Available online at www.dergipark.org.tr/en

INTERNATIONAL ADVANCED RESEARCHES and ENGINEERING JOURNAL

Journal homepage: www.dergipark.org.tr/en/pub/iarej

e-ISSN: 2618-575X

International Open Access

Volume 09
Issue 02

August, 2025

Research Article

Feature selection and classification of brain midline tumors using tiny vision transformers and neighborhood component analysis

Uğur Demiroğlu [a] ***and Bilal Şenol*** [b,*]

[a]*Department of Software Engineering, Faculty of Engineering, Architecture and Design, Kahramanmaraş İstiklal University, Kahramanmaraş, Türkiye*

[b]*Department of Software Engineering, Faculty of Engineering, Aksaray University, Aksaray, Türkiye*

ARTICLE INFO

Article history:
Received 07 February 2025
Accepted 30 June 2025
Published 20 August 2025

Keywords:
Brain midline tumors
Image classification
Neighborhood component analysis
Vision transformers

ABSTRACT

Brain midline tumors pose significant diagnostic challenges due to their complex structure and variability in presentation. In this study, we propose a robust classification framework for detecting brain midline tumors using image classification and feature selection techniques. The images, categorized into positive and negative classes, were processed using the Google Tiny Model Vision Transformer without any additional training. Feature extraction was performed by utilizing the head layer of the Tiny Model Vision Transformer, which yielded 1,000 features from the fully connected layer based on the pre-trained weights of the network. These features were initially classified using classical machine learning classifiers such as support vector machines, k-nearest neighbors and decision trees. To improve classification accuracy and reduce computational costs, Neighborhood Component Analysis was applied as a feature selection method. Neighborhood Component Analysis selected the top 460 most informative features from the initial set of 1,000 features. These selected features were subsequently used for classification, and the performance was compared with the results obtained using the full feature set. The comparative analysis revealed that the Neighborhood Component Analysis-based feature selection significantly enhanced classification accuracy and reduced processing time without sacrificing model reliability. The findings demonstrate that combining Tiny Model Vision Transformers with Neighborhood Component Analysis is an effective approach for brain midline tumor classification, offering a balance between accuracy and computational efficiency. This method holds promise for improving early diagnosis and aiding clinical decision-making, making it a valuable tool in medical image analysis and brain tumor detection.

1. Introduction

Brain midline tumors are a critical health concern, with accurate and early diagnosis being essential for improving patient outcomes. These tumors can lead to serious complications due to their location, which often affects vital brain functions. Traditional diagnostic techniques, such as biopsy and manual radiological evaluation, are invasive, time-consuming, and prone to variability based on radiologist experience [1]. Therefore, the application of automated image classification systems has emerged as a powerful tool for assisting medical professionals in diagnosing such conditions accurately and efficiently.

In recent years, advancements in deep learning and computer vision have significantly improved the performance of medical image analysis systems. Among these advancements, Vision Transformers (ViTs) have demonstrated superior performance in various classification tasks compared to traditional convolutional neural networks (CNNs) [2]. Unlike CNNs, ViTs leverage self-attention mechanisms that model global dependencies in images, making them highly suitable for complex medical imaging applications [3, 4]. However, standard ViT models are often computationally expensive, limiting their practical use in resource-constrained environments such as hospitals and clinical settings [5].

* Corresponding author. Tel.: *+90 505 2984717*
E-mail addresses: ugurdemiroglu@istiklal.edu.tr *(Uğur Demiroğlu),* bilal.senol@aksaray.edu.tr *(Bilal Şenol)*
ORCID: 0000-0002-0000-8411 *(Uğur Demiroğlu),* 0000-0002-3734-8807 *(Bilal Şenol)*
DOI: 10.35860/iarej.1635430

To address this challenge, this study leverages Google's Tiny Model Vision Transformer, which offers a lightweight architecture while maintaining strong classification performance. Feature extraction from the Tiny Model ViT allows for the capture of high-dimensional image representations without the need for extensive training, thus reducing computational costs and time [6]. These features can be classified using traditional machine learning classifiers such as support vector machines (SVM), k-nearest neighbors (KNN), and decision trees, which are known for their simplicity and robustness in small-scale datasets [7, 8].

