Journal of Technology Research (JTR)
Volume 3, Issue 2, (2025), pp 155-163, ISSN 3005-639X

A Robust Face Recognition Framework Based on CapsNet–PCA–ELM Integration

Khadeja S. Alessawi
Department of Information Technology, College of Industrial Technology, Misrata, Libya.
*Corresponding author email: khadija.alessawi@cit.edu.ly.

Received: 29-10-2025 | Accepted: 10-12-2025 | Available online: 20-12-2025 | DOI:10.26629/jtr.2025.18

ABSTRACT

In this paper, we propose a hybrid face recognition model based on Capsule Networks (CapsNet) for feature extraction, integrating Principal Component Analysis (PCA) for dimensionality reduction and an Extreme Learning Machine (ELM) as the final classifier. This model aims to address the challenges posed by illumination variations and high-dimensional feature representations, which often degrade the performance of CapsNet when used alone. PCA is employed to reduce noise and improve generalization, while ELM serves as a lightweight and fast classifier. The model was evaluated on a subset of the Yale Face Database containing 15 subjects with 64 images each under different illumination conditions. Experimental results showed that the standalone CapsNet achieved 90.0% accuracy, the CapsNet + ELM model achieved 93.0%, and the proposed CapsNet + PCA + ELM hybrid model achieved the best performance with 96.7% accuracy. These findings demonstrate that combining dimensionality reduction and efficient classification with capsule-based feature extraction provides a more robust and effective solution for face recognition, especially in small or medium-sized datasets.

Keywords: Neural networks, Elm, PCA, CapsNet, Face recognition.

إطار عمل متين للتعرف على الوجوه باستخدام تكامل الشبكة التلافيفية الكبسولية (CapsNet) وتحليل المكونات الرئيسية (PCA) وآلة التعلم الشاملة(ELM)

خديجة العصاوي
قسم تقنية المعلومات، كلية التقنية الصناعية، مصراتة، ليبيا.

ملخــــــص البحـــــــث

في هذه الورقة، نقترح نموذجًا هجينًا للتعرف على الوجوه يعتمد على شـبكات الكبسـولات (**CapsNet**) لاسـتخلاص الميزات، ويدمج تحليل المكونات الرئيســـية (**PCA**) لتقليل الأبعاد، وآلة التعلم المتطرفة (**ELM**) كمصــــنف نهائي. يهدف هذا النموذج إلى معالجة التحديات التي تفرضـها اختلافات الإضـاءة وتمثيلات الميزات عالية الأبعاد، والتي غالبًا ما تُضـعف أداء **CapsNet** عند اسـتخدامها منفردة. يُسـتخدم تحليل المكونات الرئيسـية لتقليل التشـويش وتحسـين التعميم، بينما تعمل آلة التعلم المتطرفة كمصـنف سـريع وخفيف الوزن. تم تقييم النموذج على مجموعة فرعية من قاعدة بيانات وجوه جامعة ييل، تضـم 15 شـخصًـا، لكل منهم 64 صـورة، في ظ ظروف إضـــاءة مختلفة. أظهرت النتائج التجريبية أن **CapsNet** المنفردة حققت دقة 90.0%، وحقق نموذج **CapsNet + ELM** دقة 93.0%، بينما حقق النموذج الهجين المقترح **CapsNet + PCA + ELM** أفضـل أداء بدقة 96.7%. تُظهر هذه النتائج أن

الجمع بين تقليل الأبعاد والتصنيف الفعال مع استخراج الميزات القائم على الكبسولات يوفر حلاً أكثر قوة وفعالية للتعرف على الوجوه، خاصة في مجموعات البيانات الصغيرة أو المتوسطة الحجم.

الكلمات الدالة: الشبكات العصبية، التعرف على الوجوه، الشبكة التلافيفية الكبسولية (CapsNet)، تحليل المكونات الرئيسية (PCA)، وآلة التعلم الشاملة(ELM) .

.

