IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 1, February 2026, pp. 695~706
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i1.pp695-706

Deep feature-based multi-class Alzheimer's disease classification with statistical performance evaluation

Maysaloon Abed Qasim[1], Marwa Mawfaq Mohamedsheet Al-Hatab[2], Lubab H. Albak[1]
[1]Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University, Mosul, Iraq
[2]Technical Engineering College, Northern Technical University, Mosul, Iraq

Article Info

Article history:

Received Oct 8, 2024
Revised Jan 4, 2026
Accepted Jan 22, 2026

Keywords:

Alzheimer's disease
K-nearest neighbors
Principal component analysis
SqueezeNet
Wilcoxon signed-rank tests

ABSTRACT

This study evaluated the performance of multiple machine learning classifiers for the classification of Alzheimer's disease (AD) stages using deep features extracted from a pre-trained SqueezeNet model. Magnetic resonance imaging (MRI) scans were processed through SqueezeNet to generate high-dimensional feature vectors, which were then used as achieved an accuracy of 94.78% input to six classifiers: k-nearest neighbors (KNN), decision tree (DT), support vector machine (SVM), neural network (NN), naive Bayes (NB), and logistic regression (LR). Models were assessed using a 70/30% training-testing split and 5-, 10-, and 20-fold stratified cross-validation. Principal component analysis (PCA) was applied to retain 99% of variance. On the original dataset consisting of 6,400 images, KNN has achieved 97.48% accuracy and 0.998 area under the curve (AUC), and when a larger dataset of 44,000 images was used it achieved an accuracy and of 94.78% and an AUC of 0.987, demonstrating the system's robustness across scales. Statistical tests, including paired t-tests and Wilcoxon signed-rank tests, confirmed that KNN has significantly leveraged from PCA. These outcomes demonstrate that combining deep feature extraction with PCA improved the reliability and efficiency of the classifier for AD stage prediction.

Corresponding Author:

Maysaloon Abed Qasim
Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University
Mosul, Iraq
Email: maysloon.alhashim@ntu.edu.iq

1. INTRODUCTION

Alzheimer's disease (AD) is considered the leading cause of dementia worldwide, it can be characterized by symptoms like memory loss, cognitive decline, and changes in mood or personality [1]. The most affected regions of the brain include the hippocampus, amygdala, and other components of the limbic system, which play a crucial role in cognitive functioning [2]. Many individuals experience a transitional stage of cognitive decline called mild cognitive impairment (MCI) before the appearance of the severe Alzheimer's symptoms; this period represents an intermediate condition between normal aging and AD [3]. Therefore, MCI is considered a significant indicator for early diagnosis of the AD [4]. Physicians use a group of techniques in collaboration with neurologists and neuropsychologists in diagnosing AD [5]. These diagnostic approaches include reviewing clinical histories, conducting physical and neurological examinations, performing diagnostic tests, and administering cognitive assessments such as the mini-mental state examination (MMSE) [6]. However, these traditional methods may be time-consuming and can lead to inconsistent results.

Because AD primarily affects the gray matter of the brain [7], imaging of the brain has become an essential tool in analyzing the functional and structural changes associated with the disease [8]. Artificial neural networks (ANNs), which are a computational approach inspired by biological neural systems, have exhibited a great potential in the early detection of AD. ANN can quantify distinct patterns and biomarkers associated with the condition by analyzing large datasets of brain images [9], [10]. Training these networks on huge imaging data, can help clinicians in diagnosing and monitoring AD more effectively and efficiently. Once trained, ANNs can classify new unseen data accurately. Deep neural networks (DNNs), which are a subgroup of ANNs, containing multiple hidden layers between the input and output that enable them to learn complex representations and obtain high level accuracy in different applications [11], [12]. Systems using DNNs can recognize patterns, make predictions, and solve a variety of complex problems in different fields such as computer vision, speech recognition, and natural language processing [13]–[15].

2. LITERATURE REVIEW

Recently, the utilization of machine learning (ML) and deep learning (DL) techniques in diagnosis of AD has advanced remarkably, especially through brain imaging data analysis. In 2022, AlSaeed and Omar [16] proposed a hybrid system combining convolutional neural network (CNN) with traditional ML for AD classification using magnetic resonance imaging (MRI) data. The approach used CNN as feature extraction and ML as a classifier. Their results showed that combining features derived from CNN with algorithms such as support vector machines (SVM) and random forests significantly enhanced diagnostic accuracy, especially when using limited data samples. In 2023, Khalid *et al.* [17] developed improved feature extraction strategy that fused CNN-based and handcrafted texture features for predicting of multiple AD stages starting from normal cognitive states to MCI and advanced AD. Public MRI datasets were used in study. Their system has obtained a higher robustness and stage-specific accuracy compared to systems using traditional CNN only. In the same year, Cherian *et al.* [18] studied the potential of different ML approaches like, decision trees (DT), k-nearest neighbors (KNN), and logistic regression (LR) in early detection of AD. Their study highlighted that traditional ML algorithms remain valuable when implemented in low-complexity systems and can complement the working of deep networks in hybrid diagnostic pipelines.

Also in 2023, Vashishtha *et al.* [19] designed a hybrid DL model to detect early signs of AD from MRI scans, using advanced CNN architectures such as InceptionV2 and ResNet50. These models' hybridization has led to improved feature representation and enhanced classification accuracy through both training and validation datasets. In 2024, Nasir *et al.* [20] developed a hybrid system using deep and meta-learning models for MRI-based AD classification, demonstrating improved accuracy when combining CNNs with meta-learners. More recently, in 2025, Liu *et al.* [21] introduced a multi-modal transformer-based framework combined MRI, genotypic. This framework predicted brain ageing, cognitive decline, and amyloid pathology, extending diagnostic capability beyond simple classification. In the same year, Arya *et al*. [22] worked on a systematic review, which summarized recent DL and ML techniques, including CNNs, transfer learning, and hybrid architectures across imaging modalities for AD detection, with focusing on challenges such as dataset imbalance and model interpretability.