The primary motivation for selecting Tiny ViT stems from its balance between computational efficiency and classification performance. Standard ViTs have demonstrated superior accuracy in various image classification tasks by leveraging self-attention mechanisms to model global dependencies. However, they are computationally expensive, requiring extensive memory and high-end GPU resources, which can be impractical in real-time medical applications. Tiny ViT, a lightweight version of ViTs, significantly reduces the number of parameters while maintaining competitive performance, making it well-suited for resource-constrained environments such as hospitals and diagnostic centers. Unlike larger ViT variants that require millions of parameters (e.g., ViT-Base with 86M parameters), Tiny ViT has only 5.7M parameters, leading to faster inference times and lower computational costs without requiring additional training.

Compared to CNNs, which are traditionally used for medical image analysis, Tiny ViT provides superior feature extraction capabilities due to its ability to capture long-range dependencies through self-attention rather than relying on local receptive fields. Additionally, CNNs often struggle with limited generalization in small datasets, while Tiny ViT leverages pre-trained weights from large-scale datasets, improving robustness in medical image classification. Another alternative, MobileNet, is commonly used for lightweight applications but has been shown to underperform compared to transformer-based models in complex image classification tasks due to its depthwise separable convolution constraints.

Despite its advantages, Tiny ViT also has limitations. While it significantly reduces computational costs compared to full-scale ViTs, it still requires more resources than traditional CNNs, which might be a concern for deployment in extremely low-power settings. Furthermore, ViTs generally require a larger amount of training data to fully exploit their self-attention mechanisms, though our study mitigates this by using a pre-trained model. To acknowledge these trade-offs, we will add a comparative discussion and a table summarizing the key differences between Tiny ViT, CNN-based models, and other transformer variants.

Although deep learning models can extract a large number of features, not all features contribute equally to classification performance. Some features may be redundant or irrelevant, increasing computational costs without improving accuracy. Feature selection methods are essential for addressing this problem. In this study, Neighborhood Component Analysis (NCA) is employed as a feature selection technique to reduce the initial set of 1,000 features extracted from the Tiny Model ViT to 460 features [9]. NCA is a supervised dimensionality reduction technique that identifies features most relevant to improving classification performance, thereby optimizing both accuracy and efficiency [10, 11].

The objective of this study is to investigate the classification performance of brain midline tumor images using Tiny Model Vision Transformer-based feature extraction combined with classical classifiers. Specifically, the study aims to:

- Extract 1,000 features from brain midline tumor images using the head layer of the Tiny Model ViT without additional training.
- Classify these features using traditional classifiers such as SVM, KNN, and decision trees to establish a baseline performance.
- Apply NCA to select the most relevant 460 features and compare the classification performance before and after feature selection.

By integrating deep learning feature extraction with classical machine learning and feature selection, this study offers a novel framework for accurate, efficient, and cost-effective brain tumor classification. The results are expected to contribute to the growing body of research on hybrid machine learning approaches for medical image analysis [12–15].

Structure of this paper is as follows. Section 2 gives information about existing studies in the literature. Section 3 introduces the methodology of the paper including information of the ViTs and the dataset used in the paper. A thorough case study is given in Section 4 with illustratives and the last section gives the conclusions.

2. Related Work

The automatic detection of brain tumors using machine learning and deep learning techniques has gained significant attention in recent years. Traditional approaches primarily relied on handcrafted features and classical machine learning algorithms for tumor classification [16]. Although these methods achieved moderate success, they often suffered from limited accuracy due to their inability to capture complex image features [17]. The advent of deep learning, particularly CNNs, revolutionized medical image analysis by enabling automatic feature extraction and improved classification

performance [18]. However, CNNs are constrained by their local receptve fields and difficulty in capturing long-range dependencies in images, leading to the emergence of ViTs as a powerful alternative [19].

ViTs were first introduced for natural image classification tasks and have since been adapted for various medical imaging applications [20]. Unlike CNNs, ViTs rely on self-attention mechanisms, which enable them to model global relationships across an entire image. Studies have demonstrated the superiority of ViTs in tasks such as tumor segmentation, breast cancer detection, and retinal disease classification [21, 22]. Although standard ViT models provide excellent accuracy, their high computational cost limits their applicability in clinical practice. To address this limitation, lightweight models such as the Tiny Vision Transformer have been developed, offering a balance between performance and efficiency [23].