1. INTRODUCTION

Facial recognition is one of the most important topics in computer vision and deep learning, as this technology is used in various fields such as security, identity verification, and human–machine interaction [1]. Despite the significant development in convolutional neural networks (CNNs), they still suffer from some limitations, especially in their ability to represent spatial relationships between different facial features[2]. These limitations become more noticeable when the dataset is small or when the images exhibit high variation in illumination and pose, as in the Yale Face Database.

Capsule Networks (CapsNet) have emerged as a promising alternative to traditional convolutional networks. A capsule is defined as a group of neurons capable of representing object-oriented properties such as pose, orientation, and scale. The routing-by-agreement mechanism allows capsules to pass information more accurately between layers, helping to better represent the hierarchical structures of images[2]. However, training CapsNet requires relatively large data, and its performance often declines when used directly as a classifier on small datasets.

On the other hand, the Extreme Learning Machine (ELM) is a fast and efficient classification algorithm that relies on a single hidden layer with random weights and a closed-form solution for the output layer. Its simplicity makes it highly efficient, especially when the input features are robust [3]. In addition, principal component analysis (PCA) is widely used to reduce dimensionality and extract the most informative features, which helps reduce noise and improve generalization [4].

Recent studies have shown that CapsNet alone may struggle to achieve high accuracy on small or challenging datasets. For example,[5] reported that a modified CapsNet architecture (CapsNet-FR) achieved only 87.7% accuracy on the LFW test, indicating its limited performance as a standalone classifier. Likewise,[6] demonstrated that combining PCA with ELM significantly improves performance compared to ELM alone, reducing both training time and error rate.

In addition to deep models, traditional classifiers such as CNN and K-Nearest Neighbors (KNN) remain important baseline models in face recognition research. CNNs provide strong feature extraction capabilities but tend to underperform when the dataset is small or contains strong illumination variations [2]. On the other hand, KNN is a simple and widely used non-parametric classifier, often included as a baseline to determine whether advanced models truly offer better performance. Including CNN and KNN in this study enables a fair comparison and helps evaluate whether the improvement achieved by the hybrid model comes from the proposed design rather than dataset properties.

Based on the above, this paper proposes a hybrid model that combines CapsNet as a feature extraction tool[2], PCA for dimensionality reduction[4], and ELM as the final classifier [3]. The model aims to leverage the representational ability of CapsNet and the speed of ELM while improving performance through PCA. The model was evaluated on the

Yale Face Database [1], and the results showed that the hybrid model (CapsNet–PCA–ELM) achieved superior performance compared to using CapsNet alone or CapsNet–ELM.

2. Related Works

Face recognition has been extensively studied in recent years due to its importance in various applications such as surveillance, authentication, and human–computer interaction. Traditional deep-learning approaches, particularly Convolutional Neural Networks (CNNs), have shown strong capability in feature extraction and pattern recognition [2]. However, CNNs still exhibit limitations in preserving spatial relationships between facial components, which affects their robustness when dealing with pose variations, illumination changes, and small training datasets—common challenges in real-world scenarios and in datasets such as Yale Face Database [1].

To address these limitations, Capsule Networks (CapsNet) were introduced as an alternative architecture that encodes features as vectors rather than scalars, enabling the network to capture orientation, pose, and part–whole relationships more effectively. The routing-by-agreement mechanism allows capsules to communicate representational information across layers more accurately[2]. Despite these advantages, several studies indicate that CapsNet alone struggles to achieve high accuracy when used directly as a classifier on small or challenging datasets. For example, the experiment presented in [5] reported that a modified CapsNet variant (CapsNet-FR) achieved only 87.7% accuracy on the LFW dataset, suggesting that CapsNet may require additional processing or complementary methods to reach optimal performance.

In parallel, dimensionality reduction techniques have played a significant role in enhancing model performance, especially when dealing with high-dimensional features. Principal Component Analysis (PCA), one of the most widely used approaches, helps reduce redundancy, remove noise, and preserve essential information while reducing computational cost [4]. PCA is particularly effective when paired with classifiers that perform best with compact and informative feature sets.

