Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

A Classical Machine Vision Pipeline for MNIST Handwritten Digit Recognition: LBP and HOG Features with From-Scratch J2 Discriminant Compression

Arooba Akhtar
Department of Computer Science, Hazara University Dhodial, Mansehra, Pakistan
Shahzeb Jadoon
Department of Computer Science, Abbottabad University of Science and Technology (AUST), Pakistan
Shanza Mehboob
Department of Computer Science, Abbottabad University of Science and Technology (AUST), Pakistan
Sohair Sultan Fayyaz Jeelani
Department of Computer Science, Abbottabad University of Science and Technology (AUST), Pakistan
Iqra Khalid
Department of Computer Science, Abbottabad University of Science and Technology (AUST), Pakistan
Nida Zainab (Corresponding Author)
Department of Computer Science, Abbottabad University of Science and Technology (AUST), Pakistan Email: nidazainab1995@gmail.com

ABSTRACT

This paper presents the full pipeline for MNIST handwritten digit recognition using Machine Learning. The pipeline consists of three stages: (1) feature extraction using Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG) descriptors; (2) feature compression using the J2 criterion (Fisher's discriminant criterion), without using any sklearn LDA; (3) classification using Support Vector Machine (SVM) with an RBF kernel and Linear Discriminant Analysis (LDA). The MNIST dataset is used with 10,000 and 2,000 samples for training and testing, respectively, across six methods. The maximum accuracy of 95.95% was obtained with combined LBP+HOG features, J2 compression, and SVM classification, reducing the 342-dimensional features to 9 discriminant components. The results show that HOG features significantly improve digit recognition compared to LBP alone, and that the J2 criterion can be useful for maximizing class separability and drastically reducing dimensionality.

Index Terms—MNIST, LBP, HOG, J2 Criterion, Fisher Discriminant, Feature Compression, SVM, LDA, Handwritten Digit Recognition

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Introduction

HANDWRITTEN digit recognition is one of the most widely researched problems in computer vision and pattern recognition. It has a wide range of applications, including postal address sorting, bank cheque processing, automated
form digitization, meter reading systems and examination paper marking [19], [12]. Since its release in 1998 by LeCun et al. [15], MNIST has remained the standard dataset for assessing handwritten digit recognition algorithms and consists of 70,000 greyscale images of handwritten digits (0-9) at a resolution of 28×28 pixels. Due to its well-defined training and test sets and its wide popularity, it has served as the foundation for decades of methodological development, from traditional feature engineering pipelines to contemporary end-to-end deep architectures, as thoroughly reviewed in recent benchmark surveys [2].

With the advent of Convolutional Neural Networks (CNNs), many systems have achieved near-human accuracy on MNIST in fully end-to-end trainable systems [14], [1]. However, traditional feature-based pipelines remain quite relevant when classical pipelines are required for resource-limited embedded hardware, model interpretability constraints, insufficient quantities of labelled training data, or deployment on a target that cannot support GPU acceleration [20]. The ability to perform handcrafted feature extraction, supervised dimensionality reduction, and discriminative classification remains an essential skill in machine vision education and research [8].

For any classical recognition process, feature extraction is the important first step. The Local Binary Pattern (LBP) descriptor, introduced by Ojala, Pietikäinen and Mäenpää [17], is designed to capture local micro-texture by thresholding the neighborhood of pixels within a circle and summing the resulting binary codes into a normalized histogram. Due to its computational simplicity and robustness to monotonic illumination changes, LBP has attracted significant interest in a variety of face recognition, texture analysis, and scene classification applications [13]. The complementary approach is the Histogram of Oriented Gradients (HOG) proposed by Dalal and Triggs [7], which is based on the distribution of gradient orientations in spatially structured cells, using block normalization with overlapping. HOG encodes the global stroke geometry in terms of edge direction and curvature, which are much more discriminative than local texture statistics for character-level recognition tasks and are very successful in object detection pipelines [9]. In addition to LBP and HOG, a large group of classical handcrafted descriptors, such as SIFT [16], are scale-invariant key point descriptors. A combination of complementary descriptors into a single feature vector has consistently been found to boost recognition accuracy compared to single descriptors alone.

High-dimensional feature representations often contain redundant features, which can affect the generalization of classifiers. The J2 criterion is based on Fisher's original linear discriminant formulation [10]. It has been extended to multi-class problems by Belhumeur et al. [3], which projects features into the subspace where the trace criterion

J_2 = trace $\left(S_w^{-1} S_b\right)$ is maximum, where Sw and Sb are within-class and between-class scatter matrices, respectively. In
a ten-class problem, this projection can reduce the feature dimension to at most C−1 =

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

9 discriminant components, while maintaining the class-relevant structure. The resulting compressed representations are then classified by two standard classifiers: the Support Vector Machine (SVM), which uses the maximum margin separating hyperplane in a kernel-induced feature space and the Linear Discriminant Analysis (LDA), which is a linear classifier that is particularly appropriate for the low-dimensional J2 compressed subspace [11], [18].

This report provides a systematic, fully reproducible evaluation of 6 method combinations, including three feature types (LBP, HOG, and LBP+HOG) and 2 classifiers (SVM and LDA), with J2 compression performed from scratch using NumPy without any external linear discriminant analysis library. The main contributions are: (1) a careful comparative evaluation showing the complementary nature of LBP and HOG within a single J2 compression structure; (2) a clear and library-free implementation of the J2 criterion using NumPy which has been evaluated on the MNIST benchmark; and (3) quantitative analysis of recognition on a per digit basis, for all the 10 digit classes and all the 6 combinations of methods.

Dataset Description

Dataset Overview and Historical Background

The MNIST database (Modified National Institute of Standards and Technology) was created by LeCun et al. [15] from two existing NIST (National Institute of Standards and Technology) sets of handwritten digit images collected by US Census Bureau employees and American secondary school students. Sample images from MNIST dataset is shown in Fig.1. To minimize preprocessing burden, keeping the data set as simple as possible, while still retaining enough stylistic diversity to challenge the pattern recognition systems over a wide range of writing styles and stroke widths, the following were taken into account: After its public release in 1998, MNIST has become one of the most cited benchmarks in the pattern recognition literature, referenced in thousands of peer-reviewed papers, ranging from classical feature-based methods, to kernel machines, to modern deep learning architectures [2]. A more recent extension of the MNIST dataset, called EMNIST [5] was then proposed to include handwritten letters as well as digits, thus expanding the range of evaluation of the MNIST dataset. However, the original ten-class digit recognition problem remains the standard benchmark for systematic comparisons of machine vision pipelines in controlled, repeatable settings.

Data Characteristics and Format

The full MNIST corpus consists of 70,000 grey images of the digits, all normalized to 28×28 and resulting in a raw feature dimensionality of 784 scalar values/image. During the original dataset construction, all digit images were size-normalized and center-of-mass-aligned within a 28×28 bounding frame [15], which significantly reduces spatial variations due to differences in writing styles, digit scale, and digit position. The pixel intensities are encoded as an unsigned 8-bit integer ranging from 0 (white background) to 255 (maximum foreground ink density).

As a common practice for machine learning pipelines [4] all pixel values are linearly rescaled to the continuous interval [0, 1] before any feature extraction. This normalization step avoids some pixel magnitudes from dominating gradient

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

calculations when computing the HOG descriptor and keeps the numbers numerically stable when computing the J2 scatter matrix. The dataset consists of ten mutually exclusive digit classes (0-9), with around 7,000 samples per digit class, resulting in a quite even distribution and no significant skew towards any particular digit class.

Experimental Dataset Partition

In this study, stratified random sampling is used to select 12,000 samples from the entire MNIST corpus, with 10,000 used for training and 2,000 for testing demonstrating in Table I. Stratification is performed at the class level to ensure that the proportions of each digit class in both the training and test partitions match those of the parent population, thus avoiding any systematic bias that might be introduced by accidental class imbalance in the reported evaluation metrics [8]. The resulting configuration consists of about 1,000 training samples per digit class and 200 test samples per digit class.

All sampling, model fitting and evaluation is done using a fixed random seed of 42 to ensure full reproducibility of all reported results. The hyperparameters of the SVM classifier are not data-driven; they are fixed a priori using previously determined values from the literature [6], without a held-out validation partition. There are no free hyperparameters that must be cross-validated and selected in the J2 compressor and LDA classifier.

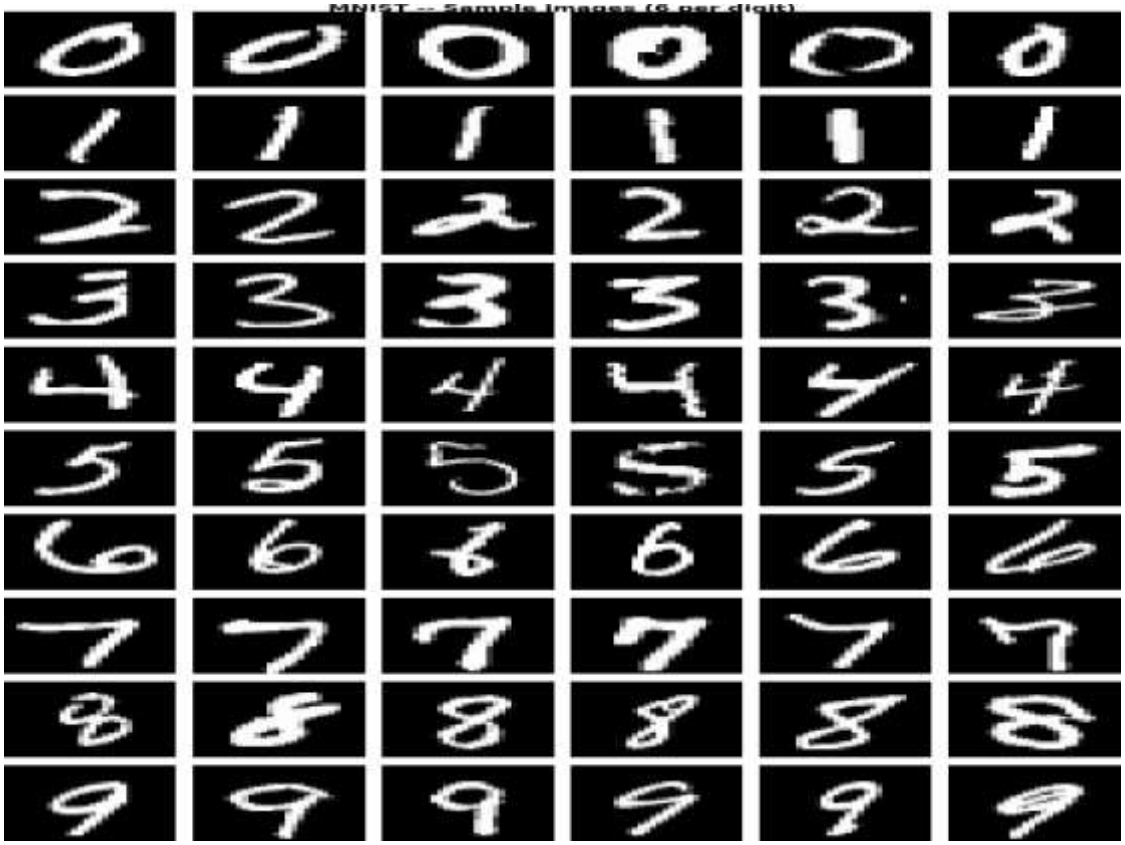


Fig. 1: Sample images from MNIST dataset. Six typical examples of each digit class (0–9) are depicted; for each class, the range of stroke width, slant, and style in

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

which they are written is very large.

