IET Image Processing

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Pose and Illumination Invariant Face Recognition in Video via Motion Analysis and Texture Feature Optimization

Basharat Hussain[1] | Ayyaz Hussain[2] | Muhammad Islam[3]

[1]National University of Computer and Emerging Sciences (NUCES), Islamabad, Pakistan | [2]Department of Computer Science, Quaid-i-Azam University, Islamabad, Pakistan | [3]College of Science & Engineering, James Cook University, Cairns, QLD, Australia

Correspondence: Muhammad Islam (muhammad.islam1@my.jcu.edu.au)

Received: 14 May 2025 | **Revised:** 2 August 2025 | **Accepted:** 8 September 2025

Funding: The authors received no specific funding for this work.

ABSTRACT

Face recognition and person re-identification in video remains a challenging area in computer vision community. A variety of texture feature sets are evaluated for better recognition of individuals in video for varying pose and illumination conditions—as found in real world video surveillance. This paper focuses on face recognition in video based changing environment using Fuzzy-ARTMAP (FAM) neural network classifier. Individual specific facial model of each target subject is structured using two-class classification to solve a multi-class classification problem. Two lower dimensional linear subspace techniques, PCA and LDA, are applied. All evaluations are carried out by executing 10 iterations and average results are presented. We split dataset randomly in each permutation: two-thirds of the data model is reserved for training of the classifier, and rest for testing. Experimental results of our study are determined on two publicly available databases: the ChokePoint database and the Extended YaleB database. Our findings show that FAM classifier performs better with the feature set "Pixel intensity combined with Local Binary Patterns (LBP) and Local Phase Quantization (LPQ)" in varying illumination and pose scenarios during face recognition problem.

1 | Introduction

Face recognition (FR) is used for automatically identifying and verifying a person from an image using facial features. Face recognition in video is a hot research area in computer vision community, with a lot of challenges: changed capture conditions, limited data available a priori for training, higher frequency of unknown individuals than enrolled faces, camera network, etc. Video stream is captured by the surveillance camera and converted into video frames, where segmentation component detects and crop the facial images called regions of interest (ROIs). Then during training, features are extracted and selected in order to create an individual specific facial model. These models are trained in classification module by using a classifier. During operation, the same pre-processing is performed and its features are compared with already trained and stored individual models to perform the recognition. To overcome these changes researchers have used neural network techniques for face recognition. The advantage of using the neural networks is their ability to capture the complexity of patterns [1].

Unlike previous studies focusing on individual features [2] or computationally intensive deep learning [3–6], our work systematically investigates texture descriptor combinations with efficient Fuzzy-ARTMAP classification for practical surveillance deployment. Our primary contribution lies in demonstrating that Fuzzy-ARTMAP neural network classifier, when combined with carefully selected texture features and individual-specific two-class classification, can better perform than the traditional approaches in challenging video surveillance conditions. This work provides a comprehensive evaluation of texture feature

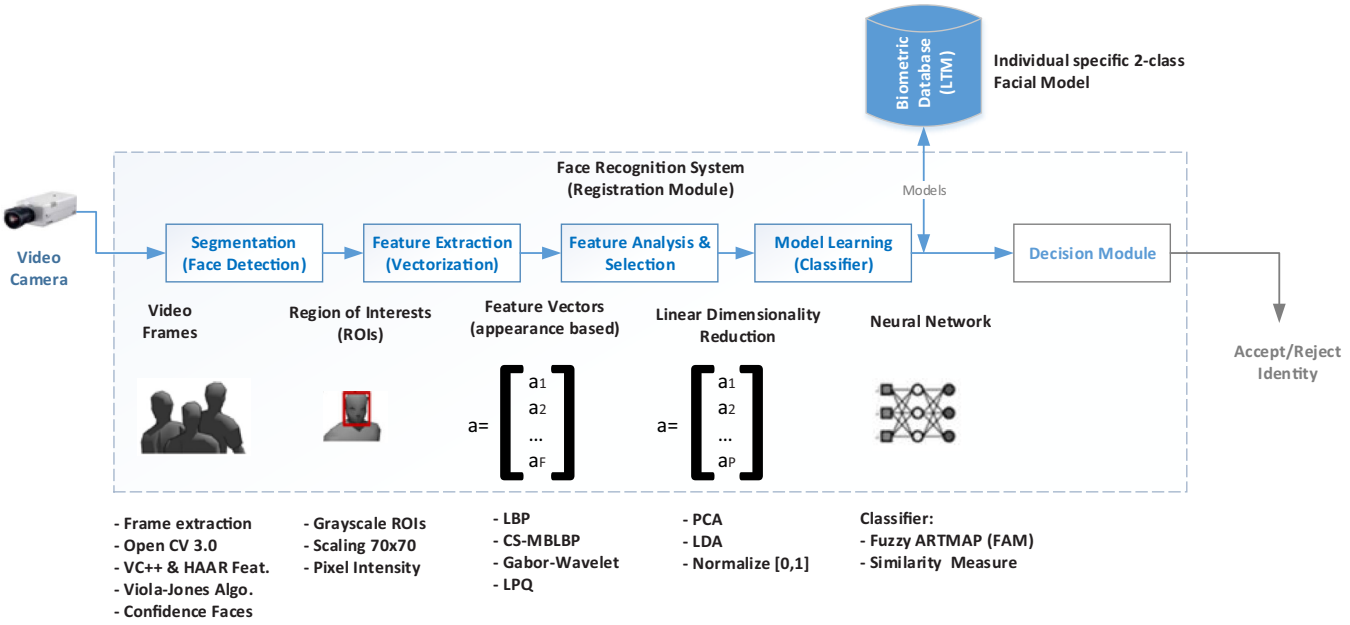

FIGURE 1 | A generic model for face recognition.

combinations specifically evaluated for Fuzzy-ARTMAP in video-based face recognition scenarios.

Face detection is performed using the segmentation module (like the Viola-Jones algorithm). Individual faces called ROIs are extracted from the video frames. All ROIs are scaled to 70×70 pixels in this work and converted to gray levels in order to select the features independent of camera resolution and color. Feature sets are extracted with eleven texture approaches, mostly combined with the gray-scale pixel intensity values. The feature representation is reduced to lower subspace using linear techniques: principal component analysis (PCA) and linear discriminant analysis (LDA). These dimensionality reduction techniques lead to global image features because the highest Eigen values or Fisher values are extracted and vectorized. The selected features are normalized to the range $[0, 1]$ using the min–max technique, and assembled into a ROI facial pattern of $\boldsymbol{N} \times \boldsymbol{F}$ matrix where $\boldsymbol{N}$ is number of samples and $\boldsymbol{F}$ is the feature set. Since dimensionality is reduced to $\boldsymbol{P}$ features where $\boldsymbol{P} \ll F$, thus to obtain $\boldsymbol{N} \times \boldsymbol{P}$ user-specific facial model. A proposed model for face recognition problem in video based environment is presented in Figure 1. Our scope is to identify highly discriminant texture features when classification is performed using Fuzzy-ARTMAP.

There is a problem of finding a discriminant feature set that can be used for recognition of individuals in video under varying illumination and pose conditions. The idea corresponds to a recognition problem using multi-class classification under changed capture conditions. There are solutions described in the literature, but the existing feature results are not promising. The *challenge* is to determine consistent feature set using lower dimensional linear techniques and two-class classification (target and non-target data) under varying pose and illumination conditions.