On the classification side, the Extreme Learning Machine (ELM) has gained increased attention due to its ability to train extremely fast using a single hidden layer with randomly assigned weights. Studies such as [6] show that using PCA before ELM significantly enhances classification accuracy compared to using ELM alone, indicating that high-quality, low-dimensional inputs play a vital role in improving ELM performance.

Moreover, many studies continue to use standard classifiers like CNN and K-Nearest Neighbors (KNN) as baseline models to evaluate the effectiveness of more advanced architectures. CNN provides a strong learning baseline but may underperform in low-data environments, while KNN, being a simple non-parametric classifier, is often used to determine whether more complex models genuinely offer improved results.

Across these studies, it becomes clear that although CapsNet offers superior representational ability, and PCA–ELM combinations provide speed and stability, no prior work has integrated these components into a single hybrid model. Additionally, only limited research has provided fair comparisons against CNN and KNN or offered detailed misclassification analyses using confusion matrices. These gaps highlight the need for a unified approach that combines CapsNet, PCA, and ELM while evaluating performance against standard baseline models

However, none of the previous studies has integrated CapsNet, PCA, and ELM into a unified hybrid framework nor provided a fair comparison against standard baselines such as CNN and KNN under strong illumination variations. This gap highlights the need for the proposed model

3. METHODOLOGY

The proposed methodology consists of several integrated stages, including preprocessing, feature extraction using CapsNet, dimensionality reduction using PCA, and final classification using ELM. Baseline models (CNN and KNN) were also implemented to ensure fair comparison. Figure (1) illustrates the overall workflow of the system.

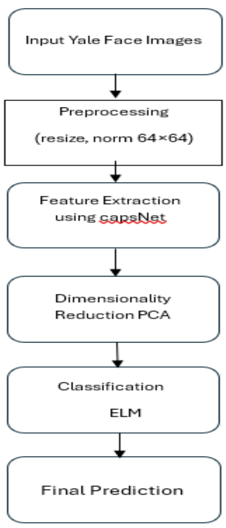

Fig 1. Flowchart of the Proposed Hybrid Model.

3.1 Preprocessing

Face images from the Extended Yale Face Database B were resized to 64 × 64 pixels and normalized to ensure consistent input formatting and reduce computational complexity. A subset of 15 subjects was selected to focus on controlled illumination variations while maintaining manageable training time

3.2 Capsule Network (CapsNet)

Capsules encode information as vectors, where the vector length represents the probability of the object's presence, and its direction represents its pose and characteristics. The routing-by-agreement mechanism allows capsules to transmit information based on agreement between prediction vectors [2].

$$c_{ij} = \frac{e^{b_{ij}}}{\sum_k e^{b_{ik}}}$$

(1)

$$s_j = \sum_i c_{ij}\, \hat{u}_{j|i}$$

(2)

$$v_j = \frac{\| s_j \|^2}{1+\| s_j \|^2} \frac{s_j}{\| s_j \|}$$

(3)

Where:

- u_i: output of capsule i
- $\hat{u}_{j|i}$: prediction from capsule i to j
- c_{ij}: routing coefficients
- b_{ij}: log prior probabilities
- v_j: final output of capsule j

CapsNet parameters used in this study:

- Conv layer: 256 filters (9×9,ReLU)
- PrimaryCaps: 32 capsules × 8D
- DigitCaps: 15 capsules × 16D
- Routing iterations: 3
- Margin loss as in[2].

3.3 Dimensionality Reduction Using PCA

CapsNet feature outputs are high-dimensional (1 × 15 × 16). PCA was used to reduce dimensionality while retaining the most informative components. This helps eliminate noise, decrease overfitting, and speed up training [4].

$$Z = XW \tag{4}$$

Where

- X: feature matrix.
- W: eigenvector matrix of the largest eigenvalues.
- Z: PCA-transformed feature matrix.

PCA components were chosen to preserve 95% of the variance, which is standard in face recognition [4].