TABLE I: Dataset Split and Distribution

Split	Total Samples	Approx. Samples per Class	Proportion
Training	10,000	1,000	83.3%
Test	2,000	200	16.7%
Total	12,000	1,200	100%

Scope and Representational Limitations

While MNIST is considered a solved benchmark for deep convolutional architectures [14], it still has significant value as a principled test of a classical feature-extraction and dimensionality-reduction pipeline. The homogeneous white background, centered digit placement, and the even distribution of classes across the dataset all provide a mechanism for isolating the effect of each stage of the pipeline from background clutter, occlusion, multi-scale variation, and uncontrolled illumination. These properties are very beneficial for studying the J2 criterion, as the clean, discriminative signal allows proper evaluation of the interaction between feature type and compression quality, free of scene-level complexity.

It must be said, however, that findings from MNIST may not be directly applicable to unconstrained real-world handwriting recognition tasks, as other factors, such as font variations, writing medium, background texture, and stroke

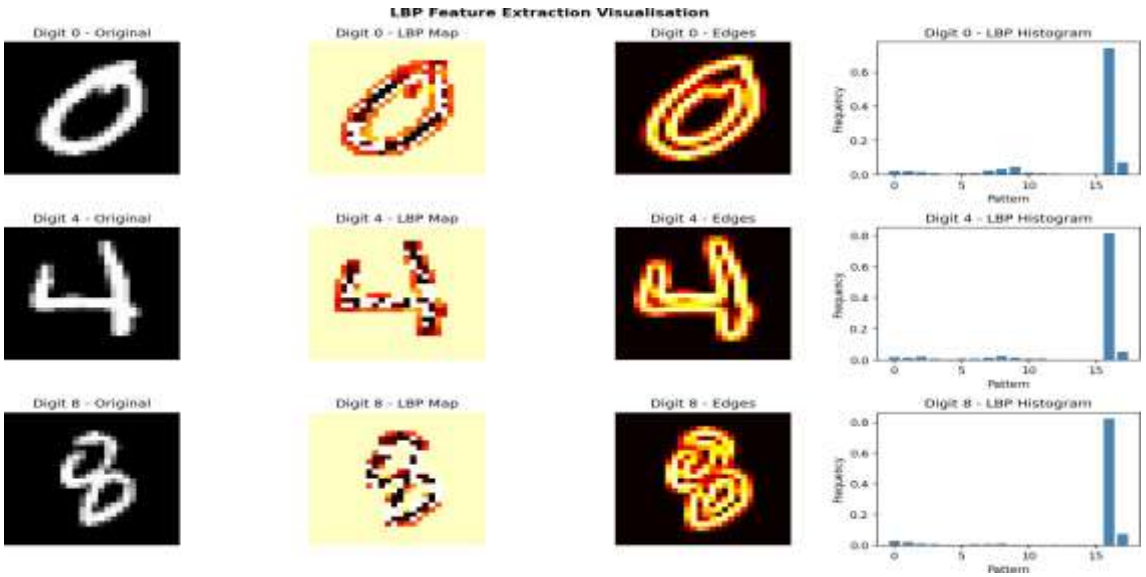


Fig. 2: Visualization of LBP features extracted for digits 0, 4 and 8. The image is displayed from left to right: Original greyscale image, LBP map (brighter pixels

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

represent a higher LBP code), Sobel edge map, normalized 18-bin LBP histogram (feature vector).

degradation, pose additional challenges [19], [12]. In addition, to ease the computational burden and better separate the impact of feature quality from that of large-sample training, the study uses a 10,000-sample training partition instead of the full 60,000-sample training partition. It is indicated that the effects of training set size on the compression quality of J2 and classification accuracy need to be investigated.

Methodology

The recognition pipeline consists of the following 3 steps: feature extraction, feature compression using J2 criterion, and classification. The raw 28×28 images are first mapped to small feature vectors and then to a low-dimensional discriminant space before being fed to a classifier. All stages are coded in Python using the NumPy and scikit-learn libraries [18], except the J2 compression stage, which is coded from scratch without using any external LDA library.

Feature Extraction

1) Local Binary Pattern (LBP):: The Local Binary Pattern (LBP) is another widely used texture descriptor in image recognition. The LBP descriptor was later introduced by Ojala, Pietikäinen, and Mäenpää [17] and encodes local texture by comparing each pixel to a circular neighborhood of P equally spaced sampling points with radius r. A binary code is generated from each neighbor p_i by thresholding it for each central pixel p_c:

$$LBP(P, r) = \sum_{i=0}^{P-1} s(p_i - p_c) \times 2^i$$

In this report, the uniform LBP variant is used, where only patterns with at most 2 bitwise transitions are considered valid, reducing the histogram to P+2 bins and discarding statistically rare non-uniform patterns [13]. An 18-dimensional normalized histogram represents each image as a feature vector, based on P = 16 sampling points at radius r =

2. According to Fig.2, the LBP map, edge response, and the resulting histogram of representative digit samples are displayed.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

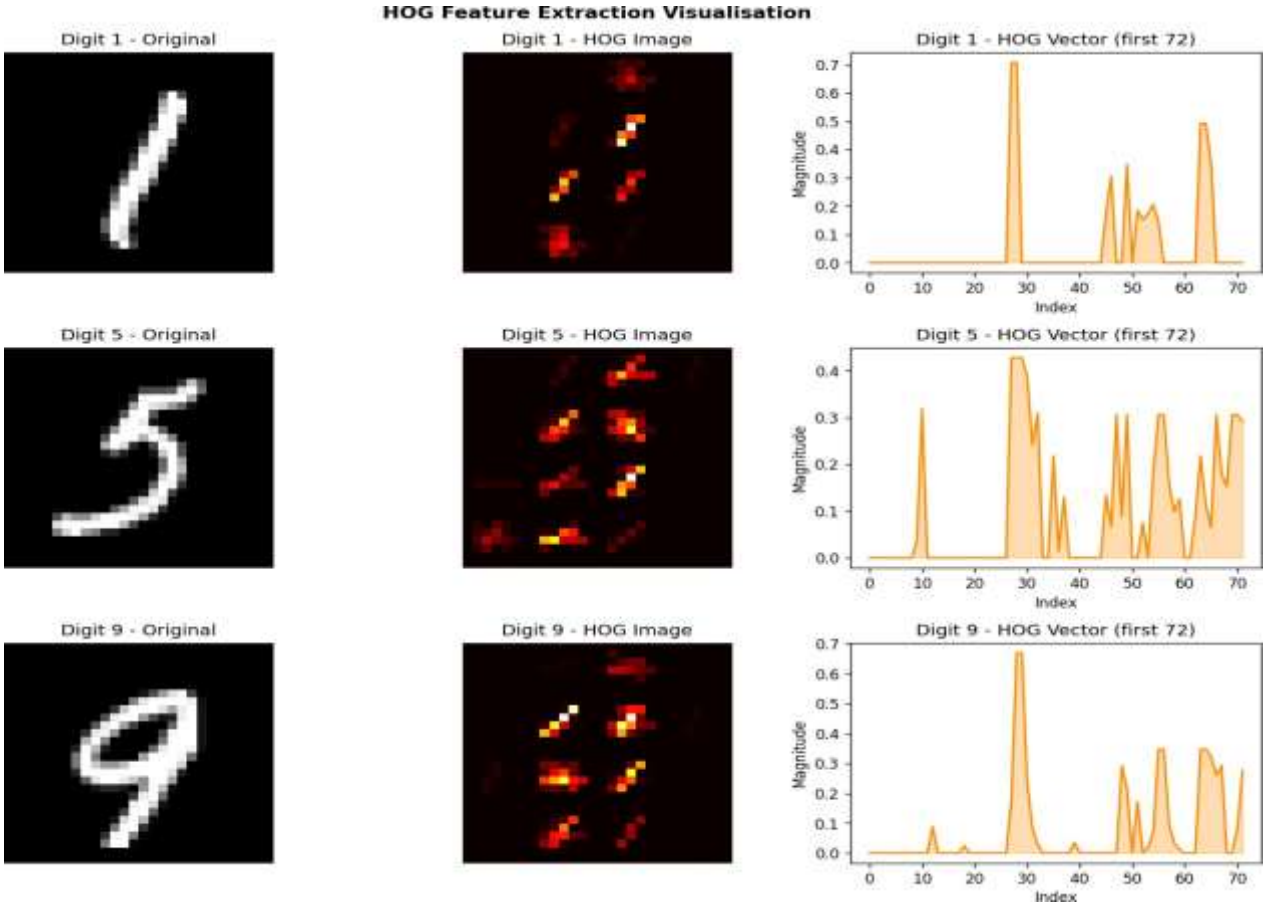


Fig. 3: HOG feature extraction visualization of digits 1, 5 and 9. For each row, original image, HOG visualization (lines are the gradient directions) and the first 72 values of the 324-dimensional HOG feature vector.

Histogram of Oriented Gradients (HOG): The local distribution of gradient orientations within a spatially structured grid of cells is captured by the HOG descriptor proposed by Dalal and Triggs [7]. For each 7×7 pixel cell, the magnitude and direction of the gradients are summed up in a histogram of 9 orientation bins ranging from 0° to 180°. Histograms of successive 2×2 cell blocks are joined together, and L2-Hys is normalized to minimize sensitivity to local illumination variation. Using a 28×28 image and the above configuration, HOG generates a 324-dimensional feature vector for each image. In contrast to LBP, HOG explicitly represents the directional structure of digit strokes, which is key information for distinguishing visually similar digit classes [9]. In Fig.3, the HOG visualizations of selected digits and their respective feature vectors are displayed.