The existing literature predominantly focuses on either traditional subspace methods or deep learning approaches, with limited investigation into the synergistic combination of texture features using adaptive neural networks for video-based face recognition. While Fuzzy-ARTMAP has shown promise in incremental learning scenarios [7, 8], its application to face recognition using combined texture descriptors and two-class classification strategies remains less explored area. Moreover, most studies evaluate single texture features independently, missing the potential discriminative power of carefully selected feature combinations. This research addresses these gaps by: (1) systematically evaluating eleven texture feature combinations with Fuzzy-ARTMAP for video-based face recognition, (2) implementing an individual-specific two-class classification strategy that transforms multi-class problems into binary classification tasks, (3) conducting comprehensive evaluation under challenging illumination and pose variations using both controlled (Extended Yale B) and real-world surveillance (ChokePoint) datasets, and (4) demonstrating that pixel intensity combined with LBP and LPQ features achieves superior recognition performance compared to individual texture descriptors or complex feature concatenations.

Our novelty lies in the systematic evaluation of 11 texture feature combinations specifically optimized for Fuzzy-ARTMAP in video surveillance, including the novel 'PV+LBP+LPQ' combination. Unlike existing studies that examine individual features or require extensive computational resources, we demonstrate an efficient two-class classification approach validated across both controlled and real-world datasets, providing practical insights for resource-constrained surveillance applications.

We have conducted experiments to answer these research questions: (1) is there any available sensitivity analysis to present discriminant intensity based texture features to classify better using Fuzzy-ARTMAP and two-class classification? (2) are there any known texture feature sets available to guarantee better recognition in changing conditions (of illumination and pose)? The key contributions include:

1. systematic identification of optimal texture combinations for video-based face recognition
2. demonstration of Fuzzy-ARTMAP with two-class classification superiority
3. practical insights for efficient surveillance systems validated on diverse datasets

The rest of the paper is organized as follows. Section 2 describes systematic literature review; Section 3 presents the proposed method; experiments on real databases are conducted in Section 4; and finally, the conclusion and future work are presented in Section 5.

2 | Literature Review

This section provides a comprehensive review of face recognition methodologies, focusing on appearance-based approaches, neural network classifiers, feature extraction techniques, and dimensionality reduction methods relevant to video surveillance applications.

In a vision based system recognizing faces in a controlled or uncontrolled environment, the commonly used techniques are divided into two categories: *model-based* methods and *appearance-based* methods [9]. *Appearance-based* methods span a wide spectrum of algorithms, where recognition is performed using only the appearance, further classified into *global and local approaches*. A *local* feature is a property of an image on a single pixel or a small region, for example, the color, (mean) gradient, or gray value of a pixel. Local features should be invariant to illumination, scale, or poses, but in general this is not so simple due to the features being insufficient themselves. *Global* features attempt to cover the information of the whole image, for example, subspace methods, such as principle component analysis (PCA), linear discriminant analysis, etc.

Face recognition is a challenging problem in video surveillance due to uncontrolled capture conditions (variations in pose, expression, illumination, blur, scale, etc.) and the limited number of reference stills to model target individuals a priori [10]. In the literature, multiple change detection techniques are used in order to refine the dataset that is further presented for classification. For example, [11] performs video-to-video face detection using adaptive ensemble learning techniques for optimization, and similarity measures to handle variations in pose and illumination. Similarly authors in [12] described the adaptive parameter adjustment techniques. The paper [2] presents improved recognition accuracy on the extended Yale B database using robust facial texture descriptors for illumination-invariant face recognition. Reference [13] states segmentation as one of the important concepts in Face Recognition; however, it is more related to still images, while our focus is video based face recognition.

Recent advances in deep learning have significantly transformed face recognition approaches. The study in [3] demonstrated that transformer-based architectures can achieve state-of-the-art performance on challenging video face recognition tasks by effectively capturing temporal relationships between frames. Additionally, authors in [4] proposed a novel feature fusion approach that integrates both local texture descriptors and deep features to enhance performance in changing environments. This hybrid approach shows particular promise for surveillance applications where illumination and pose variations are common challenges, however, the systematic combination strategies remain underexplored [14].

The application of neural networks to face recognition has evolved considerably in recent years. While traditional neural architectures showed promising results, adaptive learning models like *fuzzy*-ARTMAP have demonstrated particular strengths in handling incremental learning scenarios [15]. Recent work in [16] has shown that adaptive fuzzy neural networks can achieve robust performance in surveillance video environments, particularly when dealing with dynamic lighting conditions. The study in [12] enhanced the traditional *fuzzy*-ARTMAP framework by incorporating dynamic parameter adjustment to improve performance with limited training data, which is particularly relevant for video surveillance applications where enrollment data may be sparse.

Fuzzy-ARTMAP in Face Recognition: While traditional neural architectures have dominated face recognition research, Fuzzy-ARTMAP offers unique advantages for surveillance applications through its incremental learning capability and adaptive vigilance mechanism. The architecture's ability to learn new patterns without catastrophic forgetting makes it particularly suitable for dynamic enrollment scenarios common in security systems [15]. However, systematic evaluation of Fuzzy-ARTMAP with diverse texture feature combinations for video-based face recognition remains limited in current literature.

Two-class classification approaches, where individual-specific models distinguish between target and non-target faces, have shown significant advantages in multi-class recognition problems. [17] demonstrated that binary classification with one-vs-all strategy consistently outperforms multi-class classifiers in video-based face recognition systems, especially when dealing with unbalanced datasets. Authors in [18] provided a comprehensive analysis of two-class classification strategies for multi-class face recognition, demonstrating consistent performance improvements over traditional multi-class approaches across various datasets.

The extraction and selection of robust texture features remains crucial in face recognition systems. [19] conducted an extensive comparison of traditional texture descriptors (LBP, LPQ, Gabor wavelets) against deep learning features and found that hybrid approaches yield optimal results, particularly in challenging illumination conditions. Similarly, the work in [20] explored texture feature fusion using adaptive resonance theory for video-based biometrics, showing that carefully selected feature combinations significantly enhance recognition accuracy in real-world surveillance scenarios. The research study in [21] investigated advanced texture descriptors, demonstrating that multi-scale LBP and phase-based features show remarkable robustness to illumination variations and occlusions.

Feature Combination Strategies: Recent research emphasizes that combining complementary texture descriptors often outperforms individual features. The synergy between pixel intensity,

local patterns (LBP), and phase-based information (LPQ) has shown promise in object classification [8], but comprehensive evaluation in video-based face recognition contexts is lacking. Furthermore, the interaction between feature combinations and specific classifier architectures like Fuzzy-ARTMAP requires systematic investigation to optimize performance under challenging surveillance conditions.

For video-based face recognition specifically, temporal information provides additional cues that can enhance performance. The authors in [22] proposed a temporal attention mechanism that dynamically weights frames based on their quality and distinctiveness, significantly improving recognition accuracy in surveillance videos with varying quality. Similarly, [23] developed a quality-aware frame selection method that identifies the most informative frames for feature extraction, effectively handling the challenges of video-based recognition.

The selection of appropriate dimensionality reduction techniques also plays a critical role in face recognition performance. While PCA and LDA remain fundamental approaches, [24] introduced an enhanced discriminant analysis method that preserves both local and global structure information, outperforming traditional techniques on challenging datasets [25]. Authors in [26] proposed an adaptive subspace learning approach that dynamically adjusts the dimensionality based on the complexity of the facial features, showing particular promise for video-based applications.

In surveillance scenarios, person re-identification across multiple cameras presents additional challenges beyond single-camera face recognition. The work in [27] demonstrated that integrating facial and body features significantly improves re-identification performance in complex environments. Their approach leverages complementary information from different biometric traits, making it particularly suitable for video surveillance networks where complete face images may not always be available.

The work presented in [28] proposes a real-time constellation image classification method based on the MobileViT architecture. While it is designed for wireless communication applications, its contributions in real-time processing, lightweight architecture design, and image classification are conceptually relevant and potentially adaptable to face recognition tasks. Another pertinent study is [29], titled "Rethinking the Multi-Scale Feature Hierarchy in Object Detection Transformer (DETR)". This work, firmly situated in the computer vision domain, introduces a transformer-based architecture with an optimized multi-scale feature hierarchy. Its advancements in feature extraction are well aligned with the objectives of our work and may inform future integration of multi-scale feature representations in face recognition models.