3. MATERIALS AND METHODS

3.1. Dataset

In this study, a publicly accessible AD MRI dataset from Kaggle has been utilized, which consists of structural MRI images groups into four clinically relevant classes: non-demented, very mild demented, mild demented, and moderate demented. The training-testing split provided on the website was intentionally avoided because it used a selection criterion may introduce hidden biases that weaken reproducibility. Instead, a subset of 6,400 images (176×208 pixels, JPG format) has been collected from the complete dataset, distributed as follows: 3,200 non-demented, 2,240 very mild demented, 896 mild demented, and 64 moderate demented. A 70-30 train-test division was implemented, in order to ensure fairness and robust evaluation. The problem of imbalanced data, especially the moderate demented limited representation of samples, is recognized. This inequality reflects the real clinical distribution of Alzheimer's stages, which enhances the ecological validity of the dataset. In addition, multiple cross-validation experiments have been performed to reduce overfitting and confirm the reliability of the results. This methodological decision enhances our results credibility and addresses directly concerns regarding dataset clarity and class distribution and distribution of the class. Figure 1 shows representative examples of the four diagnostic categories [23]: Figure 1(a) mild demented, Figure 1(b) moderate demented, Figure 1(c) non-demented, and Figure 1(d) very mild demented.

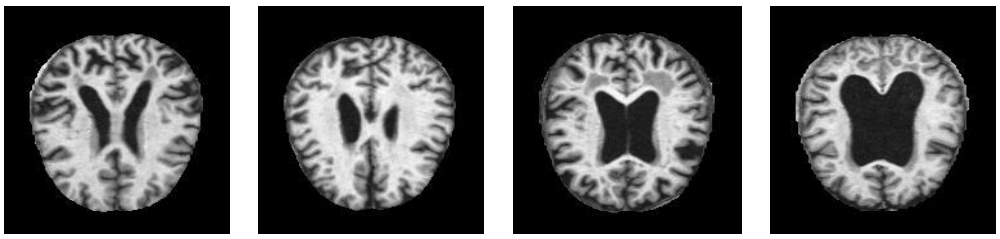


Figure 1. Class distribution in the dataset: (a) mild demented, (b) moderate demented, (c) non-demented, and (d) very mild demented

3.2. Deep feature extraction

A pre-trained SqueezeNet model was applied for deep feature extraction. Images were sent to the network through a dedicated feature layer, then the resulting feature vectors were use in the analysis and classification. SqueezeNet is a compact CNN architecture, which designed to provide high accuracy with significantly fewer parameters, making it suitable for memory-constrained environments and for efficient model deployment. Its design depends on 1×1 and 3×3 convolutions to keep strong performance [24].

The architecture is built around the fire module, which consists of a squeeze layer with 1×1 convolutions and an expand layer integrating 1×1 and 3×3 filters to capture multi-scale spatial features. Images were processed through conv1 and fire2–fire9, max-pooling was applied after conv1, fire4, fire8, and conv10 to gradually decrease spatial resolution and maintain essential information [24]. Nonlinear feature representation was supported by rectified linear unit (ReLU) activation, while overfitting was reduced by using dropout after fire9. Through the strategic downsampling and modular structure of SqueezeNet, a compact and expressive model was achieved, which is able to learn complex features efficiently [25]. Figure 2 presents the detailed architecture.

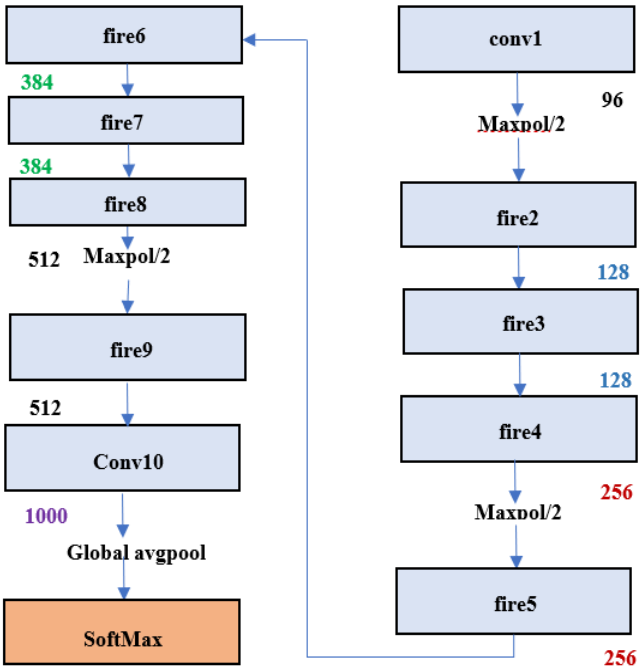


Figure 2. Squeeze Net architecture with a squeeze and expand process

3.3. Machine learning and model optimization

The study utilized different ML algorithms to classify the stages of AD using MRI data features extracted from CNN to improve the diagnostic accuracy [26], [27].

i) KNN: KNN identify the class depending on the majority among k nearest samples, which measured using Euclidean distance [27]:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

ii) NB: NB uses Bayes' theorem to estimate the posterior probability of class A given features B [28], [29]:

$$P\ (A|B) = \frac{P\ (B|A)P\ (A)}{P\ (B)} \quad (2)$$

iii) LR: LR models the positive class probability using the sigmoid function [30]:

$$P(X) = \frac{1}{1+e^{-z}} \quad (3)$$

and reduces the cross-entropy loss:

$$J(\theta) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(P(X_i)\right) + (1-y_i)\log\left(1-P(X_i)\right)\right] \quad (4)$$

iv) SVM: SVM employs the linear decision function to determine the optimal hyperplane that maximizes the margin between classes [31].

$$f(x) = w.x + b \quad (5)$$

where w and b are the weight and bias terms, respectively. Kernel functions enable nonlinear separation.

v) DT: DT repeatedly divides data into subsets based on entropy [31]:

$$\mathrm{E} = -\sum_{i=1}^{c} p(i) log_2\left(p(i)\right) \quad (6)$$

NN was also utilized to model non-linear relationships among high-dimensional features, involving 100 hidden units with ReLU activation.