Combined LBP+HOG Features: To benefit from the complementary properties of both descriptors, the LBP and HOG feature vectors are concatenated along the feature axis, resulting in a 342-dimensional combined descriptor (18
+ 324 = 342). LBP provides local micro-texture information, and HOG provides global stroke geometry information; the combination of these two is expected to provide a more comprehensive and discriminative feature space than either descriptor alone [16]. Each of the three configuration modes (LBP, HOG, and LBP+HOG) is

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

tested separately so they can be compared in a controlled way.

J2 Criterion Feature Compression (Implemented from Scratch)

The classical feature-compression method, J2 criterion (Fisher's trace criterion), finds the low-dimensional projection that maximizes between-class separability with respect to within-class variability. This implementation does not use sklearn LDA; only NumPy is used for compression.

2) Mathematical Foundation: Fisher [10] and Belhumeur et al. [3] proposed a principled approach to supervised dimensionality reduction, known as the J2 criterion. For a given labeled data set, two scatter matrices are calculated: the within-class scatter matrix S_w, which captures the variance of samples within each class around class centers, and the between-class scatter matrix S_b, which captures the variance of class centers around the global mean.

$$S_w = \sum_{c} \sum_{i \in c} (x_i - \mu_c)(x_i - \mu_c)^T$$

The criterion value of J2 is:

$$S_w = \sum_{c} n_c (x_i - \mu_c)(x_i - \mu_c)^T \quad (3)$$

$$J_2 = \text{trace}\left(S_w^{-1} S_b\right) \quad (4)$$

To maximize J2, we will obtain the generalized eigenvalue problem $S_w^{-1} S_b \mathbf{w} = \lambda \mathbf{w}$. Each feature vector is mapped into a 9-dimensional discriminant representation by the projection matrix W, which is composed of the top C = 10 eigenvectors of the C = 10 digit classes.

3) Implementation Details: These are the steps that are followed to implement the J2 compressor from scratch using NumPy only:

- Standard Scaler is used to standardize the features, so that each feature has a mean of 0 and a variance of 1.
- To address numerical ill-conditioning in S_w before eigen decomposition for high-dimensional inputs such as HOG and LBP+HOG, a PCA preprocessing step retaining 98% of variance is applied.
- To explicitly compute the scatter matrices S_w and S_b, we need to iterate over the class labels, create outer products, and accumulate.
- To avoid exact singularity, S_w is regularized with a ridge term $S_{w,reg} = S_w + \alpha \times \mathbf{I}$ with $\alpha = 1e{-}4$.
- The eigen decomposition of $S_w^{-1} S_b$ is computed using numpy.linalg.eig, and the first 9 real eigenvectors, sorted by decreasing eigenvalue, are used to construct the projection matrix $\mathbf{W}$.
- The same $\mathbf{W}$ is used to project both training and test features.

4) J2 Eigenvalue Analysis: The eigenvalue spectra and cumulative discriminability curves for LBP, HOG, and combined LBP+HOG features are shown in Fig.4, Fig.5, and Fig.6 respectively. LBP, HOG, and LBP+HOG accordingly have J2 criterion

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

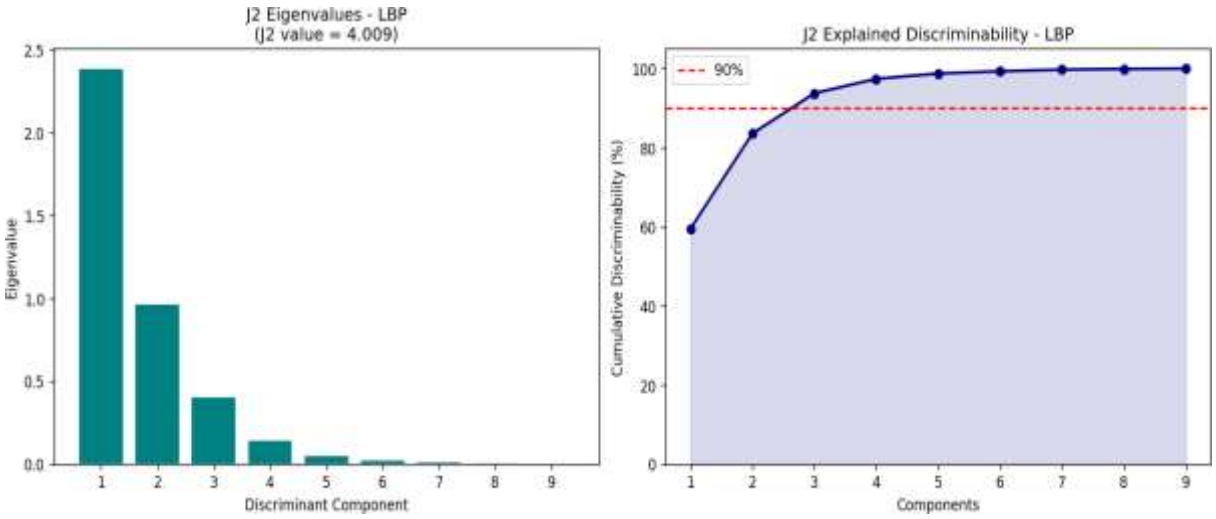
values of 4.01, 23.92, and 26.20, respectively. The significantly higher J2 value for the HOG-based features indicates that gradient orientation descriptors are more class-separable than local binary texture patterns at this scale. The best J2 value is obtained with the combined LBP+HOG configuration, indicating that combining features yields a discriminant structure beyond the individual descriptors.

5) J2 Feature Space Visualization: Scatter plots of training samples in the first two J2 discriminant components for the respective feature types are shown in Fig.7, Fig.8, and Fig.9 respectively. As shown in the LBP feature space, there is significant overlap between the classes, consistent with the low J2 value and poor classification accuracy. The HOG and combined LBP+HOG spaces exhibit distinct cluster structures, and digit 1 is outstandingly well separated across all feature types due to its unique single-stroke vertical feature geometry.

Classification

Support Vector Machine (SVM): The main classification technique used is SVM (Radial Basis Function (RBF) kernel) [6]. SVM finds the maximum-margin separating hyperplane in a kernel-induced feature space, with good generalization when the data are sparse relative to the number of features. To classify multi-class (10-digit classes), the one-vs-rest (OvR) approach is used. As shown in Table II, the regularization parameter is set to C = 10, and the kernel width is gamma = scale, according to the standard values recommended for digit recognition problems [8]. The features are J2-compressed and re-standardized and fed into the SVM.

Linear Discriminant Analysis (LDA) Classifier: As the second classifier, LDA provides a linear baseline that is computationally efficient [11]. LDA models the class-conditional distribution as Gaussian, with a common covariance matrix and classifies new samples by assigning them to the nearest class centroid in the projected discriminant space. The compression in J2 maximizes between-class scatter relative to within-class scatter, and hence the compressed features are well suited to the assumptions of LDA, leading to a natural pairing. The LDA training time is less than 0.03 seconds for all the feature configurations, which provides a good speed-accuracy trade-off for scenarios where the computational resources are limited [20].



Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Fig. 4: LBP features J2 eigenvalues and cumulative discriminability. J2 criterion value = 4.01. The small J2 value validates that 18-dimensional features of LBP map give poor class separation.

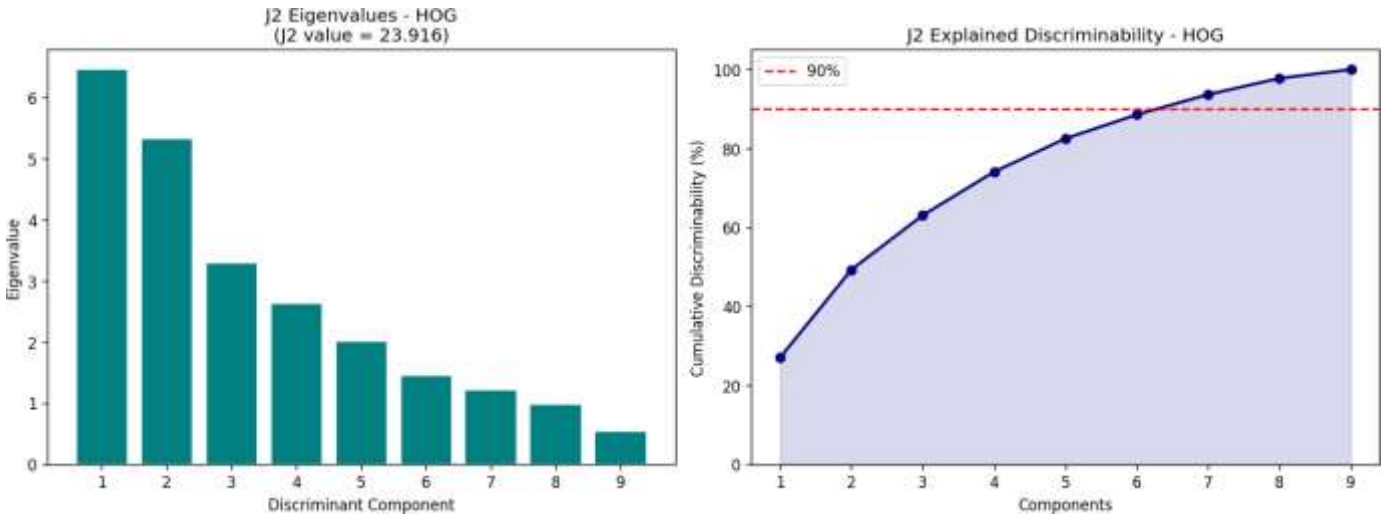


Fig. 5: J2 eigenvalues and cumulative discriminability for HOG features. J2 criterion value = 23.92. The value of J2 is significantly larger (6 times larger than LBP), indicative of better class discriminability of HOG. The first two components account for about 50% of the total discriminability.

Experimental Results

In this section, the quantitative and qualitative results for all six method combinations on a test set of 2000 samples are presented. Results are presented in terms of classification accuracy, confusion matrices, per-digit precision and recall, and a summary table. The same J2-compressed 9-dimensional feature space has been used for all experiments to allow a direct and fair comparison between the different feature types and classifiers.

Confusion Matrices

All six method combinations are presented in the confusion matrices. The raw count form is shown on the horizontal axis, and the row-normalized form on the vertical axis to emphasize per-class recall. A good classifier will generate a matrix in which the diagonal elements have high intensity and the non-diagonal elements have close to zero intensity.