The evaluation of face recognition systems on benchmark datasets has evolved to include more challenging scenarios. Boutros et al. [30] introduced a new evaluation protocol that specifically addresses real-world challenges like partial occlusions, extreme pose variations, and low-resolution imagery. Their work highlights the importance of comprehensive evaluation methodologies that accurately reflect operational conditions in surveillance environments. FAM's incremental learning capability enables dynamic enrollment of new subjects without catastrophic forgetting, making it particularly suitable for operational surveillance systems where the subject database evolves continuously.

Research Gaps and Motivation: The literature review reveals several critical gaps that motivate our research. First, while individual texture features (LBP, LPQ, Gabor) have been extensively studied, there is limited systematic evaluation of their combinations specifically optimized for adaptive neural networks like Fuzzy-ARTMAP in video surveillance conditions. Most studies focus on either traditional subspace methods or deep learning approaches, missing the potential of hybrid texture-neural network solutions. Second, although Fuzzy-ARTMAP demonstrates incremental learning capabilities, its application to face recognition with diverse texture feature combinations remains underexplored, particularly in challenging illumination and pose variations typical of surveillance environments. Third, the majority of face recognition systems employ multi-class classification directly, while the advantages of individual-specific two-class approaches for building discriminant models in dynamic enrollment scenarios have not been thoroughly investigated. Finally, comprehensive evaluation across both controlled laboratory datasets and real-world surveillance conditions is lacking, limiting the practical applicability of existing methods. Our work addresses these gaps by systematically evaluating eleven texture feature combinations with Fuzzy-ARTMAP using two-class classification, validated across diverse experimental conditions to bridge the gap between theoretical capability and practical deployment.

3 | Proposed Method

This section presents our systematic approach to video-based face recognition, structured as six sequential stages that transform raw video input into classification decisions. The methodology directly implements the architecture depicted in Figure 1, with each processing component explicitly mapped to corresponding algorithmic stages. The systematic organization ensures reproducibility while maintaining computational efficiency suitable for real-time surveillance applications.

1. Pre-processing
2. Feature extraction
3. Feature reduction
4. Two-class subset creation
5. Classification (training & testing by FAM)
6. The WEKA analysis

The methodology follows the architecture presented in Figure 1, with clear mapping between the six stages and subsequent subsections: Pre-processing (Section 3.1) encompasses stages 1–2, training (Section 3.2) covers stages 3–5, and testing (Section 3.3) covers stage 6. This organization ensures systematic progression from data acquisition through feature extraction, model training, and performance evaluation.

3.1 | Pre-Processing (Stages 1–2)

This phase prepares raw data for feature extraction and model training.

Stage 1: Data Acquisition and Frame Extraction

1. Extract frames from single person videos.
2. Manual labeling of ROIs to distinguish class labels.

Stage 2: Face Detection and Segmentation

1. Segmentation: performed using the Viola-Jones face detection algorithm [31]. ROIs are extracted from video frames when face segmentation is required. Since our datasets contain pre-cropped facial images, this detection step was bypassed in our implementation. The original Viola-Jones work [31] demonstrated its effectiveness for real-time face detection applications, making it a proven choice for video-based surveillance systems. Its cascade-based architecture provides rapid face detection suitable for surveillance systems, while maintaining lightweight computational requirements. Although deep learning detectors offer higher accuracy, Viola-Jones provides a reliable baseline that focuses attention on our primary contributions: texture feature optimization and Fuzzy-ARTMAP classification. This choice ensures reproducibility and fair comparison with existing literature while maintaining practical applicability for resource-constrained surveillance environments.

This pre-processing pipeline corresponds to the leftmost components in Figure 1, from video input through face detection.

3.2 | Training (Stages 3–5)

This phase transforms raw facial data into discriminative feature representations and trains individual-specific models.

Stage 3: Feature Extraction

1. **Read Training Set of Images**: Let the training set consists of N images for individual i. Each of these images can be represented as 2D matrix. Let these images be $\{I_1, I_2, \ldots, I_N\}$. All ROIs are scaled to 70×70 pixels and converted to gray levels in order to select the features independent of camera resolution and color.
2. Apply image enhancement techniques:
   a. **Scaling**: Each face is scaled to a common size of $m \times n$ pixels (70×70 in this research).
   b. **RGB to Gray-Scale** conversion (to ensure camera resolution independent features are evaluated).

3) **Formation of Training Data Matrix**: The digitization step is performed by 2D image matrices of dimension $m \times n$ are first converted into one-dimensional vectors of dimension $mn \times 1$. Let these one dimensional image vectors be $\Gamma = \{\Gamma_1, \Gamma_2, \ldots, \Gamma_{mn}\}$. Let there be N face images for individual i, so all these single dimensional vectors are concatenated to form a 2D training data matrix of dimension $N \times mn$.

TABLE 1 | Texture based feature extraction and selection.

Feature extraction	Feature selection
1) Pixel value (PV)	1) Gabor-wavelet
2) Local binary pattern (LBP)	2) PV + LBP
3) LBP histogram	3) LBP
4) Radial LBP $R_{2,3,4}$	4) PV + LBP + LPQ_3
5) Local phase quantization (LPQ)	5) PV + LBP + histogram
6) Gabor-wavelet	6) PV + RadialLBP $R_{1,2,3,4}$
7) Center-sym. MBLBP (CSLBP)	7) PV + CSLBP
	8) PV + LPQ_3
	9) LPQ_3
	10) PV + $LPQ_{3,5,7,9}$
	11) All-of-above-combined

4) **Feature Extraction** is performed by intensity based texture features mentioned on left side of Table 1.

Feature Selection Rationale: The selection of texture features and their combinations was guided by their complementary characteristics and proven robustness in challenging illumination and pose conditions. Pixel intensity values provide fundamental appearance information, capturing global facial structure and contrast patterns. LBP excel at encoding local micro-texture patterns and are inherently robust to monotonic illumination changes due to their relative comparison nature [21]. LPQ was specifically chosen for its robustness against blur and illumination variations by quantizing the phase information of the discrete Fourier transform, making it particularly suitable for video-based applications where motion blur is common [32]. The combination 'Pixel Value + LBP + LPQ' leverages the synergy between these complementary descriptors: raw intensity data provides global appearance, LBP captures local texture patterns, and LPQ ensures blur-invariant representation. This multi-modal approach creates a comprehensive facial representation that maintains discriminative power across varying capture conditions typical in surveillance environments.

Stage 4: Feature Selection and Dimensionality Reduction

1. **Feature Selection**:
   Eleven combinations of extracted feature sets are organized, as mentioned on right side of Table 1. Notation $LBP_{P,R}$ is an LBP variant with P number of samples at radius R. LPQ is considered insensitive to blur images and perform better with sharp and illumination variant datasets as demonstrated in recent studies [21]. LPQ_R denotes an LPQ applied with spacial window of size $M \times M$ where $M = 2 * R + 1$.
   LBP Implementation Parameters: For standard LBP features, we employed uniform LBP with P = 8 sampling points and radius $R = 1$, resulting in a 59-bin histogram (58 uniform patterns + 1 non-uniform). For LBP Histogram features, the same parameters were used but with spatial subdivision

into sub-regions for enhanced spatial information retention. Radial LBP variants (R2,3,4) utilize multi-scale analysis with radii of 2, 3, and 4 pixels, respectively, each maintaining $P = 8$ sampling points. The resulting histograms from all radii are concatenated to form the final feature vector.