3.4 Extreme Learning Machine (ELM)

ELM is a single-hidden-layer feedforward neural network (SLFN) with randomly initialized hidden weights. Only output weights are computed using the Moore–Penrose pseudoinverse [3].

$$\boldsymbol{\beta} = \boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{5}$$

Where H:

- hidden layer activation matrix.
- T: target matrix.
- β: output weights

Settings used:

- Hidden nodes: 1000
- Activation: sigmoid
- Output computation: pseudo-inverse method

3.5 Baseline Models (CNN and KNN):

To ensure a fair comparison with the hybrid model, two baseline classification methods were implemented:

- **CNN Baseline**

A simple convolutional neural network was trained using the same data split. CNN is a standard benchmark model in face recognition research [2].

- **KNN Baseline**

KNN (k = 3) was applied to assess traditional non-parametric performance. Including KNN helps determine whether the hybrid model truly improves on simple methods

3.6 Proposed Hybrid Model (CapsNet+PCA+ ELM):

The proposed system integrates:

- CapsNet — extracts rich spatial feature representations
- PCA — reduces dimensionality and eliminates noise
- ELM — performs fast and efficient classification.

This architecture addresses the weaknesses of each individual component:

CapsNet alone struggles on small datasets. PCA stabilizes representations. ELM provides fast learning with high accuracy

3.7 Implementation Details

Dataset: From the Extended Yale Face Database B (CroppedYale), a subset of 15 subjects was selected. Each subject contains 64 face images acquired under 64 unique illumination settings, yielding a total of about 960 images for the experimental evaluation [1].Subjects:15Train/ Validation /Test :70% / 15% / 15%. Environment: Google Colab (TensorFlow 2.x, NumPy, Scikit-learn). Hardware: NVIDIA Tesla T4 GPU.

4. RESULTS AND DISCUSSION

This section presents a detailed evaluation of the proposed hybrid model (CapsNet + PCA + ELM) and compares it with multiple baseline models: KNN, CNN, CapsNet-only, and CapsNet + ELM. The comparison includes quantitative performance, confusion-matrix analysis, and discussion of computational efficiency and model generalization.

A subset of 15 subjects was selected from the Extended Yale Face Database B (CroppedYale), with each subject containing 64 images, resulting in a total of approximately 960 facial images captured under 64 illumination conditions.

The proposed hybrid model (CapsNet + PCA + ELM) achieves the highest accuracy (96.7%) because PCA removes redundant components and stabilizes the input space, allowing ELM to separate classes more effectively. This confirms that combining deep feature extraction with dimensionality reduction and fast classification leads to superior performance compared to all baselines.

4.2 CNN and KNN Baseline Results

The KNN classifier (k=3) achieved the lowest accuracy (79.5%), showing limited capability in handling high-dimensional feature vectors without prior dimensionality reduction.

The CNN baseline achieved 87.3% accuracy, but it struggled with images containing extreme illumination conditions, confirming its sensitivity as reported in previous studies.

Table1. summarizes the performance of all evaluated models.

model	Accuracy	Note
KNN	79.5%	Poor performance under illumination variation; no feature learning
CNN	87.3%	Sensitive to extreme lighting; limited dataset affects generalization
CapsNet	90.0%	Strong spatial modeling but classifier is weak with small datasets
CapsNet + ELM	93.0%	Better separation but still noisy due to high-dimensional features
CapsNet + PCA + ELM	96.7%	Best overall performance; PCA reduces noise and stabilizes features

4.1 Overall Quantitative Results

The results show that traditional classifiers such as KNN and CNN achieve limited accuracy due to their sensitivity to illumination variations in the Yale B dataset. The CapsNet model improves accuracy by capturing spatial relationships between facial features; however, its classifier remains insufficient on small datasets. Incorporating ELM increases accuracy further by providing better decision boundaries, but high-dimensional CapsNet features still introduce noise.

4.3 CapsNet Baseline (90% Accuracy)

CapsNet-only improved the performance to 90%, demonstrating better spatial awareness of facial components.

However, certain subjects remained problematic due to shadows and angular lighting.