LBP-Based Methods: The confusion matrix for the LBP + J2 + SVM combination is shown in Fig.10, which achieved an overall test accuracy of 56.25%. This matrix shows pervasive off-diagonal errors in the matrix for almost all digit classes, suggesting that the 18-dimensional LBP histogram is not adequate to distinguish structurally different digit classes in the J2 compressed space. Digits 0 and 1 achieve the highest individual recall; this may be due to their shapes, which generate unique global texture histograms even in the simple LBP representation. The digits 2-9, however, are very confusing to one another, with many classes having less than 50% recall.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

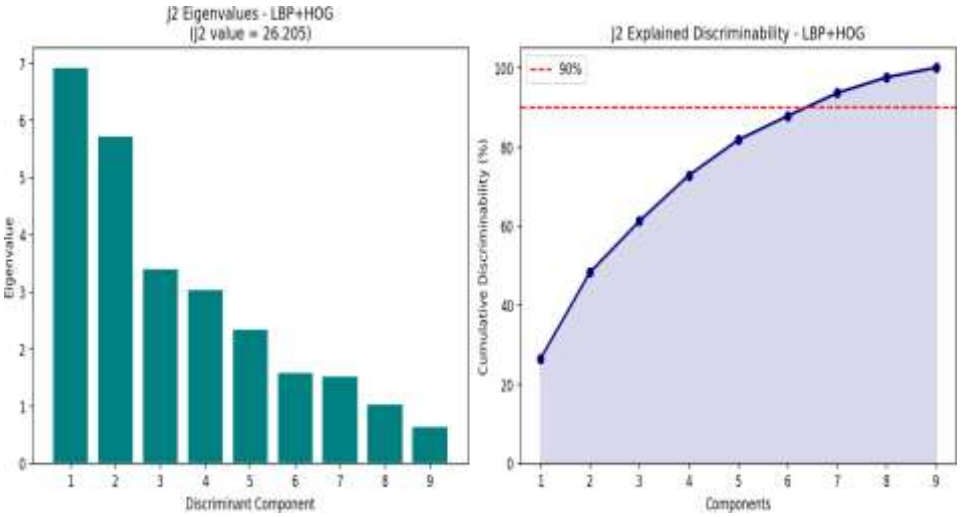
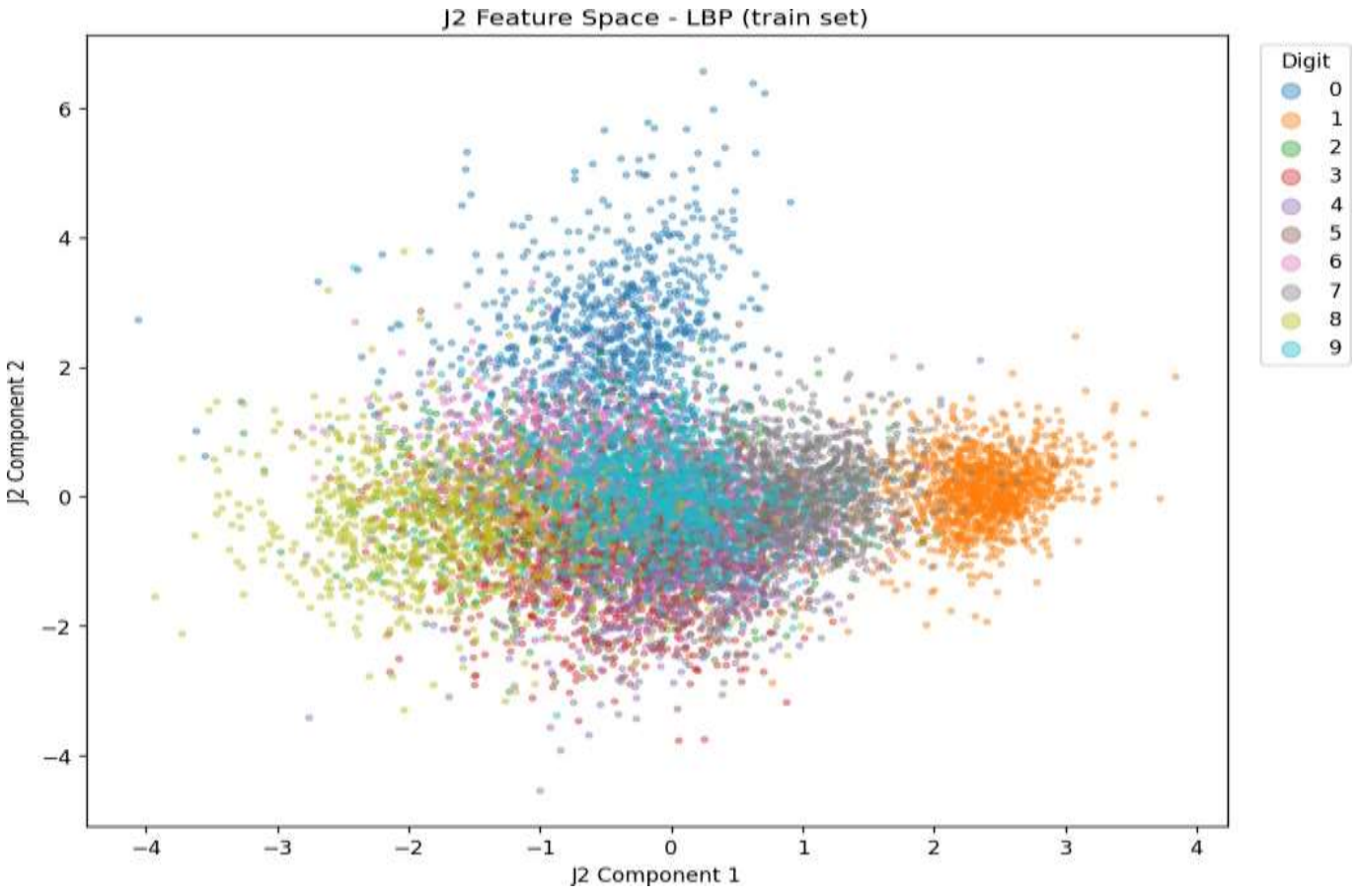


Fig. 6: J2 eigenvalues and cumulative discriminability for Combined LBP+HOG features. The result shows that the J2 criterion value is 26.20 (the best among all features), indicating that feature fusion increases class separability. The discriminability level is 90%, all 9 components are required.



Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Fig. 7: J2 feature space scatter plot for LBP features (10 000 training samples). However, there is significant overlap between digit classes, which accounts for the poor accuracy (56.25%) of LBP-based methods.

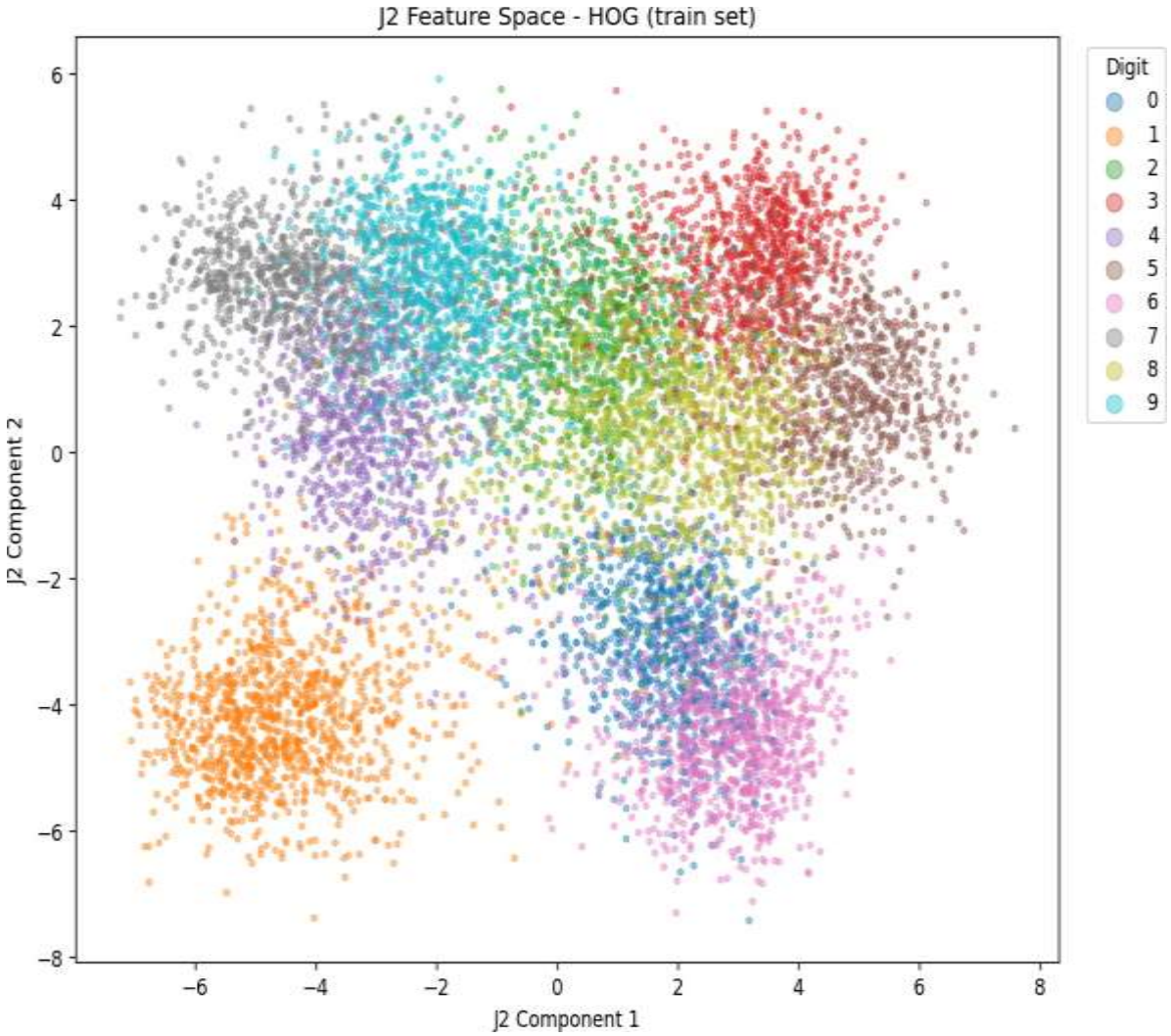


Fig. 8: J2 feature space scatter plot for HOG features (training set). There is a clear cluster structure, with digit 1 (orange) well separated from all other digits. The remaining classes cluster, but not in a separate manner.

