer dataset reveal the drastic differences in stability and informativeness for class-specific objects among filtered normalized features (NORM), PCs (PCA), and INEs. The 2 plots for the 6-class dermatology dataset reveal that, while the PSO DR method results in dimensions that discern class labels, the INEs are more stable within classes. The two plots for the 6-class glass dataset indicate the UANN provided dimensions that appear to be informative for class, but not as strongly as the INEs. Finally, the remaining plots for the the 10-class optchar dataset for grayscaled pixel values of hand-written digits {0, 1, 2, 3, 4, 5, 6, 7, 8, 9} reveal that the normalized pixel values are not nearly as informative for discerning class as the INEs are.

4. Discussion

Improved class prediction performance from using INEs for DR is primarily due to increased within-class stability of estimates when compared with other DR methods. The increased stability of INEs based on lower s.d. of accuracy and smoother curvilinear behavior when randomly shuffled is due to several factors. First, positive values of INEs are enforced by use of normalized feature values with range [0, 1] in the numerator of (12) along with a positive sign of eigenvector elements in the denominator. Through analogous borrowing from root MUSIC, the denominator in (9) is the Euclidean distance (squared) from a vector $\mathbf{y}$ to the noise eigenspace in the form $d^2 = \mathbf{y}^\top \mathbf{E}_n \mathbf{E}_n^\top \mathbf{y}$. Therefore, the noise component needs to be positive. Second, INEs exploit the fact that the noise eigenspace varies over classes, and therefore they scale data values by noise eigenvectors, which are much smaller than, for example, the much larger eigenvectors in the signal (correlation) subspace employed by PCs from PCA. Third, the noise subspace is often larger than the signal space with more orthogonal dimensions from which to capture the subtle class-specific differences. The overall advantage for using INEs is that noise eigenvalues are smaller in magnitude, are more ubiqitous, and are unique to classes. Therefore, scaling by stable-small noise eigenvectors yields stable ratios which inflate class differences.

Dropping features whose values are invariant within a class is another important characteristic of INEs. Constant within-class feature values are present in the dermatology, glass, hepatitis, optchar, and soybean datasets. Constant within-class feature values result in a standard deviation of zero. A zero standard deviation leads to indeterminate correlation. Indeterminate correlation results in a correlation matrix with all zeroes in the rows (columns) for such features. Subsequent eigendecompostiion of a correlation matrix with zeroed out

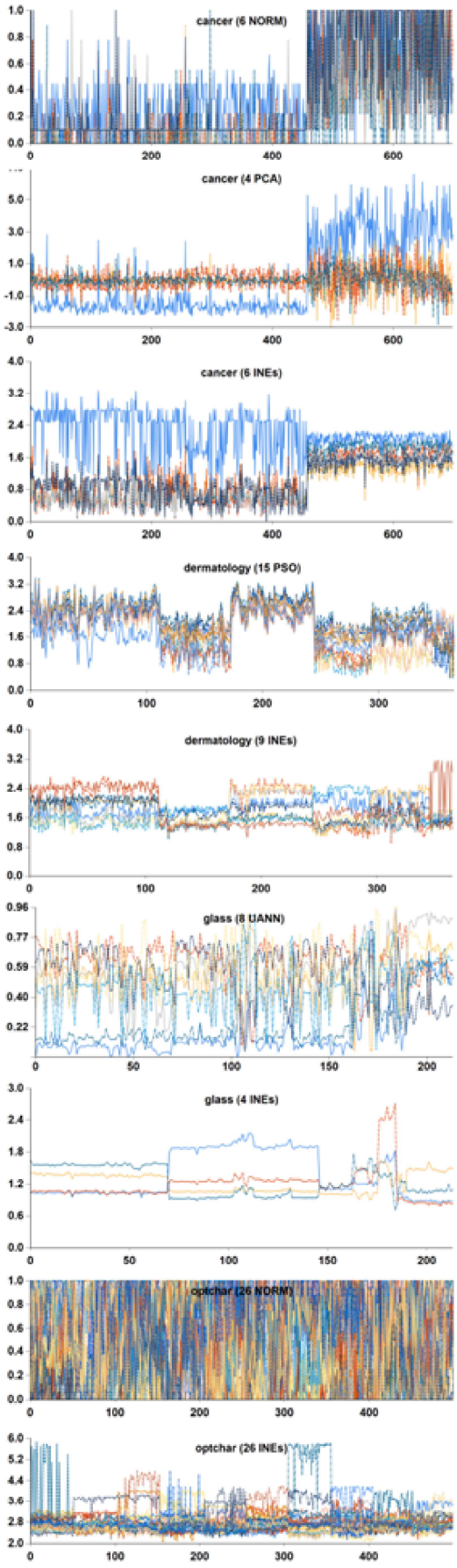

Figure 6: Plots of Filtered Dimensions (OAA-MW) Used as Inputs to Class Prediction