3.4. Model optimization

3.4.1. Grid Search for traditional machine learning models

SVM, KNN, DT, LR, and NB hyperparameters were optimized using grid search. A 5-fold stratified cross-validation to avoid overfitting and retain balanced representation of AD classes. Table 1 summarizes the configurations of selected hyperparameter.

Table 1. Hyperparameters and grid search settings for traditional ML models

Model	Key hyperparameters	Justification
SVM	C=1.0, ε=0.1, polynomial kernel (degree 3)	Balances margin and misclassification; captures non-linear patterns
KNN	k=2, Metric: Manhattan, distance weighting	Sensitive to nearest neighbors while minimizing influence from distant points
DT	Max depth=100, Pruning: ≥4 instances in leaves, ≥5 internal nodes	Limits overfitting while capturing complex patterns
LR	L2 regularization, C=650	Reduces overfitting in high-dimensional feature space
NB	Default	Effective for large, independent feature sets

3.4.2. Bayesian optimization for NN model

To classify AD stages, 1000 deep features per image extracted from SqueezeNet, were processed using NN. Bayesian optimization was applied to fine-tune hyperparameters, such as the size of hidden layer and strength of regularization, based on cross-validated F1-score and accuracy. The NN used 100 hidden layers with ReLU activation and Adam optimizer, incorporating L2 regularization (α=0.0001) to reduce overfitting and limited to 100 iterations for replicable training. It took 1,000 deep features from SqueezeNet across 6,400 images, including meta-attributes like image name, size, width, and height. The model aimed to classify the AD stage category.

3.5. Principal component analysis

Principal component analysis (PCA) is a dimensionality reduction technique based on ML. It converts a large dataset to smaller set of components while maintaining essential patterns and variance. In this study 1000 features extracted from the SqueezeNet deep learning model were reduced to 100 principal components using PCA.

$$\text{Final Data} = \text{RowFeatureVector} \times \text{RowDataAdjusted} \quad (7)$$

3.6. Statistical significance analysis

The paired t-test and the Wilcoxon signed-rank test have been utilized, to identify whether the observed enhancements in the performance of the classification after applying PCA were statistically significant. These tests made a comparison between accuracies before and after PCA using multiple cross-validation folds to confirm that the gains were due to genuine improvements and not just random variation. Accuracies of the classification resulted from each fold corresponded as paired samples for statistical comparison.

i) Paired t-test: the paired t-test determines the mean difference between two paired samples, assuming the normality. The statistic of the test is given in:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \quad (8)$$

where, $\bar{d}$ is the mean difference between observations paired, it is computed as:

$$\bar{d} = \frac{1}{n} \sum_{1=1.}^{n} (x_i - y_i) \quad (9)$$

and s_d represents the standard deviation of differences, which can be determined as:

$$s_d = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} ((x_i - y_i) - \bar{d})^2} \quad (10)$$

where n represents the number of observations paired (cross-validation folds) x_i, y_i are the accuracies before and after applying PCA, respectively. The small value of p (<0.05), reflects the statistically significant improvements after PCA [32].

ii) Wilcoxon signed-rank test: the Wilcoxon signed-rank test does not assume a normal distribution; therefore, it considers as a non-parametric alternative to the paired t-test. It assesses if average of the differences between paired samples is zero. The test statistic can be computed using:

$$w = \sum_{i=1}^{n} R^+{}_i \quad (11)$$

where di is the difference between paired samples (xi−yi), and $R^+{}_i$ is the rank of the absolute differences for positive differences (di>0). Average ranks were assigned to the ties. When the value of p is less than 0.05, this refers that the observed differences are statistically significant, which confirms that PCA has contributed to improve the performance of the model [33].

4. RESULTS AND DISCUSSION

4.1. Evaluation criteria indicators

The performance of the model has been evaluated using accuracy, recall, precision, F1-score, and area under the curve (AUC). Among all models, KNN has obtained the best results with 93% accuracy and an AUC of 0.97, showing a great discriminative ability. Its configuration, which consists of two neighbors, Manhattan distance, and distance-weighted voting has taken part in this high performance. In contrast, DT demonstrated the lowest performance of 55% accuracy, AUC 0.64, referring to a limited stage classification. Its performance can be enhanced by tuning pruning thresholds and depth constraints. Moderate AUC of 0.68 low accuracy 43% have been achieved by SVM, suggesting the requirement for further kernel and regularization optimization. The performance of NN was well with 83% accuracy and AUC 0.94, which enable it to capture MRI data patterns effectively through its 100 hidden layers, ReLU activation, and Adam optimizer. The weak performance was recorded using NB with 42% accuracy, AUC 0.67, this is because of its independence assumption mismatch.

4.1.1. Confusion matrix evaluation

Figure 3 illustrates the confusion matrix for the KNN model, revealing the class-wise predictive performance as follows:

i) Mild demented: the percentage of correctly classified were 92.4%with little misclassifications.
ii) Moderate demented: with 92.4% correctly classified, and low false positives.
iii) Non-demented: with 95.0% correctly identified, and little misclassifications (3.3% mild, 6.6% very mild).
iv) Very mild demented: reached 91.7% correctly classified, and limited overlap (4.1% non-demented, 3.9% mild).

These results confirmed KNN robustness through all AD stages, especially for the identification of early dementia.