TABLE II: Support Vector Machine Hyperparameters

Parameter	Value
Kernel	RBF (Radial Basis Function)
C (regularisation)	10
Gamma	scale $(1/(n_{features} \times X.var()))$
Multi-class strategy	One-vs-Rest (OvR)

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

The corresponding matrix for LBP + J2 + LDA is shown in Fig.11, achieving 54.95% accuracy. The error pattern is also similar to that of SVM, suggesting that the performance limitation stems from the feature representation itself rather than the classifier selection. The normalized confusion matrix shows the confusion was especially high between pairs of digits with similar curvature patterns in their strokes, for example, 3 and 5, 4 and 9, and 7 and 2, because the global LBP histogram poorly captures this kind of information.

1) HOG-Based Methods: The confusion matrix for HOG + J2 + SVM, with an accuracy of 95.55%, is shown in Fig.12. The matrix is highly diagonal, with most digit classes achieving a recall > 0.93. Since digit 1 is geometrically very simple, with a single vertical line, it has a very distinctive HOG gradient orientation profile and therefore almost perfect recall. The remaining errors are mostly structurally similar ”confusion pairs”: digit 3 is sometimes mis predicted as 5, digit 4 as 9, and digit 7 as 2, all of which exhibit partial stroke curvature and loop structures.

The confusion matrix for HOG + J2 + LDA is shown in Fig.13, with an accuracy of 94.60%, which is just lower than that of SVM. The error distribution shows a similar spatial distribution to the SVM case, with digit 9 demonstrating

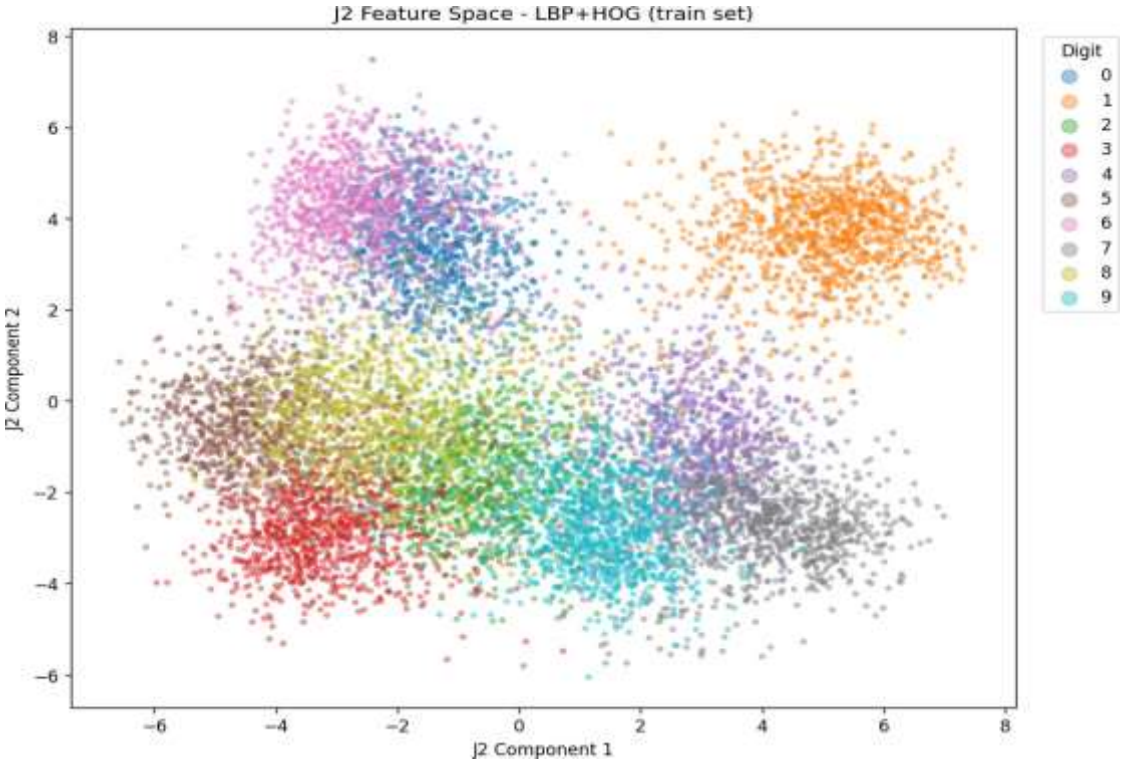


Fig. 9: J2 feature space scatter plot for Combined LBP+HOG features (training set). This is the space that presents the most clearly defined cluster structure among the three feature types, with the greatest J2 criterion value (26.20) and the best classification accuracy (95.95%).

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

the lowest individual recall of about 0.86. The small difference in performance between SVM and LDA for the HOG features indicates that the space of J2-compressed HOG features is mostly linearly separable, and that the non-linear RBF kernel for the SVM offers only a slight advantage.

2) Combined LBP+HOG Methods (Best Results): The confusion matrix for the best configuration (LBP+HOG + J2 + SVM), with a test accuracy of 95.95%, is shown in Fig.14. Digit 1 scores 1.00, indicating perfect recall. All other digit classes have recall values greater than 0.93, with digit 2 having the lowest recall of about 0.94 and digit 3 having the lowest of about 0.93, because both digit 2 and 3 have a closed loop structure. The off-diagonal entries are sparse and tend to occur near structurally similar class pairs, corresponding to the expected human confusion patterns.

The confusion matrix for LBP+HOG + J2 + LDA (with accuracy 95.60%) is shown in Fig.15 and is just 0.35 percentage points lower than the corresponding SVM result. The diagonal structure is also robust, and the distribution of residual errors is quite similar to that of the SVM configuration. This proximity between classifiers suggests that the combined LBP+HOG descriptor, compressed using J2, has a nine-dimensional feature space that is almost near-linearly separable for all ten-digit classes.

Overall Accuracy Comparison

A bar chart comparing the test accuracy for all six method combinations is shown in Fig.16 and Table III. The results show a distinct two-tiered distinction. The accuracy obtained by the LBP-based method is much lower (56.25% for SVM and 54.95% for LDA) than the HOG-based results (a difference of more than 38%). The performances of the HOG and combined LBP+HOG methods are clustered within a very narrow range from 94.60% to 95.95%, indicating that the

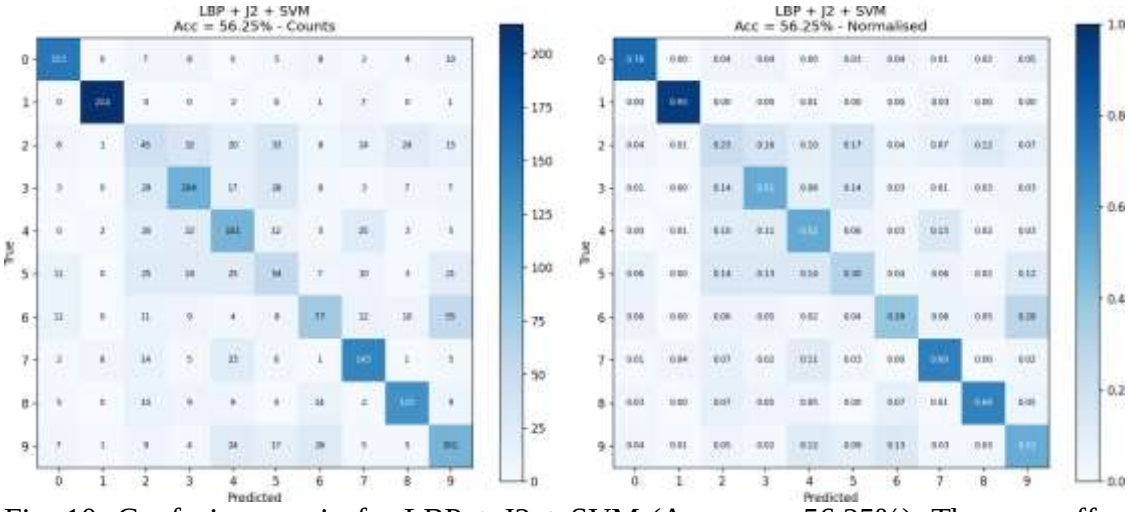


Fig. 10: Confusion matrix for LBP + J2 + SVM (Accuracy: 56.25%). The very off-

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

diagonal entries make it clear that LBP features alone are not sufficient to ensure good digit recognition. It is observed that digits 0 and 1 are well classified, whereas digits 2-9 suffer from severe confusion.

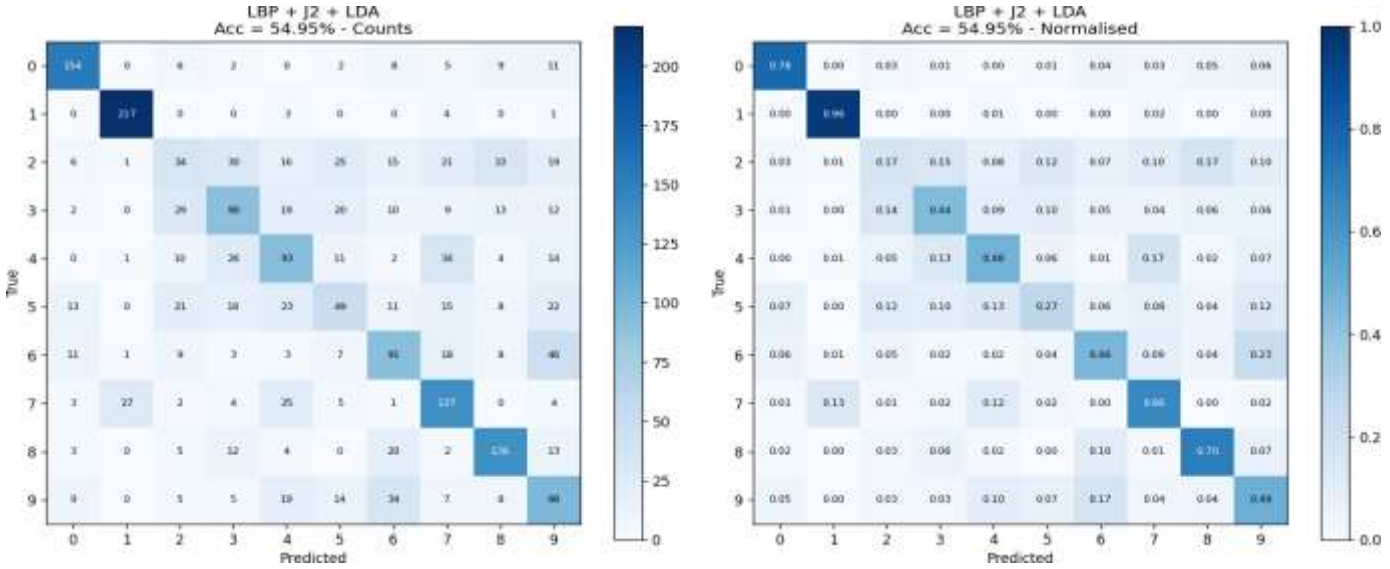


Fig. 11: Confusion matrix for LBP + J2 + LDA (Accuracy: 54.95%). This is comparable to SVM with LBP features, where there is large variation in the percentage of digits misclassified in a particular category, with structurally similar digits being largely misclassified (e.g. 3 vs 5, 4 vs 9).