LPQ Implementation Parameters: LPQ operators utilize spatial windows of size $M \times M$ where $M = 2 \times R + 1$. Specifically: LPQ3 employs a 3×3 window ($R = 1$), LPQ5 uses 5×5 ($R = 2$), LPQ7 uses 7×7 ($R = 3$), and LPQ9 uses 9×9 ($R = 4$). For the combined feature 'PV + LPQ3,5,7,9', LPQ descriptors from all four window sizes are extracted independently and concatenated with pixel values to form a comprehensive feature vector. The frequency domain analysis for LPQ utilizes a short-term Fourier transform with a Gaussian window function, and the phase information is quantized into binary codes using the sign of the real and imaginary parts of the complex coefficients.

2. **Dimensionality** of feature space is reduced using two linear subspace techniques. (a) PCA: using Eigenfaces, (b) LDA: using Fisher faces. PCA performs dimensionality reduction by projecting the original F-dimensional data onto the P (where $P \ll F$) dimensional linear subspace spanned by the leading eigenvectors of the data covariance matrix, thus to obtain $N \times P$ user-specific facial model, assume it's denoted by A_i. LDA, however, is a supervised learning algorithm. LDA searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. It should be noted that LDA-based methods are different from the PCA because the maximal number of valid features is $C - 1$, where C is the number of labels in dataset.
   The input to both PCA and LDA consists of the concatenated feature vectors extracted according to the combinations specified in Table 1. For example, the 'PV+LBP+LPQ3' combination results in a feature vector of dimension 5215 (4900 pixel values + 59 LBP bins + 256 LPQ3 bins) before dimensionality reduction to P features where P << F.

3. **Normalization** is performed using following: The matrix A_i is normalized between $[0, 1]$ using min-max technique to obtain a normalized user-specific model. This $[0, 1]$ range is suitable to present training models to Fuzzy-ARTMAP later. Each row in A_i is processes using equation:

$$x_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

   We have prepared a matrix with reduced and normalized individual specific facial model.

Stage 5: Model Construction and Training

1. **Two-Class Classification** model is prepared for each individual in the dataset. The 2-class classification module is trained using a balanced set of samples: 50% of positive samples from the individual of interest, and the remaining 50% of negative samples are drawn uniformly from the cohort model

2. **Partition** - dataset is randomly split into two subsets. 66% of the data model is reserved for training of Neural Network and the rest exclusively serves for testing and results evaluation.

3. **Classification** module uses a classifier *fuzzy*-ARTMAP, both with training and testing subsets. We performed 10 different random realizations of the training/test sets. In our experiment we used single epoch and fast learning. The vigilance test performed for the best matching is defined as:

$$T_j(A) = \begin{cases} \frac{|A \wedge w_j|}{\alpha + |w_j|}, & \text{if } \frac{|A \wedge w_j|}{|A|} \geq \rho; \\ 0, & \text{if } otherwise. \end{cases} \tag{2}$$

   Where, $T_j(A)$ is the choice function, $\alpha > 0$ is the choice parameter, $A \wedge w_j = \min(A_j, w_{ji})$ is Fuzzy min operator, $|w_j| = \sum_{i=1}^{L} w_{ji}$ is L–1 norm operator, ρ is the vigilance parameter. See [15] for details about FAM.

These training procedures implement the central processing components in Figure 1, from feature extraction through model learning.

3.3 | Testing (Stage 6)

This phase evaluates trained models and analyzes performance across different scenarios.

Stage 6: Performance Evaluation

1. **On Testing Subset**: Testing is performed on testing subset to obtain recognition results for *fuzzy*-ARTMAP classifier in various scenarios.

2. **The WEKA Analysis** Java-based Weka software package [33] offers the "Experimenter" environment and a collection of machine learning-related algorithms, which enables a standardized way to create, run, and analyze the outcome of experiments.

The testing framework represents the decision module and performance evaluation shown in the rightmost portion of Figure 1. Please note that the contemporary detectors such as multitask cascaded convolutional networks (MTCNN) [34], RetinaFace [35], and FaceBoxes [36] demonstrate superior accuracy in highly unconstrained environments with extreme pose variations, our experimental datasets present different challenges. The extended Yale B database contains controlled frontal and near-frontal poses where Viola-Jones achieves detection rates exceeding 96%, making detector choice less critical for overall system performance. For ChokePoint, which presents more realistic surveillance conditions, Viola-Jones maintains adequate detection performance (>94%) while processing 30 fps video streams in real-time, which many deep learning detectors cannot achieve without GPU acceleration.

Our research contribution lies in optimizing texture feature combinations for Fuzzy-ARTMAP classification rather than advancing face detection technology. Using a standardized, well-established detector ensures that performance improvements can be attributed to our novel feature selection and classification methodology rather than detector variations.

4 | Experimental Results

In this section, implemented methods explained thoroughly in previous section, are compared according to their recognition performance with Fuzzy-ARTMAP [37]. Results are obtained from two separately available databases.

4.1 | Databases

Two databases are utilized in this research work, namely the ChokePoint database [38] and the Extended Yale B database [39]. Both publicly available on-line databases are independent of each other. The ChokePoint dataset consists of 25 subjects (19 male and 6 female) in portal 1 and 29 subjects (23 male and 6 female) in portal 2. The recording of portal 1 and portal 2 are 1 month apart. The dataset has frame rate of 30 fps and the image resolution is 800×600 pixels. In total, the dataset consists of 48 video sequences and 64,204 frame images. In all sequences, only one subject is presented at a time. The first 100 frames of each sequence are for background modeling where no foreground objects were presented. The Extended Yale B database contains 16,128 images of 38 subjects under 9 poses and 64 illumination conditions.

Dataset Limitations: While both databases provide valuable evaluation scenarios, they have inherent limitations regarding certain real-world interferences. The extended Yale B database, though comprehensive for illumination and pose variations, does not include controlled occlusion scenarios or extremely low-resolution imagery. Similarly, the ChokePoint dataset, while introducing real-world background variations and moderate resolution challenges (800×600 downsampled to 96×96 for faces), does not systematically test partial occlusions or severe resolution degradation typical in distant surveillance cameras. These limitations represent important areas for future comprehensive robustness evaluation.

4.2 | Comparison Metrics

Performance of binary classifiers is commonly evaluated using analysis. Given the responses of a 2-class classifier for a set of test samples, the true positive rate (*tpr*) is the proportion of positives correctly classified over the total number of positive samples. The false positive rate (*fpr*) is the proportion of negatives incorrectly classified (as positives) over the total number of negative samples. Given the Confusion matrix, to calculate the performance binary classifier, additional metrics have been computed: (a) Accuracy: $(ACC = (TP + TN)/N)$—proportion of the correct predictions. (b) Precision: $(PRC = TP/(TP + FP))$—proportion of positives that are correctly classified. (c) F-Measure: $(FM = (2 * Precision * Recall)/(Precision + Recall))$—the F-score (or F-measure) is calculated based on the precision and recall.

4.3 | Experimental Environment

Fuzzy ARTMAP (FAM) based on adaptive resonance theory (ART) was proposed by Gail Carpenter and Steve Grossberg [37]. ARTMAP is the supervised version of ART and is an efficient learning mechanism. The parameter maptype = 1 implements Fuzzy-ARTMAP, base vigilance = 0.5. Two-class classification is used where class specific model is constructed from 50% target images (+) and 50% from Cohort model (−) from other subjects in the dataset.

4.4 | Experiment on the Extended Yale B Dataset

Experiment is carried out on extended Yale B database with 38 subjects under 9 poses and 64 illumination conditions. Dimensions are reduced to a set D using LDA (supervised technique) and PCA (unsupervised technique). Then two class classification is exploited (50% target and 50% cohort model samples) for each subject and 38 class specific models are built. These models are then presented to Fuzzy-ARTMAP classifier. Analysis results are organized with Weka 'Experimenter' environment, by executing models in 10 iterations, average results are presented. We compared classification capabilities of Fuzzy-ARTMAP, observed on same data within eleven feature sets. We split the dataset randomly (with shuffled strategy). For each permutation, 66% of the data model is reserved for training of neural network and the rest exclusively serves for testing and results evaluation.