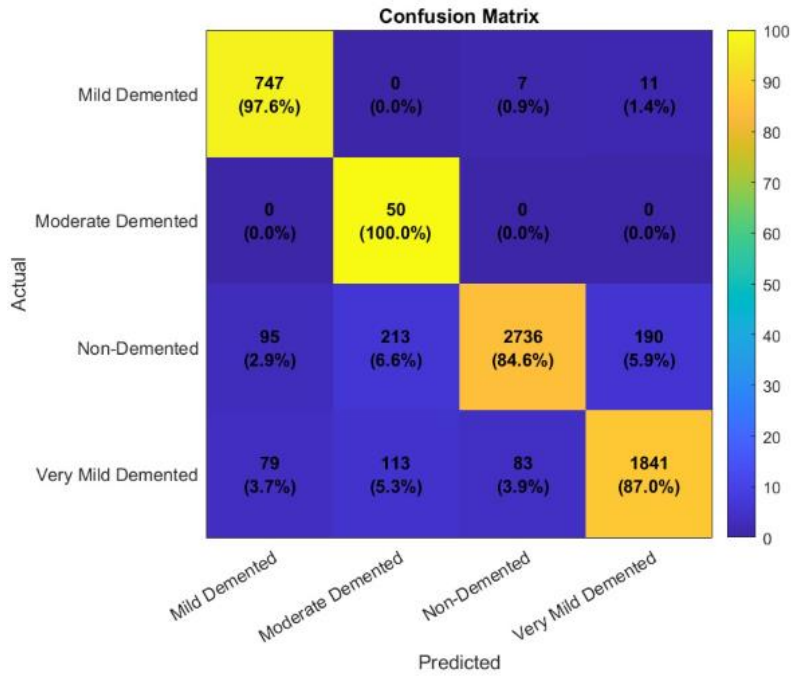


Figure 3. Confusion matrix for KNN model

4.1.2. ROC curve analysis

Figure 4 depicts the receiver operating characteristic (ROC) analyses the ML models across stages of AD mild demented as in Figure 4(a), moderate demented as in Figure 4(b), non-demented as in Figure 4(c), and very mild demented as in Figure 4(d). False positives and negatives were equally weighted (500 each), with class probabilities of 14%, 1%, 50%, and 35%. In early detection in the mild and very mild demented stages, high sensitivity was critical, while false diagnoses have been prevented in the non-demented group with accurate classification. the moderate demented class required the identification of rare instances with little false positives. Overall, ROC results highlight the trade-off between sensitivity and specificity across classes. KNN demonstrates the best classification for mild and very mild demented stages.

4.2. PCA dimensionality reduction

PCA technique was applied to deep features extracted from pre-trained SqueezeNet to reduce dimensionality while retaining the most informative components. Different values of variance thresholds of 90%, 95%, 96%, and 99% were tested. Variance of 96% has retained (~89 components) offering the optimal balance between accuracy of the classification and computational efficiency.

4.2.1. Performance comparison across PCA variance thresholds

Figure 5 illustrates the cumulative explained variance for the components of PCA, showing the ability of the first principal components to progressively capture data variability. Specifically, Figure 5(a) corresponds to a 90% variance threshold, Figure 5(b) to 95%, Figure 5(c) to 96%, and Figure 5(d) to 99%. Table 2 shows the evaluation metrics of ML models after applying PCA.

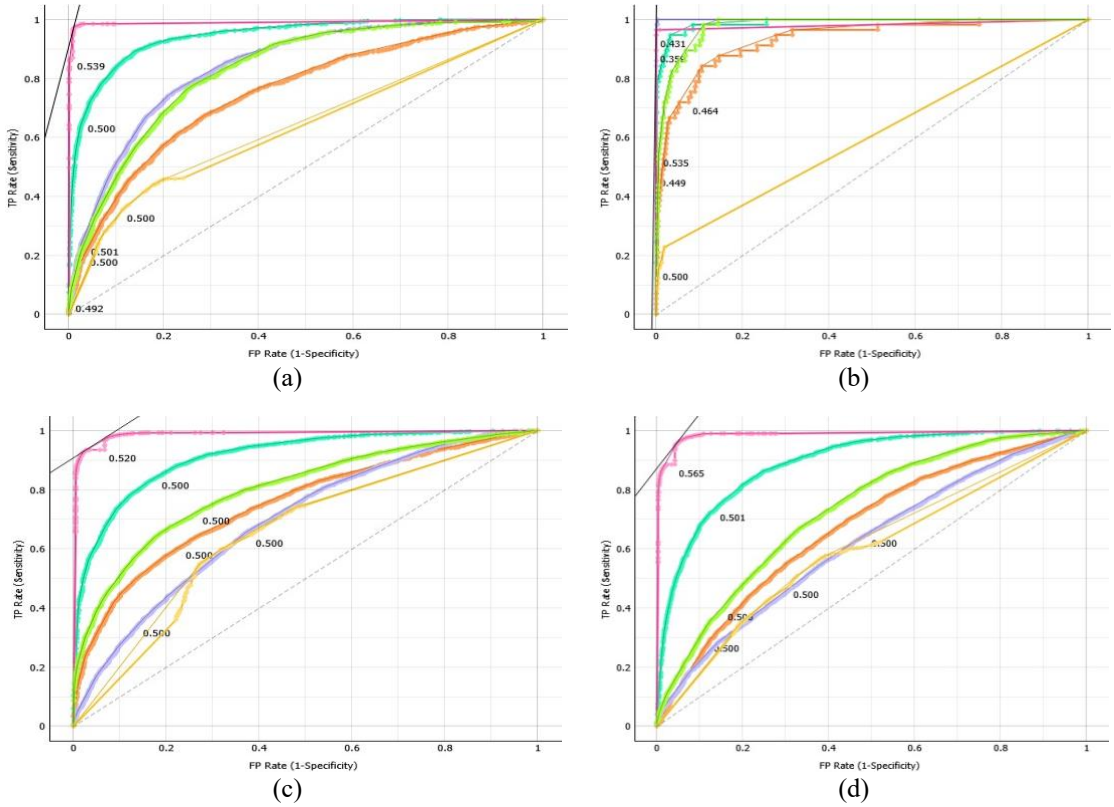


Figure 4. ROC curves for ML models (a) mild demented, (b) moderate demented, (c) non-demented, and (d) very mild demented