Several studies have explored the combination of deep learning-based feature extraction with classical machine learning classifiers to reduce computational complexity while maintaining high accuracy. For instance, Gupta et al. [24] used pre-trained CNN models to extract features from brain tumor MRI images, followed by classification with SVM, achieving significant accuracy improvements. Similarly, Zhang et al. [25] proposed a hybrid framework combining CNN feature extraction and KNN for lung cancer diagnosis, highlighting the effectiveness of such hybrid approaches.

Feature selection is a crucial step in machine learning pipelines, particularly in medical image analysis, where high-dimensional feature spaces can degrade performance. Dimensionality reduction techniques, such as principal component analysis (PCA) and linear discriminant analysis (LDA), have been widely used in previous studies [26, 27]. However, recent studies have shown that supervised feature selection algorithms, such as NCA, offer superior performance by focusing on features most relevant to the target classification task [28]. NCA has been successfully applied in various domains, including cancer detection, heart disease prediction, and diabetic retinopathy classification [29, 30].

The integration of ViTs with feature selection algorithms is a relatively unexplored area. Existing research primarily focuses on CNN-based feature extraction, while only a few studies have investigated the use of ViTs for feature extraction in combination with classical classifiers. This study aims to fill this gap by applying NCA for feature selection on ViT-extracted features and evaluating its impact on classification performance.

Recent studies on brain tumor classification have primarily relied on CNNs and hybrid approaches integrating deep learning with traditional classifiers. CNN-based methods have demonstrated strong feature extraction capabilities but are constrained by their local receptive fields, limiting their ability to capture long-range dependencies in medical images [31]. To address this limitation, ViTs have emerged as a promising alternative due to their self-attention mechanisms, which effectively model global relationships in images [2]. However, standard ViTs require substantial computational resources, making them impractical for real-time clinical applications [5]. Prior research has explored the integration of deep feature extraction with classical machine learning classifiers to improve efficiency. For instance, Gupta et al. employed CNN-based feature extraction followed by SVM classification, achieving improved accuracy but at a high computational cost [24]. Similarly, Zhang et al. combined CNN-derived features with KNN, but their approach lacked an effective feature selection step, leading to redundant features and increased processing time [25]. Unlike these methods, our study leverages Google's Tiny Vision Transformer for lightweight yet powerful feature extraction, significantly reducing computational complexity while maintaining high classification performance. Additionally, our use of NCA for feature selection eliminates redundant features, enhancing model efficiency and accuracy beyond conventional dimensionality reduction techniques such as PCA [9]. While prior studies have incorporated feature selection in tumor classification, NCA remains underexplored in ViT-based medical image analysis, highlighting the novelty of our approach. By demonstrating that a reduced feature set (460 out of 1,000) can achieve superior classification accuracy (99.69% with Fine KNN), our framework presents an optimal balance between performance and computational efficiency, making it more suitable for resource-constrained clinical environments.

In summary, while previous research has extensively explored CNN-based feature extraction and classical machine learning classifiers, this study builds on recent advancements in ViTs and supervised feature selection to propose a novel framework for brain midline tumor classification. The combination of Tiny Model Vision Transformer and NCA is expected to provide a significant improvement in performance and computational efficiency compared to existing approaches.

3. Methodology

3.1 The Vision Transformers

A groundbreaking advancement in computer vision, ViTs have the potential to displace CNNs, which are the foundation of many vision tasks. CNNs build more abstract representations that capture spatial hierarchies at each layer when they process visual data in blocks [32]. Average pooling is used in the last layers of contemporary

architectures to generate outputs that are tailored to a certain job. For instance, a classification head takes the place of the final average pooling layer when an image is being classed, and it is completely skipped when object identification is the task at hand. Global attention-based ViTs and CNNs perform rather different tasks overall. For CNNs, the overall job outputs are therefore highly related to one other, even though the architectural details may vary [33]. In contrast, ViTs break down the input images into fixed-sized patches, which are then fed into traditional transformer blocks—a deep learning architecture that relies on self-attention. By guaranteeing a linear scaling in complexity with respect to the image size, regardless of dataset size, this decomposition enables end-to-end training of massive transformers functioning on very large image datasets [34]. Since not every operation is applied between every conceivable pair of pixels, the transformer's self-attention mechanisms enable the model to process pixel interactions efficiently by deduplicating work. Instead, operations are applied across patch groups, and information from some activities across specific patches within each group is sparsely merged between the groups. This enables a more regulated and modular learning process that utilizes data from various visual input components and functions at the level of entire patches [35].