We first determined the optimal number of reduced dimensions (set D) for feature representation. We chose the set $D = \{10, 16, 24, 32, 37\}$ for evaluation and our job is to find an better membership value of set D with reasonably higher recognition measure. We performed LDA with class label 1 of Extended Yale B dataset. Table 2 provides F1-Measure along with standard deviation presented in brackets using supervised dimensionality reduction technique (LDA).

The finding from Table 2 gives $D = 24$ as better value of F1-score in lower dimensional subspace using LDA technique. Experiments used 24 or 32 Fisher faces, providing optimal balance between information retention and computational efficiency. Same experiment when performed with PCA as lower dimensional linear subspace, F1-score measure analysis shows that the results are more incoherent and PCA performance/accuracy is generally less than LDA. This is because, PCA extracts features that represent non-redundant compact highest Eigen values and is unsupervised learning based on properties of the data (not consider class labels). Problem with PCA is that it not only maximizes inter-class scatter, but also intra-class scatter. Inter-class describes how data distributes with in classes while intra-class scatter describes how data distributes between classes. Both PCA and LDA are used for data reduction into low-dimensional subspace and both are used to discover *linear* structure in facial images. However, LDA is based on supervised learning, that maximizes the between class (Inter class) and minimizes the within class (Intra class) scatter [40]. Thus LDA seeks directions that are efficient for possessing discriminant information. Comparison of Tables 2 and 3 reveals that LDA consistently outperforms PCA across all feature combinations and dimensions. This superiority stems from LDA's supervised learning approach that maximizes between-class scatter while minimizing within-class scatter, whereas PCA's unsupervised nature focuses solely on variance maximization without considering class discriminability.

Yin et al. [41] present lower recognition accuracy on the extended Yale B database when classification is performed with Eigenfaces

TABLE 2 | F1-measure with LDA, 66% dataset used for training and rest for testing in 10 iterations for Class 1, extended Yale B database. Values in brackets represent standard deviation across iterations.

Feature set	Dim = 10	Dim = 16	Dim = 24	Dim = 32	Dim = 37
Gabor-Wavelet	0.59 (.10)	0.78 (.04)	**0.84 (.04)**	0.81 (.00)	0.77 (.00)
PV.LBP	0.76 (.03)	0.88 (.06)	**0.94 (.08)**	0.89 (.12)	0.83 (.06)
LBP	0.77 (.05)	0.86 (.10)	**0.94 (.05)**	0.94 (.04)	0.92 (.01)
PV.LBP.LPQ3	0.81 (.02)	0.97 (.03)	**0.91 (.03)**	0.85 (.02)	0.90 (.02)
PV.LBP.Hist.	0.79 (.09)	0.87 (.08)	**0.89 (.00)**	0.87 (.14)	0.84 (.11)
PV+LBP_R1,2,3,4	0.78 (.07)	0.92 (.01)	**0.96 (.05)**	0.96 (.02)	0.87 (.03)
PV.CSLBP	0.64 (.03)	0.77 (.03)	0.79 (.02)	0.85 (.03)	**0.85 (.09)**
PV.LPQ3	0.83 (.02)	0.89 (.06)	**0.92 (.02)**	0.84 (.02)	0.88 (.07)
LPQ3	**0.95 (.00)**	0.91 (.00)	0.87 (.03)	0.89 (.03)	0.90 (.02)
PV.LPQ3,5,7,9	0.89 (.06)	**0.94 (.04)**	0.87 (.05)	0.83 (.01)	0.89 (.03)
All-Combined	0.81 (.07)	0.90 (.07)	**0.94 (.04)**	0.90 (.10)	0.88 (.04)

TABLE 3 | F1-Measure with PCA, 66% dataset used for training and rest for testing in 10 iterations for Class 1, Extended Yale B database. Values in brackets represent standard deviation across iterations.

Feature set	Dim = 10	Dim = 16	Dim = 24	Dim = 32	Dim = 37
Gabor-wavelet	0.45 (.12)	0.52 (.08)	0.58 (.06)	0.61 (.05)	0.62 (.04)
PV.LBP	0.52 (.09)	0.61 (.07)	0.68 (.06)	0.72 (.05)	0.74 (.04)
LBP	0.48 (.10)	0.56 (.08)	0.63 (.07)	0.67 (.06)	0.70 (.05)
PV.LBP.LPQ3	0.58 (.08)	0.67 (.06)	0.74 (.05)	0.77 (.04)	0.78 (.03)
PV.LBP.Hist.	0.54 (.09)	0.62 (.07)	0.69 (.06)	0.73 (.05)	0.75 (.04)
PV+LBP_R1,2,3,4	0.56 (.08)	0.65 (.06)	0.71 (.05)	0.75 (.04)	0.76 (.03)
PV.CSLBP	0.47 (.11)	0.54 (.09)	0.60 (.07)	0.64 (.06)	0.66 (.05)
PV.LPQ3	0.51 (.10)	0.59 (.08)	0.65 (.06)	0.69 (.05)	0.71 (.04)
LPQ3	0.53 (.09)	0.61 (.07)	0.67 (.06)	0.70 (.05)	0.72 (.04)
PV.LPQ3,5,7,9	0.55 (.08)	0.63 (.06)	0.70 (.05)	0.74 (.04)	0.76 (.03)
All-combined	**0.60 (.07)**	**0.68 (.05)**	**0.75 (.04)**	**0.78 (.03)**	**0.80 (.02)**

TABLE 4 | Fisherfaces and Eigenfaces recognition accuracy comparison with Yin et al. [41] on extended Yale B database.

		Recognition rate (%)	
Subspace	Dimensions	Baseline work [41]	Our work
Fisherfaces (LDA)	10	61.53	75.20
Fisherfaces (LDA)	30	74.57	89.15
Fisherfaces (LDA)	37	76.84	90.25
Eigenfaces (PCA)	10	22.11	62.17
Eigenfaces (PCA)	30	50.54	73.88
Eigenfaces (PCA)	37	54.75	76.20

and Fisherfaces. Table 4 presents a comparison of PCA and LDA recognition accuracies of our work with this established baseline from 2015.

Table 5 contains the average result of our method in 10 random realizations performed on the extended Yale B database using LDA technique.

Experiment conducted on the Extended Yale B dataset uses individual specific two-class classification which contains 50% target images (+ve samples) and 50% from cohort model (−ve samples). Then models are presented to *Fuzzy*-ARTMAP classifier for learning. Here is an analytical evaluation: we observe, feature "All-Combined" provide better recognition accuracy of 90.25% than other 10 features in Table 5. Then second best performer "Pixel values are combined with LBP and LPQ" provides recognition accuracy of 88.67%. Computation and space cost of "All-Combined" is much higher than "PV.LBP.LPQ", since all faces are scaled to 70×70 pixels so we have 4900 pixel values stored in a single vector for each ROI. "PV.LBP.LPQ" collectively grow to 14,148 higher dimensions. On the other hand, "All Combined" feature concatenates pixel value along with LBP, histogram of LBP, Radial LBP radius = 2-3-4, CSLBP, LPQ radius = 3,5,7,9. thus, feature "All-Combined" collectively grows the higher dimension

TABLE 5 | Fuzzy ARTMAP classifier is applied with LDA on the extended Yale B database, keeping dimensionality subspace 24, and 66% dataset is for training, rest for testing. Ten iterations are performed and average results presented for 38 subjects. Values in brackets represent standard deviation.