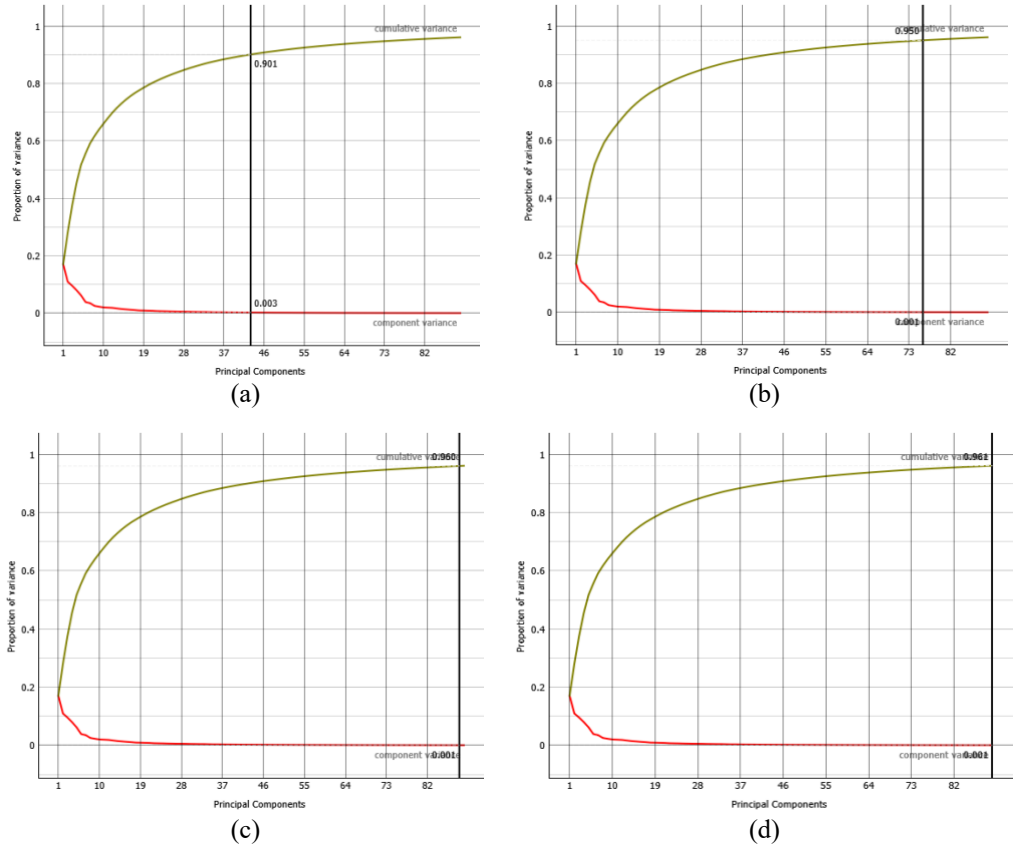


Figure 5. Cumulative explained variance at different thresholds: (a) 90%, (b) 95%, (c) 96%, and (d) 99%

Table 2. Evaluation metrics of ML models after PCA

PCA variance	Model	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)	AUC (%)
90% (~43 components)	KNN	54	54	54	54	63
	NN	72	72	72	72	87
	LR	60	60	59	58	76
	NB	57	57	57	57	72
	SVM	47	47	56	47	70
	DT	54	54	54	54	54
95% (~76 components)	KNN	92	92	92	92	98
	NN	77	77	77	77	91
	LR	61	61	60	60	77
	NB	55	55	57	55	72
	SVM	47	47	56	47	69
	DT	54	54	54	54	64
96% (~89 components)	KNN	93	93	93	93	99
	NN	79	78	78	78	91
	LR	62	62	61	61	78
	NB	54	54	57	55	72
	SVM	48	48	57	48	70
	DT	53	53	53	53	62
99% (~100 components)	KNN	93	93	93	93	99
	NN	79	79	78	78	92
	LR	62	61	62	61	79
	NB	54	54	56	57	72
	SVM	47	47	57	45	70
	DT	53	53	53	53	62

4.3. Cross-validation performance

The performance of multiple ML classifiers was assessed using stratified 5-, 10-, and 20-fold cross-validation, without applying PCA and after applying it to maintain 96% the variance. Without PCA, the highest performance was achieved by KNN, with accuracy of 97.05% and AUC of 98.99%, NN followed with accuracy of 86.83% and AUC of 96.24%. SVM and NB demonstrated poor performance with accuracy around 44% and AUC around 67%, while DT and LR shared the moderate performance.

After applying PCA, the performance of KNN improved, reaching 97.48% accuracy and 99.84% AUC, which indicates that removing redundant features enhanced generalization. NB showed a remarkable improvement, while the performance of NN, LR and DT slightly reduced due to the loss of subtle but informative features during dimensionality reduction. Table 3 summarizes the evaluation metrics across all models and folds.

Table 3. Evaluation metrics across 5-, 10-, and 20-fold stratified cross-validation (with and without PCA)

Model	Fold	AUC % (no PCA)	AUC % (with PCA)	Accuracy % (no PCA)	Accuracy % (with PCA)	F1-score % (no PCA)	F1-score % (with PCA)	Precision % (no PCA)	Precision % (with PCA)	Recall % (no PCA)	Recall % (with PCA)
KNN	5	98.03	99.32	95.38	94.89	95.37	94.88	95.38	94.89	95.38	94.89
	10	99.72	99.72	96.80	96.80	96.80	96.80	96.80	96.80	96.80	96.80
	20	98.99	99.84	97.05	97.48	97.05	97.48	97.05	97.49	97.05	97.48
DT	5	64.19	61.11	56.38	52.33	56.33	52.27	56.29	52.27	56.38	52.33
	10	62.34	62.34	52.70	52.70	52.74	52.74	52.84	52.84	52.70	52.70
	20	63.81	62.34	56.72	53.08	56.59	52.88	56.52	52.74	56.72	53.08
SVM	5	66.73	70.83	43.25	48.30	44.63	47.16	52.05	57.20	43.25	48.30
	10	71.01	71.01	47.84	47.84	46.54	46.54	57.40	57.40	47.84	47.84
	20	67.51	70.10	44.06	47.59	45.67	46.29	54.10	57.29	44.06	47.59
NN	5	95.07	92.84	84.45	80.64	84.41	80.49	84.41	80.55	84.45	80.64
	10	93.46	93.46	81.83	81.83	81.76	81.76	81.80	81.80	81.83	81.83
	20	96.24	93.82	86.83	82.11	86.79	82.02	86.78	82.05	86.83	82.11
NB	5	68.26	72.79	43.09	55.06	45.02	56.03	55.06	57.97	43.09	55.06
	10	71.01	73.59	47.84	56.33	46.54	57.06	57.40	58.76	47.84	56.33
	20	68.65	73.90	44.09	57.03	45.82	57.61	55.16	59.13	44.09	57.03
LR	5	82.91	78.39	67.38	62.06	67.28	61.42	67.21	61.29	67.38	62.06
	10	78.39	78.39	61.98	61.98	61.37	61.37	61.29	61.29	61.98	61.98
	20	83.57	78.41	68.41	61.97	68.29	61.33	68.20	61.24	68.41	61.97

4.4. Statistical significance analysis

Two tests were performed: Paired t-tests and Wilcoxon signed-rank. Most models have recorded statistically insignificant differences before and after PCA ($p>0.05$), except NB, showing significant increase

in accuracy (p=0.0144) in the t-test. This indicates that PCA is particularly useful for probabilistic classifiers. Table 4 illustrates the results obtained with and without PCA.