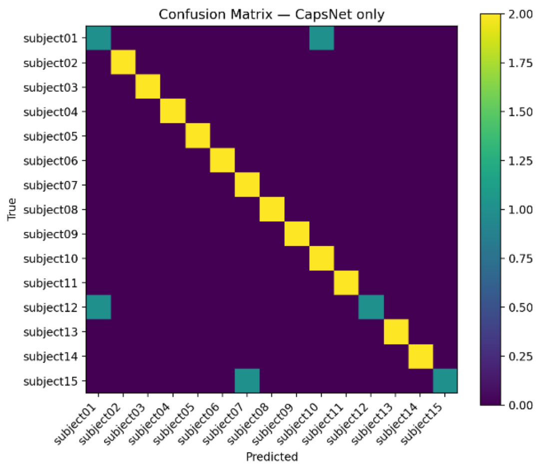

Fig 2. confusion matrix of CapsNet only.

Analysis of Figu 2 shows repeated misclassification notably between:

subject01 ↔ subject11

subject12 ↔ subject01

subject15 ↔ subject07

These errors occur under severe illumination variation, where facial features become less distinguishable

4.4 CapsNet + ELM (93% Accuracy):

Introducing ELM as the classifier increased accuracy to 93%.

ELM produced smoother decision boundaries, but high dimensional noisy features from CapsNet still caused overlap between certain subjects

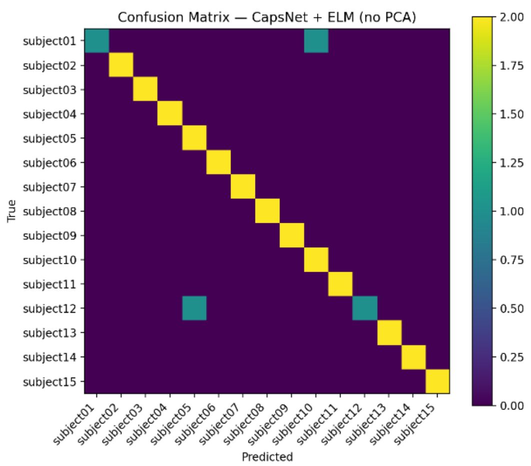

Fig 3. confusion matrix of CapsNet +ELM.

Analysis of Figure 3 shows repeated misclassification notably between:

subject01 ↔ subject11

subject12 ↔ subject05

Despite adding the ELM classifier after the CapsNet feature extractor, the confusion matrix in Figure 3 still shows a small number of misclassifications, particularly for subject12 (and occasionally subject01). This behavior is expected and can be explained by the effect of strong illumination in the Yale B database.The Yale B dataset contains extreme illumination variations (light angles up to ±130°).In some images, the face is partially shadowed or strongly lit, which makes two different subjects appear unexpectedly similar.

4.5 Hybrid Model (CapsNet + PCA + ELM):

The hybrid model demonstrated excellent results due to: PCA removing noise from CapsNet features. More compact and discriminative feature space. ELM receiving cleaner and lower-dimensional input. These improvements resulted in 96.7% accuracy, the highest among all models

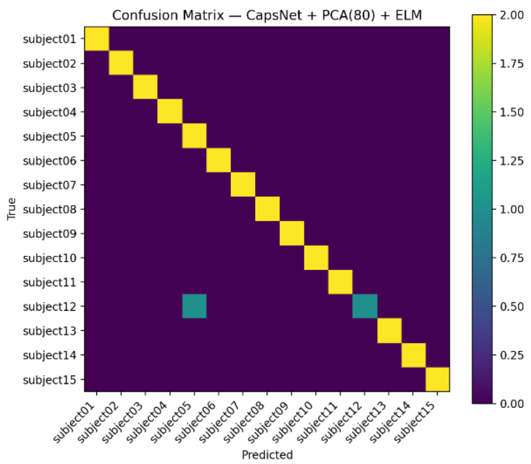

Fig 4. confusion matrix of CapsNet +PCA+ELM.