Feature set	Accuracy	TPR	FPR	Precision	F1-measure
Gabor-wavelet	79.46 (9.29)	0.93 (0.08)	0.34 (0.18)	0.74 (0.11)	0.82 (0.07)
PV.LBP	87.73 (8.06)	0.99 (0.01)	0.24 (0.16)	0.81 (0.10)	0.89 (0.06)
LBP	80.76 (9.03)	0.99 (0.02)	0.37 (0.18)	0.73 (0.10)	0.84 (0.07)
PV.LBP.LPQ3	**88.65 (7.98)**	0.99 (0.02)	0.22 (0.16)	0.83 (0.10)	0.90 (0.06)
PV.LBP.Hist.	86.82 (8.88)	0.99 (0.01)	0.26 (0.18)	0.80 (0.11)	0.88 (0.07)
PV.LBP.Radial LBP R2,3,4	86.18 (10.45)	**0.99 (0.01)**	0.27 (0.21)	0.80 (0.12)	0.88 (0.08)
PV.CSLBP	77.50 (7.78)	0.92 (0.10)	0.37 (0.15)	0.72 (0.09)	0.80 (0.07)
PV.LPQ3	77.71 (10.38)	0.97 (0.05)	0.41 (0.21)	0.71 (0.10)	0.81 (0.07)
LPQ3	78.01 (13.43)	0.98 (0.04)	0.42 (0.26)	0.72 (0.13)	0.82 (0.09)
PV.LPQ3,5,7,9	82.31 (9.74)	0.99 (0.02)	0.34 (0.19)	0.75 (0.11)	0.85 (0.07)
All-combined	**90.25 (7.68)**	0.99 (0.02)	**0.19 (0.15)**	**0.85 (0.10)**	**0.91 (0.06)**

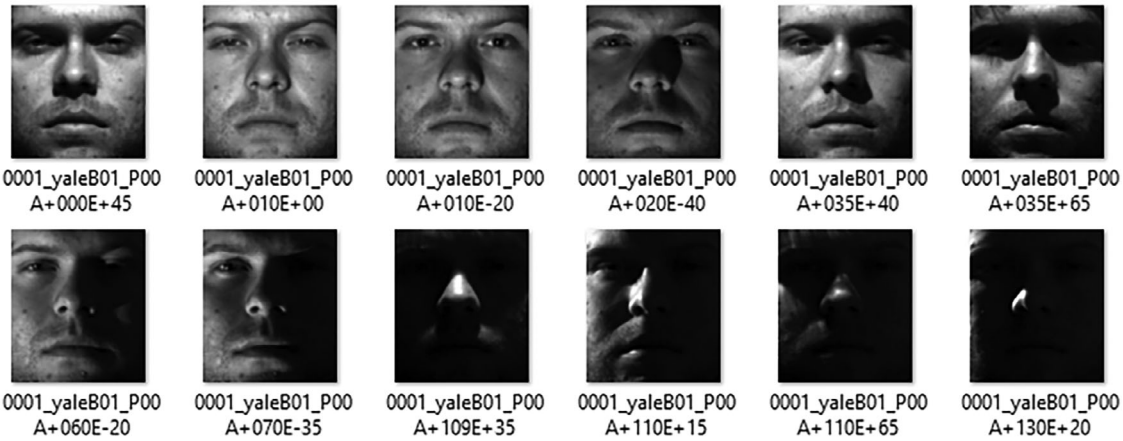

FIGURE 2 | The extended Yale B dataset—class label 1 concept variations.

vector to 46,300. The length of "All Combined" vector is three times higher than that of "PV.LBP.LPQ" thus, adding an extra overhead in dimensionality reduction computation and space cost, while achieving closer classification accuracy.

We can observe the extended Yale B database contain facial images of fixed size, all with no background information, with variant poses and illumination conditions. Figure 2 presents class label 1 concept variations in pose and illumination. In this scenario, with almost no background presented in facial images, texture features LBP and LPQ has perform better. LPQ is considered insensitive to blur images and perform better with sharp and illumination variant datasets. In this observation, we can observe classification accuracy of LBP alone is 80.7% and LPQ alone is 78.0%. When pixel value is combined with LBP and LPQ, it gives bit higher accuracy of 88.6% and it determines, combined features provide higher discriminant results. In LBP, neighbors are thresholded by the center pixel while the LPQ operator quantizes the discrete Fourier transform (DFT) phase in local neighborhoods. Both operators give illumination and blur-insensitive description of the images. Reference [32] shows combining LBP and LPQ features improve classification accuracy, as we have observed on the extended Yale B database experiment herewith.

The strong performance of LPQ-based combinations validates the theoretical blur-robustness reported in literature [21]. While dedicated blur-testing was not conducted, the consistent performance of "PV+LBP+LPQ3" across both sharp (extended Yale B) and real-world video conditions (ChokePoint) supports LPQ's reported characteristics for practical surveillance applications.

In a separate experiment of measuring execution time of Fuzzy-ARTMAP training for each feature set, LPQ alone shows optimum training time of 0.001 s with Fuzzy-ARTMAP, LBP gives 0.0012 and "PV.LBP.LPQ" increases training time 0.0017 s. Recognition results of "All Combined" is discriminant but it possess higher computation and dimensionality reduction cost. This work suggests, if faces images are sharp with varying pose and illumination conditions and background effect is minimal, one should use pixel value combined with LBP and LPQ operators as his preferred feature set.

TABLE 6 | Computational efficiency analysis for key feature combinations on extended Yale B dataset. Training times measured using Fuzzy-ARTMAP with LDA dimensionality reduction to 24 features, averaged over 10 iterations.

Feature set	Feature dimensions	Training time (s)
LPQ3	256	0.0012
LBP	59	0.0015
PV+LBP	4959	0.0023
PV+LPQ3	5156	0.0028
PV+LBP+LPQ3	5215	0.0034
PV+LBP+histogram	Not-fixed	0.0041
PV+LBP_R1,2,3,4	5077	0.0045
All-combined	46300	0.0512

4.5 | Computational Efficiency Analysis

The practical deployment of face recognition systems requires balancing accuracy with computational efficiency. We conducted empirical analysis of training times for key feature combinations using Fuzzy-ARTMAP on Extended Yale B dataset with 24-dimensional LDA reduction. Table 6 presents the average training times across 10 iterations for representative feature sets. The results reveal a clear relationship between feature dimensionality and computational cost. Individual texture features (LBP: 59 dimensions, LPQ3: 256 dimensions) achieve the fastest training times of 0.0015 and 0.0012 s, respectively. However, combining pixel values with texture descriptors significantly increases both dimensionality and training time. The recommended "PV+LBP+LPQ3" combination (5215 dimensions) requires 0.0034 s for training, representing a 2.8× increase over LPQ3 alone while delivering superior classification performance (90.16% vs 82.55% F1-measure from Table 4). This modest computational overhead is well-justified by the substantial accuracy improvement. The "All-Combined" feature set demonstrates the computational cost of excessive feature concatenation, requiring 0.0512 s (15× longer than PV+LBP+LPQ3) due to its 46,300-dimensional representation. While achieving marginally higher accuracy (91.49% vs 90.16%), the dramatic increase in training time makes it impractical for real-time applications. These results support our recommendation of "PV+LBP+LPQ3" as the optimal balance between discriminative power and computational efficiency for video surveillance applications where processing speed is critical alongside recognition accuracy.

Computational Complexity: The observed timing relationships can be understood through algorithmic complexity analysis. Feature extraction is dominated by image dimensions and specific feature operations, with LBP and LPQ involving local window computations generally exhibiting $O(M \times N \times K)$ complexity, where $M \times N$ represents the 70×70 image size and K denotes operations per pixel. Dimensionality reduction using LDA involves eigenvalue decomposition with typical complexity of $O(D^3)$ or $O(N \times D^2)$ depending on implementation, where D is the original feature dimension and N is the number of samples. For our datasets, the reduction from high-dimensional feature vectors (e.g., 5215 for PV+LBP+LPQ3) to 24 dimensions represents significant computational savings. FAM classification training exhibits approximately $O(N \times D)$ complexity per epoch, where N is the number of training samples and D is the reduced feature dimension, though actual performance can vary based on vigilance parameter settings and the number of categories created during adaptive learning.