Table 4. Statistical test results for mean accuracy with and without PCA

Model	Mean accuracy without PCA (%)	Mean accuracy with PCA (%)	Paired t statistic	Paired t p-value	Wilcoxon	Wilcoxon p-value
KNN	96.41	96.39	-0.075	0.9469	1.000	0.6547
DT	55.27	52.70	-1.992	0.1847	0.000	0.1797
SVM	45.05	47.91	1.912	0.1960	0.000	0.1797
NN	84.37	81.53	-1.967	0.1881	0.000	0.1797
NB	45.01	56.14	8.241	0.0144	0.000	0.2500
LR	65.92	62.00	-1.973	0.1872	0.000	0.1797

4.5. Confusion matrix and calibration analysis

KNN's performance after PCA with (20-fold CV) was near-perfect with accuracy=97.48%, AUC=0.998. Figure 6 shows the calibration plot. It shows that predicted probabilities align well with observed outcomes, confirming that the model is well-calibrated and clinically interpretable.

4.6. Performance comparison across datasets

To check the generalization of system, a large dataset containing 44,000 samples has been applied across four dementia stages, KNN achieved the best performance again as shown in Table 5 [34], [35]. When using larger dataset, KNN retained the excellent performance with only a minor reduction due to increased data diversity. Figure 7 presents the performance evaluation of the proposed SqueezeNet+KNN framework on the new 44,000-image dataset. The confusion matrix in Figure 7(a) depicts a little overlap between mild and very mild demented groups, while the ROC curve in Figure 7(b) confirms the strong discriminative power across all stages. These results highlighted that the proposed SqueezeNet+KNN framework is scalable and robust for the classification of AD.

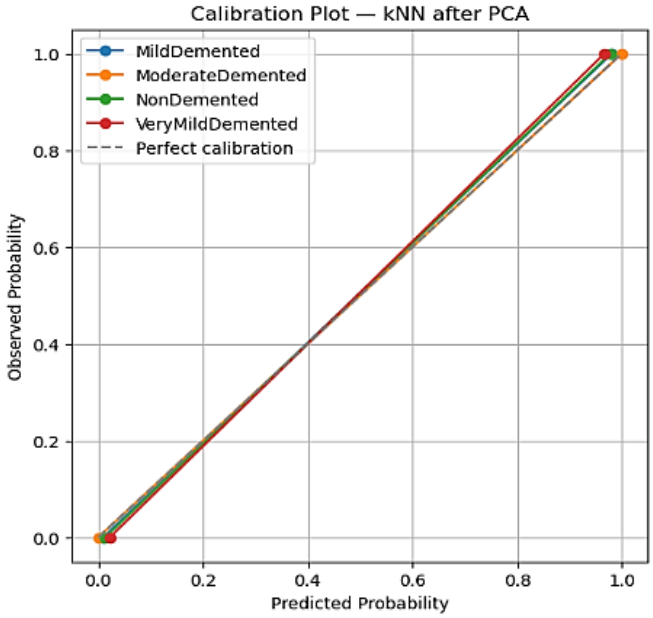
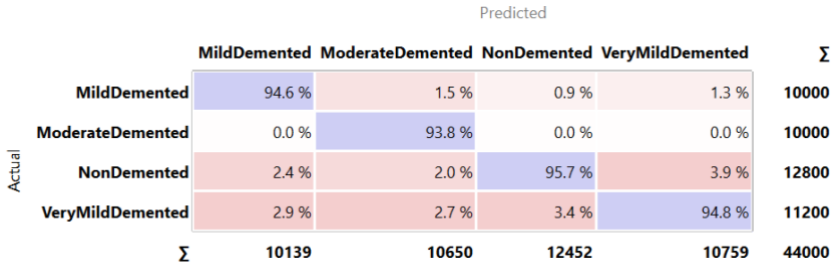


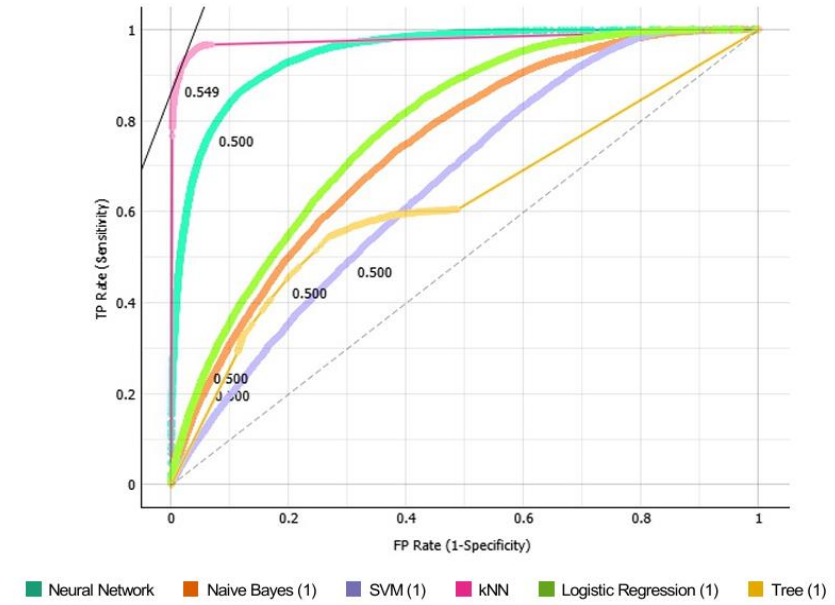
Figure 6. KNN performance after PCA: calibration plot