rows (columns) can result in eigenvalues of unity whose eigenvectors contain zero or near-zero values, for example, in the range [0, 1E-14] depending on the software used. Dividing by zero eigenvector values in the denominator of (12) is undefined, and division by near-zero eigenvector values, and then rescaling with $\log_{10}$ would result in artificially inflated INE values. If features with class-invariant values were not dropped, then INEs would merely degenerate into a form of feature selection based on within-class invariance.

Results of using an arbitrary number of dimensions for datasets with the greatest number of features reveal that the maximum value of class prediction accuracy is typically reached by use of 10 dimensions. For PCA, this is in good agreement with the rule-of thumb that inputting 10 decorrelated PC vectors into a classifier will maximize the performance. It was also observed that there was little difference between class prediction performance based on use of different DR methods. Specifically, selecting PC vectors that were informative for class separation via the OAA-MW test resulted in only a difference of 0.03 (0.83-0.80) when compared with using signal PC vectors for classification. Hence, there is also merit for employing PC vectors in the signal space for class prediction.

5. Conclusion

Until recently, pattern classification analysis has mostly focused on inputs derived from feature selection or dimension reduction. However, there has been a lack of focus on use of the class-specific noise subspace of correlation. We have shown that classification analysis of 14 datasets resulted in a 15.5% increase in class prediction performance from 84.5% to 97.6% when using INEs compared with other DR Methods. In addition, the standard deviation (s.d.) of accuracy over all datasets was 0.14-0.16 for other DR methods and 0.078 for INE, suggesting a ~48% decrease. We advocate intraclass noise eigenspace projection for pattern classification as an alternative to input feature selection or DR with other methods prior to classification analysis.

References

Alarabi, A. and Alkishriwo, O.A.S. (2021). Modulation Classification Based on Statistical Features and Artificial Neural Network. Proceedings of the 1st IEEE International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MISTA), pp. 748-751.

An, T.T. and Lee, B.M. (2023). Robust Automatic Modulation Classification in Low Signal to Noise Ratio. *IEEE Access*, 11, 7860-7872.

Belkin, M. and Niyogi, P. (2004). Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. *Advances in Neural Information Processing Systems*, Vol. 14, pp. 585-591, MIT Press, Cambridge (MA).

Cabrera, A.F., Petersen, E.B., Graversen, C., Sørensen, A.T., Lunner, T. and Rank, M.L. (2018). Individual Classification of Single Trial Eeg Traces to Discriminate Brain Responses to Speech with Different Signal-to-Noise Ratios. Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 987-990.

Deng, L. (2012). The MNIST Database of Handwritten Digit Images for Machine Learning Research. *IEEE Signal Processing Magazine*, 29(6), 141-142.

Geman, S., Bienenstock, E. and Doursat, R. (1992). Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4(1), 1-58.

Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2(5), 359-366.

Ivascu, C., Everson, R.M. and Fieldsend, J.E. (2022). Optimising Diversity in Classifer Ensembles. *SN Computer Science*, 3(191), 1-20.

Jiang, X., Mandal, B. and Kot. A. (2008). Eigenfeature Regularization and Extraction in Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3), 383-394.

Jolliffe, I.T. (1992). Principal Component Analysis and Exploratory Factor Analysis. *Statistical Methods in Medical Research*, 1, 69-95.

Kelly, M., Longjohn, R. and Nottingham, K. (2024). The UCI Machine Learning Repository. https://archive.ics.uci.edu/

Kennedy, J. and Eberhart, R.C. (1995). Particle Swarm Optimization. *Proc. of IEEE Int. Conf. Neural Networks*, 1942-1948.

Kim, K.-T., Seo, D.-K. and Kim, H.-T. (2002). Efficient Radar Target Recognition Using the Music Algorithm and Invariant Features. *IEEE Transactions on Antennas and Propagation*, 50(3), 325-337.

Kuncheva, L.I. and Whitaker, C.J. (2003). Measures of Diversity in Classifier Ensembles and their Relationship with the Ensemble Accuracy. *Machine Learning*, 51(2), 181-207.

Lafon, S. and Lee, A.B. (2006). Diffusion Maps and Coarse Graining: A Unified Framework for Dimensionality Reduction, Graph Partitioning, and Dataset Parameterization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(9), 1393-1403.