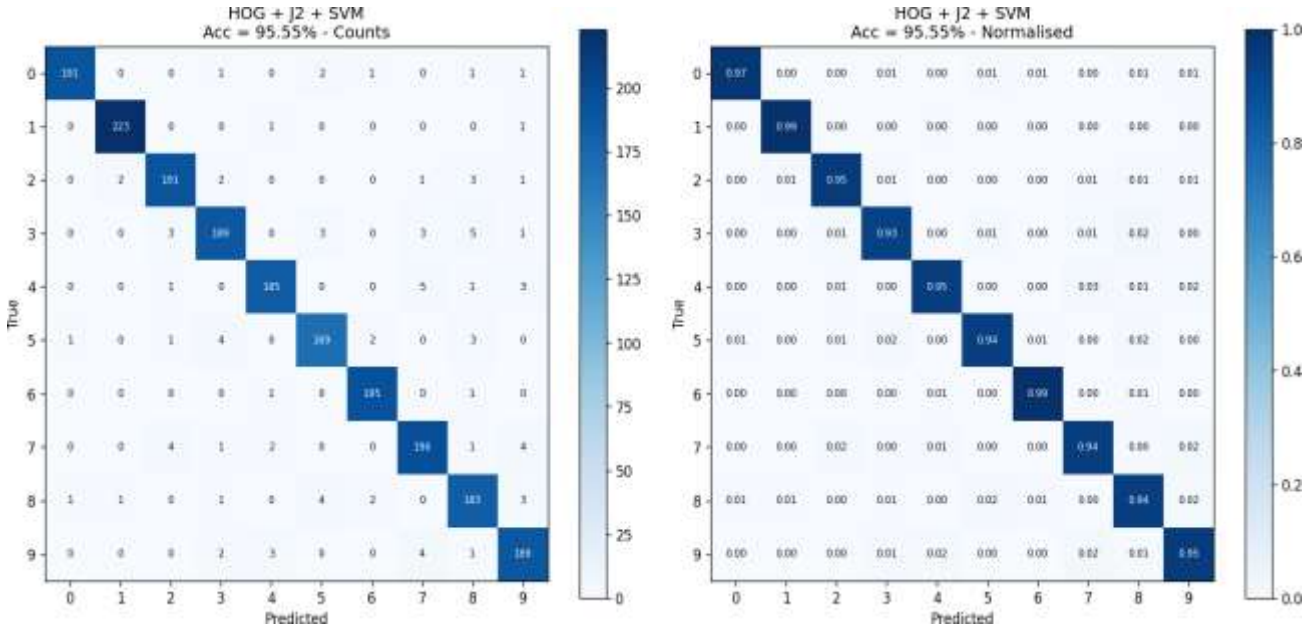


Fig. 12: Confusion Matrix for HOG + J2 + SVM (Accuracy: 95.55%). Diagonal dominance is desirable and indicates that HOG well captures the digit structure.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Digit 1 gets almost full marks for recall. There is some confusion between visually similar pairs (3/5, 4/9, 7/2).

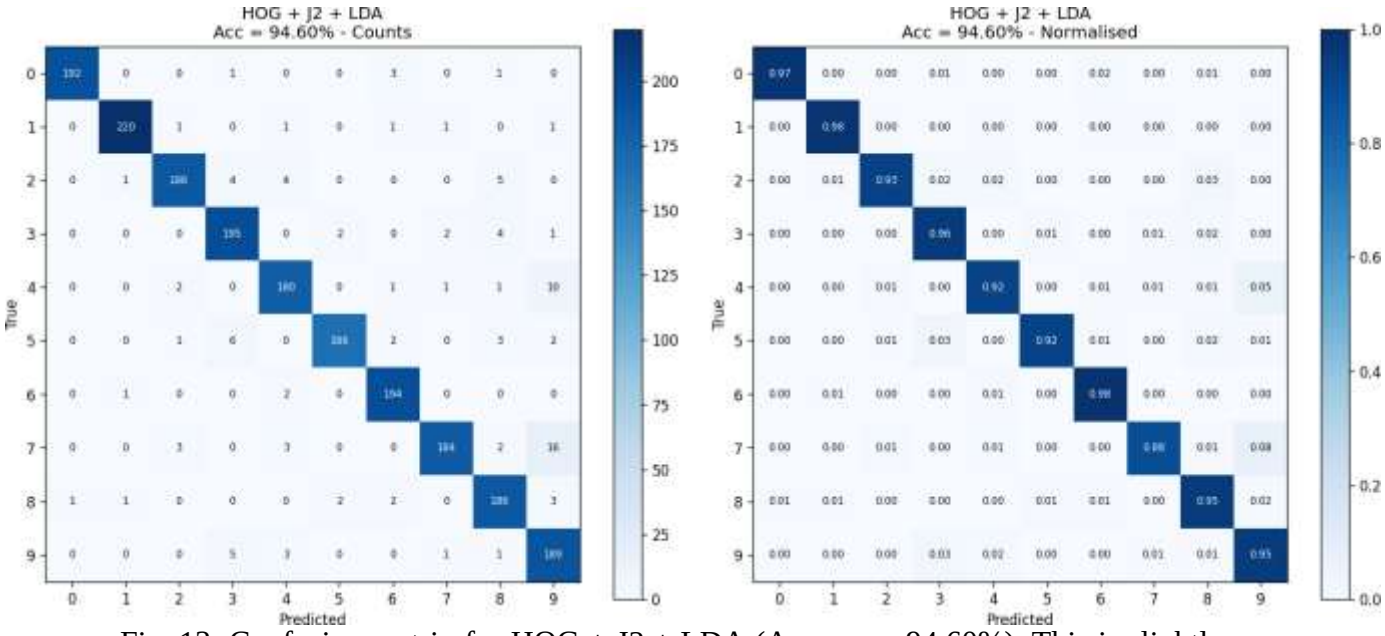


Fig. 13: Confusion matrix for HOG + J2 + LDA (Accuracy: 94.60%). This is slightly worse than HOG+SVM, suggesting that a non-linear RBF kernel has a minor benefit over a linear boundary with LDA. Digit 9 has lowest recall (around 0.86).