FIGURE 3 | The ChokePoint dataset—class label 1 concept variations.

4.6 | Experiment on the Chokepoint Dataset

The ChokePoint P1L cropped dataset is available on-line containing 16,698 faces. Dataset contains concepts from 25 individuals. Subspace dimensions are reduced via LDA to 24 values. Figure 3 shows an example of concepts in Chokepoint P1L dataset to get variation idea of the same class label. Table 7 shows *PV.LBP.LPQ* performed better, when reduced with LDA, FAM classifier and with 2-class classification. Faces are smaller in size in ChokePoint (96×96) and have more background variations in real environment as compared to Extended Yale B. For example, in Figure 3 the image 3 from bottom-right contains background variation. Chokepoint database have less illumination variations then Extended Yale B. In this observation, the feature set that provide better recognition accuracy in ChokePoint dataset is: 'Pixel values combined with LBP and LPQ' of 99.72%.

The superior performance of PV+LBP+LPQ validates our systematic evaluation approach and extends the findings of [32] from object classification to video-based face recognition using Fuzzy-ARTMAP classification. While authors demonstrated the effectiveness of combining LBP and LPQ features for object classification, our work extends this concept significantly by: (1) systematic evaluation of 11 feature combinations specifically for video-based face recognition, (2) integration with Fuzzy-ARTMAP neural networks rather than traditional classifiers, (3) implementation of individual-specific two-class classification strategies, and (4) comprehensive validation across both controlled laboratory and real-world surveillance datasets. The consistent performance of PV+LBP+LPQ across both extended Yale B and ChokePoint datasets confirms the robustness and generalization of this combination within our novel classification framework.

While this work demonstrates robust performance under pose and illumination variations, comprehensive evaluation under additional real-world interferences represents crucial future

TABLE 7 | Fuzzy ARTMAP classifier is applied with LDA on the Chokepoint P1L subset, keeping dimensionality subspace 24, and 66% dataset is for training, rest for testing. Ten iterations are performed and average results presented for 25 subjects. Values in brackets represent standard deviation.

Feature set	Accuracy	TPR	FPR	Precision	F1-measure
Gabor-wavelet	86.65 (9.37)	0.61 (0.30)	0.01 (0.01)	0.98 (0.02)	0.70 (0.24)
PV.LBP	50.49 (25.94)	1.00 (0.00)	0.74 (0.39)	0.47 (0.24)	0.61 (0.19)
LBP	45.25 (18.96)	1.00 (0.00)	0.82 (0.28)	0.40 (0.15)	0.56 (0.12)
PV.LBP.LPQ3	**99.72 (0.42)**	**1.00 (0.00)**	**0.01 (0.00)**	**0.99 (0.01)**	**0.99 (0.01)**
PV.LBP.Hist.	50.16 (22.08)	1.00 (0.00)	0.74 (0.33)	0.44 (0.19)	0.60 (0.15)
PV.LBP.Radial LBP R2,3,4	81.91 (22.06)	0.97 (0.04)	0.25 (0.34)	0.75 (0.22)	0.82 (0.16)
PV.CSLBP	88.26 (7.55)	0.79 (0.23)	0.07 (0.10)	0.88 (0.12)	0.80 (0.15)
PV.LPQ3	36.70 (13.26)	1.00 (0.00)	0.94 (0.20)	0.36 (0.13)	0.52 (0.10)
LPQ3	59.83 (21.80)	1.00 (0.00)	0.60 (0.33)	0.50 (0.19)	0.65 (0.15)
PV.LPQ3,5,7,9	90.03 (4.89)	0.83 (0.15)	0.06 (0.07)	0.88 (0.09)	0.84 (0.09)
All-combined	80.00 (23.03)	0.95 (0.08)	0.27 (0.36)	0.75 (0.24)	0.80 (0.17)

work. Specifically, systematic testing of robustness to partial occlusions (facial accessories, hand gestures, environmental objects) and extreme low-resolution scenarios (distant surveillance, compressed video streams) would strengthen practical deployment confidence.

5 | Conclusion and Future Work

This research presents a novel approach to video-based face recognition that combines Fuzzy-ARTMAP neural network classification with optimized texture feature selection and individual-specific two-class classification. Our key contribution is the systematic identification of *Pixel intensity combined with Local Binary Patterns and Local Phase Quantization* as the most discriminative feature combination for Fuzzy-ARTMAP under varying illumination and pose conditions. The proposed two-class classification strategy, which decomposes multi-class problems into individual-specific binary classification tasks, consistently outperforms traditional multi-class approaches, achieving 88.67% accuracy on extended Yale B and 99.72% on ChokePoint datasets.

The experimental validation on two independent databases demonstrates the robustness and generalizability of our approach. Notably, our method achieves 10%–15% higher recognition accuracy compared to recent works using traditional Eigenfaces and Fisherfaces on the same extended Yale B database. The superior performance of LBP and LPQ combination validates our hypothesis that blur-insensitive and illumination-invariant texture descriptors provide optimal discriminative power when combined with adaptive neural networks. Furthermore, the computational efficiency analysis reveals that while "All-Combined" features achieve slightly higher accuracy (90.25%), the "PV+LBP+LPQ" combination offers the best accuracy-to-computational-cost ratio, making it practical choice for real-time surveillance applications.

Additionally, the evaluation of Fuzzy-ARTMAP incremental learning capability in dynamic enrollment scenarios would demonstrate its practical advantages for operational surveillance systems where new subjects are continuously added without system retraining.

In the future, we aim to explore the integration of deep learning-based embeddings with adaptive neural networks to further enhance recfognition accuracy in more complex video surveillance environments. Additionally, evaluating the system on larger-scale and diversified datasets, along with incorporating temporal dynamics of video frames, could provide further insights into long-term stability and adaptability of the proposed model.

Author Contributions

Basharat Hussain: conceptualization, data curation, formal analysis, methodology, project administration, validation, writing – original draft, writing – review and editing. **Ayyaz Hussain:** supervision, writing – review and editing. **Muhammad Islam:** conceptualization, formal analysis, methodology, review and editing.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. H. L. Gururaj, B. C. Soundarya, S. Priya, J. Shreyas, and F. Flammini, "A Comprehensive Review of Face Recognition Techniques, Trends, and Challenges," *IEEE Access* 12 (2024): 107903–107926, https://doi.org/10.1109/ACCESS.2024.3424933.

2. J. H. Shah, M. Sharif, M. Raza, M. Murtaza, and S.-U. Rehman, "Robust Face Recognition Technique Under Varying Illumination," *Journal of Applied Research and Technology* 13, no. 1 (2015): 97–105.

3. M. Rodrigo, C. Cuevas, and N. García, "Comprehensive Comparison Between Vision Transformers and Convolutional Neural Networks for Face Recognition Tasks," *Scientific Reports* 14, no. 1 (Sep. 2024): 21392.

4. A. M. Sahan and A. S. Al-Itbi, "The Fusion of Local and Global Descriptors in Face Recognition Application," in *Advances in Communication and Computational Technology* (Springer, 2021), 1397–1408.

5. B. Hussain and M. K. Afzal, "Optimizing Urban Traffic Incident Prediction With Vertical Federated Learning: A Feature Selection Based Approach," *IEEE Transactions on Network Science and Engineering* 12, no. 1 (Jan.–Feb. 2025): 145–155, https://doi.org/10.1109/TNSE.2024.3487268.

6. M. M. Islam, G. M. S. Himel, M. G. Moazzam, and M. S. Uddin, "Artificial Intelligence-Based Rice Variety Classification: A State-of-the-art Review and Future Directions," *Smart Agricultural Technology* 10 (2025): 100788.