Table 5. Comparison of best model performance between original and larger dataset

Dataset	Best model	AUC (%)	Accuracy (%)	F1-score (%)	Precision (%)	Recall (%)
Original dataset (6,400 images, after PCA, 20-fold CV) [23]	KNN	99.84	97.48	97.48	97.49	97.48
New dataset (44,000 images, after PCA, 20-fold CV) [35]	KNN	98.74	94.78	94.75	94.79	94.78



(a)



(b)

Figure 7. Performance of KNN on the new 44,000-image dataset: (a) confusion matrix and (b) ROC curve

5. CONCLUSION

The results obtained from this study illustrate that integrating deep features from a pre-trained SqueezeNet with traditional ML classifiers can predict AD stages effectively. KNN has achieved the best performance consistently, with 97.48% accuracy and 0.998 AUC when using the original dataset consisting of 6,400 images, and with 94.78% accuracy with 98.74% AUC when using new larger dataset (44,000 images). Confusion matrix and ROC analyses also confirmed great discrimination, with minor early-stage misclassifications. Using PCA improved the efficiency and feature discrimination, particularly for KNN.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Maysaloon Abed Qasim	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓		✓	
Marwa Mawfaq	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓		
Mohamedsheet Al-Hatab														
Lubab H. Albak	✓			✓	✓	✓	✓		✓		✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Kaggle at https://www.kaggle.com/datasets/preetpalsingh25/alzheimers-dataset-4-class-of-images, reference [23].

REFERENCES

[1] A. L. Sosa-Ortiz, I. Acosta-Castillo, and M. J. Prince, "Epidemiology of dementias and Alzheimer's disease," *Archives of Medical Research*, vol. 43, no. 8, pp. 600–608, 2012, doi: 10.1016/j.arcmed.2012.11.003.

[2] D. A. de A. Junior *et al.*, "The association between post-traumatic stress disorder (PTSD) and cognitive impairment: a systematic review of neuroimaging findings," *Journal of Psychiatric Research*, vol. 164, pp. 259–269, 2023, doi: 10.1016/j.jpsychires.2023.06.016.

[3] R. L. Beard and T. M. Neary, "Making sense of nonsense: experiences of mild cognitive impairment," *Sociology of Health & Illness*, vol. 35, no. 1, pp. 130–146, 2013, doi: 10.1111/j.1467-9566.2012.01481.x.

[4] J. Sheng, Y. Xin, Q. Zhang, L. Wang, Z. Yang, and J. Yin, "Predictive classification of Alzheimer's disease using brain imaging and genetic data," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-06444-9.

[5] G. Waldemar *et al.*, "Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline," *European Journal of Neurology*, vol. 14, no. 1, 2007, doi: 10.1111/j.1468-1331.2006.01605.x.

[6] A. J. Mitchell, "The mini-mental state examination (MMSE): an update on its diagnostic validity for cognitive disorders," in *Cognitive Screening Instruments*, London, United Kingdom: Springer, 2013, pp. 15–46, doi: 10.1007/978-1-4471-2452-8_2.

[7] J. M. F. Montenegro, B. Villarini, A. Angelopoulou, E. Kapetanios, J. Garcia-Rodriguez, and V. Argyriou, "A survey of Alzheimer's disease early diagnosis methods for cognitive assessment," *Sensors*, vol. 20, no. 24, 2020, doi: 10.3390/s20247292.

[8] R. A. Hazarika, A. K. Maji, S. N. Sur, B. S. Paul, and D. Kandar, "A survey on classification algorithms of brain images in Alzheimer's disease based on feature extraction techniques," *IEEE Access*, vol. 9, pp. 58503–58536, 2021, doi: 10.1109/ACCESS.2021.3072559.

[9] O. I. Abiodun *et al.*, "Comprehensive review of artificial neural network applications to pattern recognition," *IEEE Access*, vol. 7, pp. 158820–158846, 2019, doi: 10.1109/ACCESS.2019.2945545.

[10] M. G. M. Abdolrasol *et al.*, "Artificial neural networks-based optimization techniques: a review," *Electronics*, vol. 10, no. 21, 2021, doi: 10.3390/electronics10212689.

[11] M. Gao and J. Mao, "A novel active rehabilitation model for stroke patients using electroencephalography signals and deep learning technology," *Frontiers in Neuroscience*, vol. 15, 2021, doi: 10.3389/fnins.2021.780147.

[12] P. K. Mallick, S. H. Ryu, S. K. Satapathy, S. Mishra, G. N. Nguyen, and P. Tiwari, "Brain MRI image classification for cancer detection using deep wavelet autoencoder-based deep neural network," *IEEE Access*, vol. 7, pp. 46278–46287, 2019, doi: 10.1109/ACCESS.2019.2902252.

[13] D. A. Arafa, H. E.-D. Moustafa, A. M. T. Ali-Eldin, and H. A. Ali, "Early detection of Alzheimer's disease based on the state-of-the-art deep learning approach: a comprehensive survey," *Multimedia Tools and Applications*, vol. 81, no. 17, pp. 23735–23776, 2022, doi: 10.1007/s11042-022-11925-0.

[14] S. Q. Hasan, "Shallow model and deep learning model for features extraction of images," *NTU Journal of Engineering and Technology*, vol. 2, no. 3, 2023, doi: 10.56286/ntujet.v2i3.449.

[15] R. H. M. Ameen, N. M. Basheer, and A. K. Younis, "A survey: breast cancer classification by using machine learning techniques," *NTU Journal of Engineering and Technology*, vol. 2, no. 1, May 2023, doi: 10.56286/ntujet.v2i1.367.

[16] D. AlSaeed and S. F. Omar, "Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning," *Sensors*, vol. 22, no. 8, 2022, doi: 10.3390/s22082911.

[17] A. Khalid, E. M. Senan, K. Al-Wagih, M. M. A. Al-Azzam, and Z. M. Alkhraisha, "Automatic analysis of MRI images for early prediction of Alzheimer's disease stages based on hybrid features of CNN and handcrafted features," *Diagnostics*, vol. 13, no. 9, 2023, doi: 10.3390/diagnostics13091654.