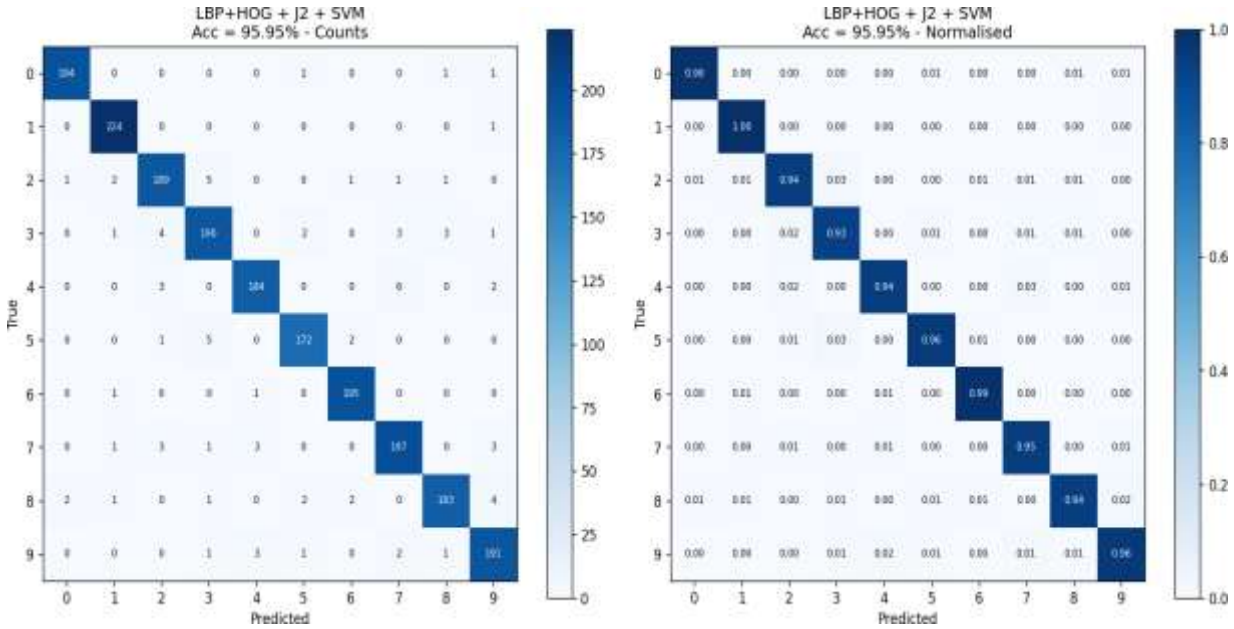


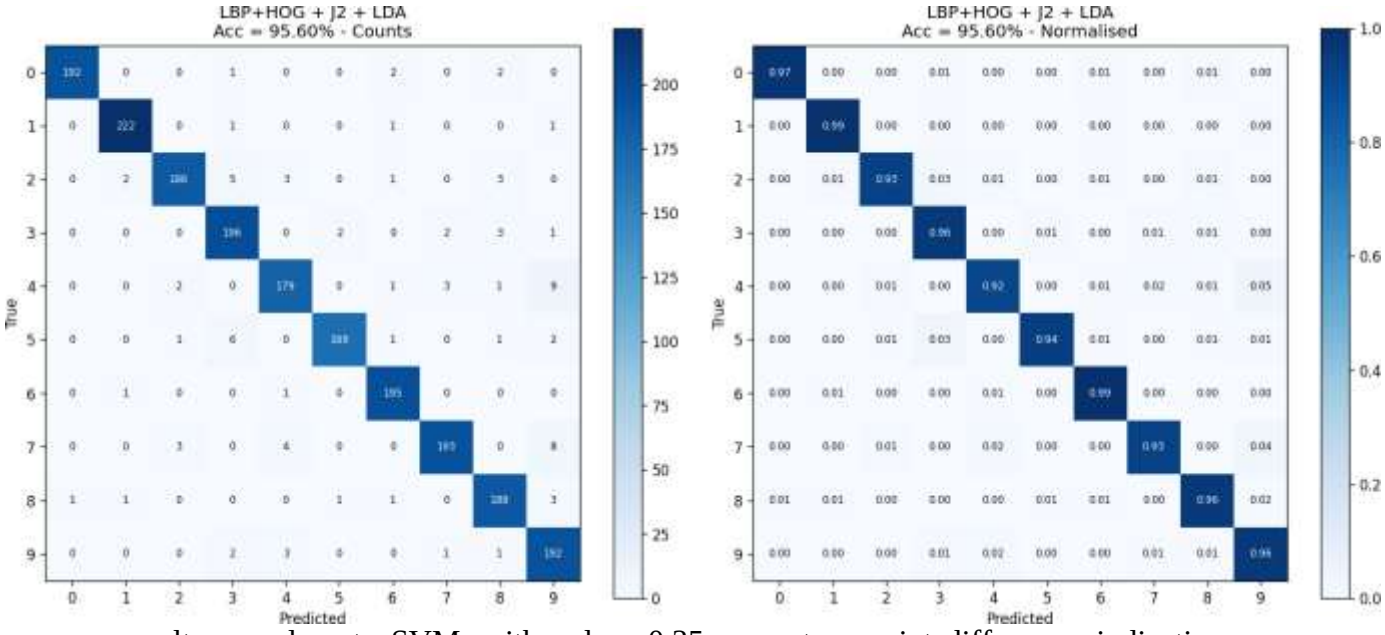
Fig. 14: Confusion matrix for LBP+HOG + J2 + SVM (Accuracy: 95.95%) – the best performing method overall. Perfect recall (1.00) is obtained by digit 1. The most remaining confusion (recall 0.93–0.94) is between digits 2 and 3, which are the most similar in structure.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

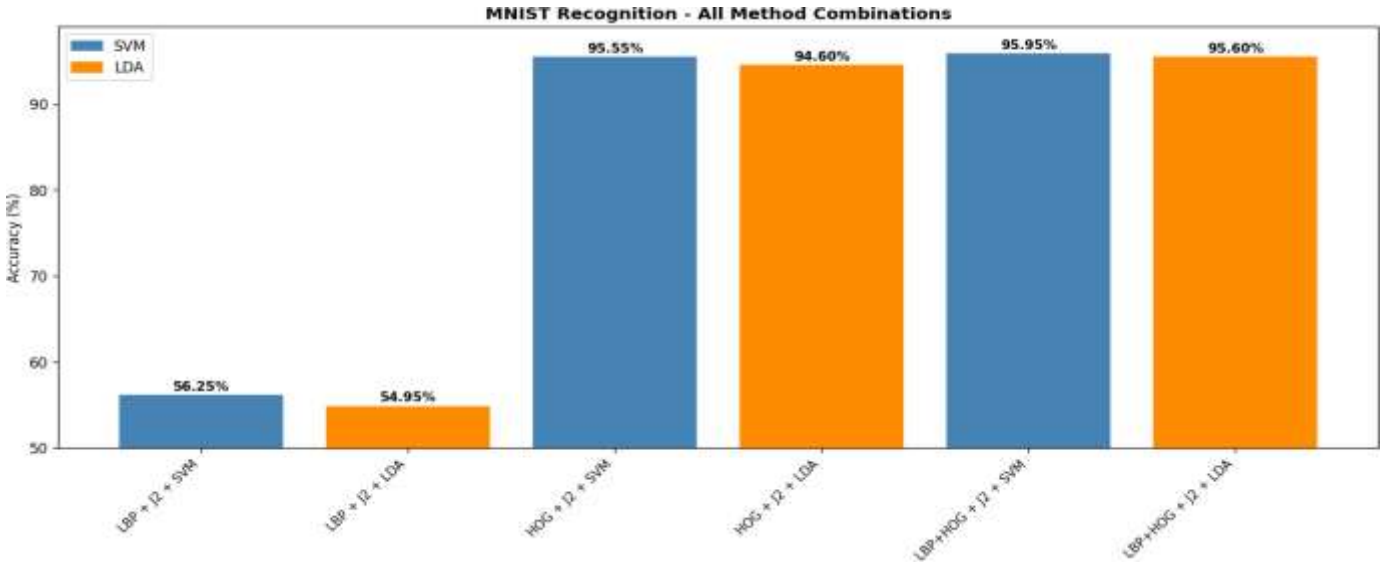
Fig. 15: Confusion matrix for LBP+HOG + J2 + LDA (Accuracy: 95.60%). Gives a



result very close to SVM, with only a 0.35 percentage-point difference, indicating that after J2 compression, the data is almost linearly separable.

TABLE III: Accuracy Rankings Across All Six Feature-Classifier Combinations

Rank	Method	Accuracy
1	LBP+HOG + J2 + SVM	95.95%
2	LBP+HOG + J2 + LDA	95.60%
3	HOG + J2 + SVM	95.55%
4	HOG + J2 + LDA	94.60%
5	LBP + J2 + SVM	56.25%
6	LBP + J2 + LDA	54.95%



Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Fig. 16: Accuracy comparison bar chart for all the 6 methods combination (blue = SVM and orange = LDA). The results of LBP-alone methods (left two bars) are much lower than all the HOG-based methods, affirming that the gradient-based HOG features are much more discriminative on digit shapes. In both types of features, SVM always gets the highest accuracy. combination of HOG and LBP offers a small but steady improvement. For each feature type, SVM performs better than LDA in all three cases, although the differences are becoming smaller as feature quality improves.

Per-Digit Precision and Recall Analysis

The per-digit precision and recall scores for all six combinations of methods when applied to each of the ten-digit classes are shown in Fig.17. There are several recurring themes among methods. In the LBP based configurations, scores per digit vary significantly with many of the classes having a precision and recall below 0.50. In LBP methods, digit 1 always gets the highest scores, and digits 2, 3, 4, 5, 6, and 8 have very low recall scores; this means that the representation of classes by the 18-D LBP histogram is not well separated for most of the digit classes.

Per digit scores are consistently good across all classes for both HOG and the combined LBP+HOG configuration. Across all classes, digit 1 remains the one known with the highest precision and recall, nearly 1.00. The digits 2 and 3 consistently give the lowest score among the more successful methods. This is due to their structural similarity in curvature direction and the presence of partial loops. Since it is a double loop form, digit 8 can be confused with digits 3 and 9, but not as much as the other two. The most well-balanced per-digit performance profile is obtained with the combined LBP+HOG feature set and SVM, in which all classes achieve a recall greater than 0.93.

Results Summary

The important performance figures summarized for the six method combinations are presented in Fig.18 and in the table below: Besides the classification accuracy, the table shows the original feature dimensionality prior to J2 compression, the compressed feature dimensionality after projection, the J2 criterion (that is, the discriminative quality of the feature space) and the training and inference time of each classifier.

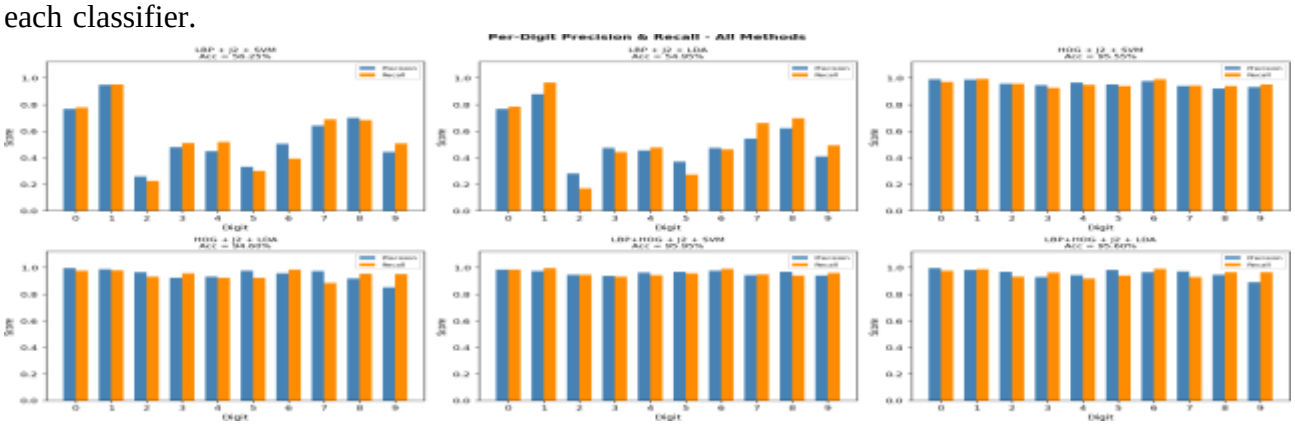


Fig. 17: All six methods per-digit precision and recall (blue = precision, orange =

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

recall). The per-digit variability for LBP methods is very high with many digits being less than 0.5. HOG method and LBP+HOG method display a stable score (>0.90) for all the digits. In all methods, the recognition of digit 1 is the easiest as it has a unique simple vertical stroke. The numbers 2 and 3 are the most difficult.

MNIST Digit Recognition - Results Summary Table

Method	Feature Dim (original)	Feature Dim (after J2)	J2 Criterion	Accuracy (%)	Train (s)	Test (s)
LBP+HOG + J2 + SVM	342	9	26.20	95.95	0.267	0.0981
LBP+HOG + J2 + LDA	342	9	26.20	95.60	0.009	0.0005
HOG + J2 + SVM	324	9	23.92	95.55	0.302	0.1145
HOG + J2 + LDA	324	9	23.92	94.60	0.026	0.0005
LBP + J2 + SVM	18	9	4.01	56.25	6.107	1.5033
LBP + J2 + LDA	18	9	4.01	54.95	0.102	0.0005

Fig. 18: Results Summary table. On the other hand, all six methods end up with the same 9-dimensional J2 subspace. However, LBP+HOG + J2 + SVM gets the best accuracy (95.95%) by training only 0.267 s which shows the computational efficiency of the J2 compression. The training time of LDA is small (less than 0.01 s) in all methods.