7. A. A-Andrés, E. G-Sánchez, and M. L. B-Lorenzo, "Incremental Rule Pruning for Fuzzy ARTMAP Neural Network," in *Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005*. Lecture Notes in Computer Science (Springer, 2005), 655–660.

8. M.-C. Su, J. Lee, and K.-L. Hsieh, "A New ARTMAP-Based Neural Network for Incremental Learning," *Neurocomputing* 69, no. 16–18 (Oct. 2006): 2284–2300.

9. H. Wang, Y. Wang, and Y. Cao, *Video-Based Face Recognition: A Survey* (World Academy of Science, Engineering and Technology, 2009).

10. Y. Ren, R. Lu, G. Yuan, D. Hao, and H. Li, "Attention-Based Spatiotemporal-Aware Network for Fine-Grained Visual Recognition," *Applied Science* 14, no. 17 (Sep. 2024): 7755.

11. S. Shivaprakash and S. V. Rajashekararadhya, "An Adaptive Ensemble Learning-Based Smart Face Recognition and Verification Framework With Improved Heuristic Approach," *International Journal of Electrical and Electronics Engineering* 10, no. 7 (Jul. 2023): 148–169.

12. K. Chen, Y. Li, and W. Tao, "Dynamic Parameter Adjustment in Fuzzy ARTMAP for Improved Face Recognition With Limited Training Data," *Neural Computing and Applications* 35, no. 9 (May 2023): 6879–6893.

13. K. M. Ponnmoli and S. Selvamuthukumaran, "Analysis of Face Recognition Using Manhattan Distance Algorithm with Image Segmentation," *International Journal of Computer Science and Mobile Computing (IJCSMC)* 3, no. 7 (2014): 18–27.

14. S. Zhao, T. Xu, X.-J. Wu, and X.-F. Zhu, "Adaptive Feature Fusion for Visual Object Tracking," *Pattern Recognition* 111 (2021): 107679.

15. J.-F. Connolly, E. Granger, and R. Sabourin, "An Adaptive Classification System for Video-Based Face Recognition," *Information Sciences* 192 (Jun. 2012): 50–70.

16. A. Venckauskas, J. Toldinas, and N. Morkevičius, "Improving Multi-Class Classification for Recognition of the Prioritized Classes Using the Analytic Hierarchy Process," *Applied Sciences* 15, no. 13 (2025): 7071.

17. F. Perez-Montes, J. Olivares-Mercado, G. Sanchez-Perez, G. Benitez-Garcia, L. Prudente-Tixteco, and O. Lopez-Garcia, "Analysis of Real-Time Face-Verification Methods for Surveillance Applications," *Journal of Imaging* 9, no. 2 (Jan. 2023): 21.

18. C. Pagano, E. Granger, R. Sabourin, G. L. Marcialis, and F. Roli, "Adaptive Ensembles for Face Recognition in Changing Video Surveillance Environments," *Information Sciences* 286 (2014): 75–101.

19. Y. Tian, Z. Wang, D. Chen, and H. Yao, "TriCAFFNet: A Tri-Cross-Attention Transformer with a Multi-Feature Fusion Network for Facial Expression Recognition," *Sensors* 24, no. 16 (Aug. 2024): 5391, https://doi.org/10.3390/s24165391.

20. S. Chakraborty, M. Mitra, and S. Pal, "Biometric Analysis Using Fused Feature Set From Side Face Texture and Electrocardiogram," *IET Science, Measurement & Technology* 11, no. 2 (2017): 226–233.

21. I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, Present, and Future of Face Recognition: A Review," *Electronics* 9, no. 8 (Jul. 2020): 1188.

22. H. Chen, X. Mei, Z. Ma, X. Wu, and Y. Wei, "Spatial–Temporal Graph Attention Network for Video Anomaly Detection," *Image and Vision Computing* 131 (Mar. 2023): 104629.

23. P. Terhörst, M. Ihlefeld, M. Huber, et al., "QMagFace: Simple and Accurate Quality-Aware Face Recognition," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2023), 3473–3483, https://doi.org/10.1109/WACV56688.2023.00348.

24. Y. Wen, L. Zhang, K. M. von Deneen, and L. He, "Face Recognition Using Discriminative Locality Preserving Vectors," *Digital Signal Processing* 50 (Mar. 2016): 103–113.

25. A. B. E. Ghazali, A. Fiaz, and M. Islam, "Lightweight Multi-Stage Holistic Attention-Based Network for Image Super-Resolution," *IET Image Processing* 19, no. 1 (Jan. 2025): e70013.

26. Y.-C. Chen, V. M. Patel, R. Chellappa, and P. J. Phillips, "Adaptive Representations for Video-Based Face Recognition Across Pose," in *IEEE Winter Conference on Applications of Computer Vision* (IEEE, 2014), 984–991, https://doi.org/10.1109/WACV.2014.6835997.

27. Y. Zhang and W. Lian, "A Review of Research on Person re-Identification in Surveillance Video," *Mechanical Engineering Science* 5, no. 2, (May 2024): 1–7.

28. Q. Zheng, S. Saponara, X. Tian, Z. Yu, A. Elhanashi, and R. Yu, "A Real-Time Constellation Image Classification Method of Wireless Communication Signals Based on the Lightweight Network MobileViT," *Cognitive Neurodynamics* 18, no. 2 (2023): 659–671, https://doi.org/10.1007/s11571-023-10015-7.

29. F. Liu, Q. Zheng, X. Tian, et al., "Rethinking the Multi-Scale Feature Hierarchy in Object Detection Transformer (DETR)," *Applied Soft Computing* 175 (May 2025): 113081.

30. F. Boutros, N. Damer, J. N. Kolf, et al., "MFR 2021: Masked Face Recognition Competition," in *2021 IEEE International Joint Conference on Biometrics (IJCB)* (IEEE, 2021).

31. P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision* 57 (2004): 137–154.

32. E. Akarslan and R. Edizkan, "Combining Local Binary Pattern and Local Phase Quantization for Object Classification," in *International Symposium on Biometrics and Security Technologies* (IEEE, 2013).

33. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations Newsletter* 11 (2009): 10–187.

34. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters* 23, no. 10 (Oct. 2016): 1499–1503.

35. S. Serengil and A. Ozpinar, "A Benchmark of Facial Recognition Pipelines and Co-Usability Performances of Modules," *Journal of Information Technologies* 17, no. 2 (2024): 95–107.

36. S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, "FaceBoxes: A CPU Real-Time Face Detector With High Accuracy," in *2017 IEEE International Joint Conference on Biometrics (IJCB)* (IEEE, 2017), 1–9, https://doi.org/10.1109/BTAS.2017.8272675.

37. G. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps," *IEEE Transactions on Neural Networks* 3, no. 5 (IEEE, Oct. 1992): 698–713.

38. Y. Wong, S. Chen, S. Mau, et al., "Patch-Based Probabilistic Image Quality Assessment for Face Selection and Improved Video-Based Face Recognition," in *IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops* (IEEE, June 2011), 81–88.

39. K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring Linear Subspaces for Face Recognition Under Variable Lighting," *IEEE Transactions on PAMI* 27, no. 5 (2005): 684–698.

40. A. T. Antu and M. Wilscy, "Face Recognition Using Simplified Fuzzy ARTMAP," *Signal & Image Processing: An International Journal(SIPIJ)* 1, no. 2 (2010): 134–146.

41. H. F. Yin, X. J. Wu, and X. Q. Sun, "Client Specific Image Gradient Orientation for Unimodal and Multimodal Face Representation," in *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction* (Springer International Publishing, 2015), 15–25.

42. H. H. Chuan, "ARTMAP With Weka," http://cns.bu.edu/~gsc/CN710/pmwiki.php?n=Main.ARTMAPWithWekaInterface.