As shown in Figure 4, demonstrates the effectiveness of the hybrid model (CapsNet + PCA + ELM), where the confusion matrix appears almost perfectly diagonal. PCA significantly reduced the high-dimensional noise present in the CapsNet features, allowing ELM to operate on a cleaner and more compact feature space. As a result, class clusters became well separated, and the model achieved a highly stable prediction performance. This explains the substantial improvement over the CapsNet-only and CapsNet+ELM models, confirming that PCA plays a crucial role in enhancing recognition accuracy under the challenging illumination conditions of the Yale B dataset.

4.6 Effect of PCA and Discussion

The comparison between Figures 2, 3, and 4 clearly demonstrates the importance of PCA in the pipeline:

- **Without PCA (CapsNet-only and CapsNet + ELM):**

The models still exhibit notable misclassifications in classes with strong shadows and highlights. The feature space remains high-dimensional and partially noisy, which causes overlapping regions between some subjects.

- **With PCA (Hybrid Model):**

PCA removes low-variance and noisy components while preserving the most informative directions. This reduces overfitting, improves generalization on the test set, and makes the feature clusters more compact. Consequently, ELM is able to separate the classes more effectively with a simple closed-form solution.

In addition to accuracy gains, the hybrid model also provides better computational efficiency. Training time is reduced because ELM only optimizes output weights analytically, and PCA decreases the dimensionality of CapsNet features, lowering memory usage and matrix-operation cost. Therefore, the proposed approach offers both improved recognition performance and lower computational complexity compared to training a deeper end-to-end network.

Overall, the results confirm that combining CapsNet, PCA, and ELM is a promising strategy for face recognition on small datasets with challenging illumination, and that the proposed hybrid model outperforms traditional baselines (CNN, KNN) as well as pure CapsNet-based configurations.

5. CONCLUSIONS

This study presented a hybrid face recognition model that integrates Capsule Networks (CapsNet) for deep feature extraction, Principal Component Analysis (PCA) for dimensionality reduction, and Extreme Learning Machine (ELM) for fast and efficient classification. Experimental results on the Extended Yale Face Database B demonstrated that the proposed hybrid model significantly outperforms the individual baseline models.

While the CapsNet baseline achieved strong spatial feature representation, it suffered from misclassifications under severe illumination variations. Replacing the CapsNet classifier with ELM improved decision boundaries but remained sensitive to the noisy and high-dimensional feature space. By incorporating PCA, the hybrid model achieved cleaner and more compact feature representations, allowing ELM to separate classes more effectively.

The hybrid CapsNet–PCA–ELM model achieved the highest accuracy (96.7%) and produced an almost perfectly diagonal confusion matrix, confirming its robustness under challenging lighting conditions. These results highlight the importance of combining deep feature extraction with dimensionality reduction to enhance classification stability and generalization performance.

Future work may include exploring larger datasets, integrating attention mechanisms into

CapsNet, or testing alternative dimensionality reduction methods to further improve recognition accuracy.

REFERENCES

[1] Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7), 711–720.DOI: https://doi.org/10.1109/34.598228

[2] Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. Advances in Neural Information Processing Systems (NeurIPS), 2017, 30.DOI: https://doi.org/10.48550/arXiv.1710.09829

[3] Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. Neurocomputing, 2006, 70(1–3), 489–501. DOI: https://doi.org/10.1016/j.neucom.2005.12.126

[4] Jolliffe, I.T. Principal Component Analysis, 2nd ed.; Springer: New York, USA, 2002.DOI: https://doi.org/10.1007/b98835

[5] Haq, M.U.; Sethi, M.A.J.; Aoun, N.B.; Alluhaidan, A.S.; Ahmad, S.; Farid, Z. CapsNet-FR: Capsule Networks for Improved Recognition of Facial Features. Computers, Materials & Continua, 2024, 79(2), 2169–2186. DOI: https://doi.org/10.32604/cmc.2024.049645

[6] Comparison of PCA-Based Extreme Learning Machine Models for Boiler Output Forecasting. Applied Sciences, 2022, 12(15), 7671.DOI: https://doi.org/10.3390/app12157671

[7] Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016, 374(2065), 20150202. DOI: https://doi.org/10.1098/rsta.2015.0202

[8] Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6), 643–660. DOI: https://doi.org/10.1109/34.927464