There are several interesting observations. First, all six methods are compressed to the same dimensionality (in this case 9-dimensional) J2 subspace, irrespective of the original feature dimensionality, thus showing the dimensionality-agnostic nature of the J2 criterion for a given number of classes. Second, the J2 criterion value is a good indicator of future classification accuracy: the higher the J2 value, the higher the recognition accuracy across all tested configurations. Third, LDA trains in less than 0.03s for all feature types, which is roughly 1 to 2 orders of magnitude faster than SVM with comparable or lower accuracy on HOG and combined features. Fourth, the training time for the LBP + J2 + SVM combination is unexpectedly long, compared to other SVM combinations, 6.107 seconds, indicating that the RBF kernel has more trouble than others in fitting decision boundaries in a poorly separated feature space.

Discussion

The experimental results shown in Section 4 exhibit distinct and meaningful trends regarding the impact of individual pipeline components on total recognition performance This section investigates these patterns across four aspects: feature type, compression behavior in J2, classifier choice, and per-digit recognition characteristics.

Effect of Feature Type

The most striking result is the gap between the performance of methods using only LBP and those using HOG, which reach around 55% and 95%, respectively. This gap is greater than 38 percentage points and is not due to the classifier or the compression stage, because all methods are reduced to the same 9-dimensional J2

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

subspace before classification. This is because the information carried by each descriptor is of a different type. LBP summarizes local micro-texture patterns into a single 18-dimensional histogram, without considering spatial arrangement information. The main reason for this is that the classes of handwritten digits are characterized mainly based on their global stroke geometry and not by their local texture statistics: two digits with completely different shapes can yield the same LBP histograms just by chance when the local neighborhood distributions coincide [17]. Therefore, the J2 scatter matrices calculated using LBP
features have a low between-class-to-within-class scatter ratio, with a low J2 criterion value of 4.01.

In contrast, HOG codes the spatial layout of gradient orientations in a structured grid of cells, which represents the layout of the digit's strokes in the direction of the gradient [7]. This structural sensitivity leads to a J2 value of 23.92, about 6 times that of LBP, and to the corresponding classification accuracy of 95.55

The highest J2 value of 26.20 and the best accuracy of 95.95% are obtained with the combined LBP+HOG descriptor, indicating that LBP and HOG contain partially non-overlapping information. While LBP alone is not sufficient to ensure digit recognizability, the texture information it encodes is a small but consistent complement that can be added to HOG features.

J2 Criterion Analysis

A particular interest in the results is the constant relationship between the J2 criterion value and classification accuracy across all three feature configurations. Comparing the scatter plots in Fig.7, 8, and 9, higher J2 values are directly related to better separation of class clusters in the compressed space. This relation verifies the theoretical basis of the J2 criterion, namely that if one maximizes the trace of $S^{-1}S_b$, one obtains a subspace in which inter-class distances are larger than intra-class distances, allowing for good classification results [10], [3].

The compression ratio that J2 can achieve is significant. The combined LBP+HOG descriptor reduces 342 dimensions to 9 discriminant components, which represents a dimensionality reduction factor of 38, while maintaining the same classification rate as if the dimensionality reduction were not done. The eigenvalue spectra in Fig.4, 5, and 6 reveal that the discriminant energy is spread across all 9 components of the eigenvalue spectrum and does not reside in a few components, indicating that all available discriminant directions are useful for separating the classes. The PCA preprocessing step, performed before calculating the scatter matrix, is crucial for ensuring numerical stability, especially with the 324-dimensional HOG features; without it, S_w would be rank-deficient.

SVM vs. LDA Comparison

In every feature configuration, SVM achieves superior results compared to LDA, but the gap is highly variable across configurations, reflecting the quality of the features. For LBPs, the accuracy gap between SVM and LDA is 1.30 percentage points (56.25% vs 54.95%), indicating that both classifiers face greater difficulty in a poorly separated feature space. For HOG features, the difference decreases to 0.95 percentage points (95.55% vs. 94.60%), and for LBP+HOG features, it further decreases to 0.35

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 4, Issue 3 (2026)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

percentage points (95.95% vs. 95.60%).

The fact that this gap is progressively reduced as feature quality increases indicates that the J2-compressed HOG and combined feature spaces are near-linearly separable. Hence, the added value of the non-linear RBF kernel in the SVM over the linear decision boundary for LDA is marginal. The practical significance of this observation is that LDA completes training in less than 0.01 seconds across all configurations, while SVM requires 0.267-0.302 seconds and provides inference time that is almost three orders of magnitude faster. If training speed, inference latency, or model transparency is prioritized in a deployment scenario, then LDA applied to the combined features with J2 compression can achieve 95.60% accuracy at a negligible computational cost compared to SVM.

It is worth noting that the SVM training time is quite long (6.107 seconds) for the LBP configuration. This is because the LBP space is poorly separated, and the SVM solver must perform many iterations to obtain support vectors that cover the overlapping regions between the classes. In contrast, the HOG and combined spaces are well separated, and the SVM solver only needs a few iterations.

Digit-Level Analysis

Fig.17 presents per-digit precision and recall for all method configurations, indicating uniform recognition behavior across most digits. Digit 1 emerges as the most reliably recognized class across all configurations, with recall values at or near 1.00 when using HOG and combined methods, and the highest absolute scores even in the LBP-only case. This superior performance can be attributed to the single-stroke vertical construction of the digit, which is both geometrically simple and produces a very compact, consistent gradient orientation profile with minimal overlap with other digit classes. Digits 2 and 3 constitute the most frequently misclassified pair among all high-performing method configurations, exhibiting the lowest recall values of approximately 0.93 and 0.94 respectively in the best-performing method (LBP + HOG + J2 + SVM). The classification confusion between these digits stems from their partial loop and curved stroke geometry, which create similar distributions of HOG gradients, particularly in handwritten samples produced under time pressure or with unconventional stroke ordering.

Digit 4 is occasionally confused with digit 9, and digit 7 with digit 2, both visually plausible confusions that human readers might also make when presented with degraded stimuli. Digit 8 demonstrates moderate susceptibility to confusion with digits 3 and 9 due to its double-loop shape, which generates gradient patterns resembling the top loops of 3 and 9. The performance profile per digit in the LBP-based configurations exhibits considerable irregularity, with some classes demonstrating reasonable individual precision while others exhibit collapsed recall-precision values. This irregularity is caused by the lack of well-defined cluster structure in the LBP J2 feature space, as illustrated in Fig.7, where different digit classes overlap significantly.

Conclusion and Future Work

The authors introduced a three-stage complete machine vision pipeline for MNIST handwritten digit recognition, including feature extraction using the LBP and HOG

Online ISSN 3007-3197
Print ISSN 3007-3189

descriptors, supervised dimensionality reduction based on the J2 criterion (implemented from scratch using NumPy), and classification using SVM and LDA. A systematic evaluation of six method combinations was undertaken on a set of 2000 test samples from the MNIST benchmark.

The results confirmed that the primary factor determining recognition rates is feature quality. The accuracies of 56.25% and 54.95% obtained using LBP alone with SVM and LDA, respectively, were attributed to the fact that neither could capture the spatial geometry of the strokes, which are the main distinguishing features of digit classes. Under the same compression and classification framework, the accuracy of HOG is significantly improved to 95.55% and 94.60%, respectively. It was observed that the two descriptors provide complementary information, as the best test accuracy with SVM was obtained with the combined LBP+HOG configuration, which was 95.95%. The J2 criterion was shown to be effective across all configurations, reducing feature vectors of up to 342 dimensions to 9 discriminant components while retaining class-related structures. It was observed that the J2 criterion value is a good indicator of downstream accuracy, with 4.01, 23.92, and 26.20 for LBP, HOG, and combined features, respectively, directly corresponding to the aforementioned performances. The SVM results show that it consistently outperforms LDA, with a comparatively small difference of 0.35 percentage points in the best feature configuration, suggesting near-linear separability in the J2-compressed space. These results showed that, with a classical pipeline, competitive recognition accuracy can be achieved on MNIST without deep learning.

Future work will focus on several extensions: One is to add spatial pyramids to LBP to capture the arrangement of local texture; training on the complete MNIST partition of 60,000 samples to investigate the impact of the amount of data on the quality of compression using J2; adding a third complementary descriptor, SIFT-based Bag-of-Words features; and testing the pipeline on more challenging datasets like EMNIST to test the generalizability of the proposed framework beyond the controlled MNIST setting.

References

Savita Ahlawat and Amit Choudhary. Hybrid cnn-svm classifier for handwritten digit recognition. Procedia Computer Science, 167:2554–2560, 2020.

Alejandro Baldominos, Yago Saez, and Pedro Isasi. A survey of handwritten character recognition with mnist and emnist. Applied sciences, 9(15):3169, 2019.

Peter N. Belhumeur, Joao P Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on pattern analysis and machine intelligence, 19(7):711–720, 2002.

Christopher M Bishop and Nasser M Nasrabadi. Pattern recognition and machine learning, volume 4. Springer, 2006.

Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. Emnist: Extending mnist to handwritten letters. In 2017 international joint conference on neural networks (IJCNN), pages 2921–2926. IEEE, 2017.

Online ISSN 3007-3197
Print ISSN 3007-3189

Corinna Cortes and Vladimir Vapnik. Support-vector networks. Machine learning, 20(3):273–297, 1995.

Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), volume 1, pages 886–893. Ieee, 2005.

RO Duda, PE Hart, and DG Stork. Introduction, in 'pattern classification', 2001.

Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 32(9):1627–1645, 2009.

Ronald A Fisher. The use of multiple measurements in taxonomic problems. Annals of eugenics, 7(2):179–188, 1936.

Jerome H Friedman. Regularized discriminant analysis. Journal of the American statistical association, 84(405):165–175, 1989.

Xiong Gan, Hong Lu, Guangyou Yang, and Jing Liu. Rolling bearing diagnosis based on composite multiscale weighted permutation entropy. Entropy, 20(11):821, 2018.

Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. Local binary patterns and its application to facial image analysis: a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41(6):765–781, 2011.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. nature, 521(7553):436–444, 2015.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

David G Lowe. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2):91–110, 2004.

Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7):971–987, 2002.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. the Journal of machine Learning research, 12:2825–2830, 2011.

Réjean Plamondon and Sargur N Srihari. On-line and off-line handwriting recognition: A comprehensive survey. IEEE Transactions on Pattern Analysis & Machine Intelligence, 22(01):63–84, 2000.

Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence, 1(5):206–215, 2019.