Lee, D.D. and Seung, S.H. (1999). Learning the Parts of Objects by Nonnegative Matrix Factorization. *Nature*, 401, 788-791.

Lloyd, S. (1957). Least Squares Quantization in PCM. *IEEE Trans. Information Theory*, 28, 128-137.

Maeda, K. and Watanabe, S. (1985). A Pattern Matching Method with Local Structure. *Trans. IEICE(D)*, J68-D(3), 345-352.

Marcenko, V.A. and Pastur, L.A. (1967). Distribution of Eigenvalues for Some Sets of Random Matrices. *Mat. Sb.*, (N.S.), 72(114), 4.

Mishra, D. and Sahu, B. (2011). A Signal-to-Noise Classification Model for Identification of Differentially Expressed Genes from Gene Expression Data. Proceedings of the 3rd International Conference on Electronics Computer Technology, pp. 204-208.

Mohammed, B. S. and Noori, N.H. (2014). Angle of Arrival and Signal to Noise Ratio Estimation for Broadband Signals by Phased Adaptive Array Antennas. The 8th European Conference on Antennas and Propagation (EuCAP 2014), pp. 347-351.

NXG Logic LLC (2024). Explorer CE for Windows 10/11, Version 23.9.4.1. https://www.nxglogic.com.

Ozcetin, A. (2018). Direction of Arrival Estimation Methods in Passive Coherent Locators. *Master's Thesis*, Air University, Report AFIT/GE/ENG/02M-18, U.S. Air Force Inst. of Tech.

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philos. Magazine*, 2, 559-572.

Peterson, L.E. and Coleman, M.A. (2006). Comparison of Gene Identification Based on Artificial Neural Network Pre-processing with k-Means Cluster and Principal Component Analysis. *Lecture Notes in Artificial Intelligence*, 3849, 267-276.

Peterson, L.E. (2013). Classification Analysis of DNA Microarrays. New York (NY), John Wiley and Sons.

Peterson, L.E., Hoogeveen, R.C., Pownall, H.J. and Morrisett, J.D. (2006). Classification Analysis of Surface-Enhanced Laser Desorption/Ionization Mass Spectral Serum Profiles for Prostate Cancer. Proceedings of the 2006 IEEE International Joint Conference on Neural Networks (IJCNN), pp. 3828-3835.

Pisarenko, V.F. (1973). The Retrieval of Harmonics from a Covariance Function. *Geophysics, J. Royal Astron. Soc.*, 33, 347-366.

Rehman, S.U., Sowerby, K. and Coghill, C. (2012). Effect of Receiver Signal to Noise Ratio on the Classification Performance of RF Fingerprinting. Proceedings of the 15th International Multitopic Conference (INMIC), pp. 143-148.

Sahu, B. and Mishra, D. (2011). A Novel Approach for Selecting Informative Genes from Gene Expression Data Using Signal-to-Noise Ratio and t-Statistics. Proceedings of the 2nd International Conference on Computer and Communication Technology (ICCCT-2011), pp. 5-10.

Sammon, J.W. (1969). A Nonlinear Mapping for Data Structure Analysis. *IEEE Trans. Computers*, 18(5), 401-409.

Schmidt, R.O. (1986). Multiple Emitter Location and Signal Parameter Estimation. *IEEE Transactions on Antennas and Propagation*, 34(3), 276-280.

Schölkopf, B. and Smola, A.J (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10(5), 1299-1319.

Tan, A.W.C., Rao, M.V.C. and Daya Sagar, B.S. (2007). A Signal Subspace Approach for Speech Modelling and Classification. *Signal Processing*, 87, 500-508.

Torgerson, W.S. (1952). Multidimensional Scaling: I. Theory and Method. *Psychometrika*, 17, 401-419.

Tu, T.-M., Chen, C.-H. and Chang, C.-I. (1998). Noise Subspace Projection Approach to Target Signature Detection and Extraction in an Unknown Background for Hyperspectral Images. *IEEE Transactions on Geoscience and Remote Sensing*, 36(1), 171-181.

Wax, M. and Adler, A. (2023). Vector Set Classification by Signal Subspace Matching. *IEEE Transactions on Information Theory*, 69, 1853-1865.

Yamaguchi, O. and Fukui, K. (2022). Image-Set Based Classification Using Multiple Pseudo-Whitened Mutual Subspace Method. Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM), pp. 296-306.

Cite this article as: Leif E. Peterson (2024). Improved Dimension Reduction for Pattern Classification Using Noise Eigenspace Projection. *International Journal of Data Science and Big Data Analytics*, 4(1), 34-51. doi: 10.51483/IJDSBDA.4.1.2024.34-51.