[18] I. Cherian, A. Agnihotri, A. K. Katkoori, and V. Prasad, "Machine learning for early detection of Alzheimer's disease from brain MRI," in *International Journal of Intelligent Systems and Applications in Engineering*, 2023, vol. 11, no. 7s, pp. 36–43.

[19] A. Vashishtha, A. K. Acharya, and S. Swain, "Hybrid model: deep learning method for early detection of Alzheimer's disease from MRI images," *Biomedical and Pharmacology Journal*, vol. 16, no. 3, pp. 1617–1630, 2023, doi: 10.13005/bpj/2739.

[20] N. Nasir, M. Ahmed, N. Afreen, and M. Sameer, "Alzheimer's magnetic resonance imaging classification using deep and meta learning models," *arXiv:2405.12126*, 2024.

[21] Y. Liu and X. Yang, "MMF-ViT: A multi-scale multi-domain frequency-aware vision transformer for MRI-based Alzheimer's classification," *Electronic Research Archive*, vol. 33, no. 10, pp. 5916–5936, 2025, doi: 10.3934/era.2025263.

[22] A. D. Arya *et al.*, "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease," *Brain Informatics*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40708-023-00195-7.

[23] P. P. Singh, "Alzheimer's dataset 4 class of images," *Kaggle*. Accessed: Jan. 10, 2025. [Online]. Available: https://www.kaggle.com/datasets/preetpalsingh25/alzheimers-dataset-4-class-of-images

[24] V. Bhole and A. Kumar, "Analysis of convolutional neural network using pre-trained SqueezeNet model for classification of thermal fruit images," in *ICT for Competitive Strategies*, 2020, pp. 759–768, doi: 10.1201/9781003052098-80.

[25] F. Shi, J. Wang, and V. Govindaraj, "SGS: SqueezeNet-guided Gaussian-kernel SVM for COVID-19 diagnosis," *Mobile Networks and Applications*, vol. 29, no. 6, pp. 1872–1885, 2024, doi: 10.1007/s11036-023-02288-3.
[26] S. J. Qin and L. H. Chiang, "Advances and opportunities in machine learning for process data analytics," *Computers & Chemical Engineering*, vol. 126, pp. 465–473, 2019, doi: 10.1016/j.compchemeng.2019.04.003.
[27] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K.-R. Muller, "Explaining deep neural networks and beyond: a review of methods and applications," *Proceedings of the IEEE*, vol. 109, no. 3, pp. 247–278, 2021, doi: 10.1109/JPROC.2021.3060483.
[28] M. Marimuthu, M. Abinaya, K. S. Hariesh, K. Madhankumar, and V. Pavithra, "A review on heart disease prediction using machine learning and data analytics approach," *International Journal of Computer Applications*, vol. 181, no. 18, pp. 20–25, 2018, doi: 10.5120/ijca2018917863.
[29] S. Goel, A. Deep, S. Srivastava, and A. Tripathi, "Comparative analysis of various techniques for heart disease prediction," in *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, Nov. 2019, pp. 88–94, doi: 10.1109/ISCON47742.2019.9036290.
[30] G. A. Ansari and S. S. Bhat, "Exploring a link between fasting perspective and different patterns of diabetes using a machine learning approach," in *Educational Research*, 2022, doi: 10.447/remie.2022.2.1.
[31] P. Tabesh, G. Lim, S. Khator, and C. Dacso, "A support vector machine approach for predicting heart conditions," in *IIE Annual Conference and Expo 2010 Proceedings*, 2010, pp. 1–6.
[32] T. Rietveld and R. van Hout, "The paired t test and beyond: recommendations for testing the central tendencies of two paired samples in research on speech, language and hearing pathology," *Journal of Communication Disorders*, vol. 69, pp. 44–57, 2017, doi: 10.1016/j.jcomdis.2017.07.002.
[33] K. Okoye and S. Hosseini, *R programming: statistical data analysis in research*. Singapore: Springer, 2024, doi: 10.1007/978-981-97-3385-9.
[34] A. K. Pallikonda and P. S. Varma, "Multi-class classification of Alzheimer's disease stages using SqueezeNet based approach for automated diagnosis," in *International Journal of Computer Information Systems and Industrial Management Applications*, 2023, vol. 15, no. 2023, pp. 321–331.
[35] A. Singhal, "Alzheimer's disease multiclass images dataset," *Kaggle*. Accessed: Oct. 22, 2025. [Online]. Available: https://www.kaggle.com/datasets/aryansinghal10/alzheimers-multiclass-dataset-equal-and-augmented

BIOGRAPHIES OF AUTHORS

Dr. Maysaloon Abed Qasim received the B.Sc. degree in Electrical Engineering from Mosul University, Iraq, the M.Sc. degree in Electronic and Communication from Mosul University, Iraq, and the Ph.D. degree Electronic and Computer Engineering from Istanbul, Turkey. She is assistant professor at Technical Engineering College for Computer and Artificial Intelligence, Northern Technical University (NTU), Mosul, Iraq. She has published multiple research papers. She can be contacted at email: maysloon.alhashim@ntu.edu.iq.

Marwa Mawfaq Mohamedsheet Al-Hatab received the B.Sc. degree in Medical Instruments Engineering from Technical Engineering College, Mosul, Northern Technical University (NTU), Mosul, Iraq, and the M.Sc. degree in Biomedical Engineering from Universita Politecnica delle Marche (UNIVPM). She is a lecturer at Technical Engineering College, Mosul, Northern Technical University (NTU), Mosul, Iraq. She has published multiple research paper. She can be contacted at email: marwa.alhatab@ntu.edu.iq.

Lubab H. Albak received B.Sc. and M.Sc. degree in Technical Computer Engineering in 1997 and 2009, respectively. During 2009 worked as an assistant lecturer at Technical College of Mosul, Iraq. In 2019, she achieved the title of lecturer in Northern Technical University. The research interests are in the recognition, security, artificial intelligence, and image processing. She can be contacted at email: lubab_harith@ntu.edu.iq.