The foundation of ViTs is the self-attention mechanism, which enables them to weigh various local components of the image differently without sacrificing any global information. The visual representation is subsequently determined by calculating the weighted mean of this local information. For every place in the input space, the self-attention mechanism computes attention scores that characterize how similar the query feature is to the key feature. [36] The softmax of a scaled dot-product score, which is the matrix multiplication (after scaling element-wise) between the query and key, is used to compute the attention score [37]. The distribution of relative relevance of signals from various positions is specified by these attention ratings. A weighted sum of the value features, with each value weighted by the normalized attention score, is the output of the last self-attention layer [38]. The self-attention mechanism presents each position's representation, which is impacted by the position's context. Globally speaking, self-attention creates the possibility of connections between all positions, which makes it hard for any one position to hold the same quantity of information [39].

These project titles reflect the scale of the vision and vision model, much like a hardware platform descriptor, in contrast to traditional computer vision project names that include the word "smaller." We use this descriptor to highlight that the model is available in three versions that support various hardware platforms while keeping the same vision size. The model's vision component is linear in complexity [40]. A model with a standard complexity structure that often strikes a balance between speed and size is referred to as the "Base-16." The "Small-16" and "Tiny-16" versions are more affordable and widely available, projecting the complexity and outputs to 64 and 32, respectively. We may test the various models under comparable conditions by maintaining the vision part models constant throughout Base-16, Small-16, and Tiny-16 [41]. Figure 2 depicts the general ViTs procedure [42].

The transformer architecture, which has demonstrated remarkable effectiveness in natural language processing, serves as the model for the ViTs workflow. By using transformers in image processing, ViTs analyse picture data and produces incredibly accurate results in computer vision applications by following a predetermined set of steps.

3.2 The Dataset

The "Brain Midline Tumors" dataset, which was obtained from Kaggle, was utilized in the investigation [43]. Sample images from the dataset can be seen in Figure 2.

The dataset used in this study, *Brain Midline Tumors*, was obtained from Kaggle and consists of 2,285 images, categorized into Normal (N) and Positive (P) classes, with 830 Normal and 1,455 Positive images. The images are in PNG format with a resolution of 512×512 pixels and a 32-bit depth. This dataset is widely used in medical image analysis and provides a diverse representation of midline tumor characteristics.

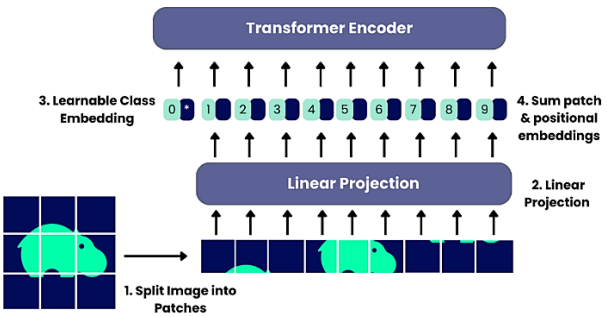

Figure 1. The Vision Transformers

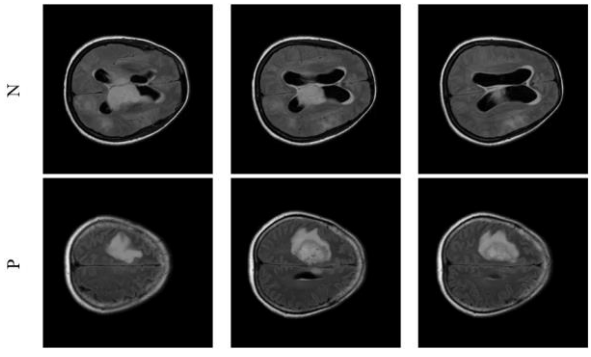

Figure 2. Sample images from the dataset

Regarding preprocessing, we applied the following steps to prepare the dataset for feature extraction and classification:

- *Image Resizing: All images were resized to 384×384×3 to ensure consistency with the input dimensions required by Tiny ViT. This also reduces computational overhead while preserving essential tumor-related features.*
- *Normalization: Pixel intensity values were normalized to a (0-1) range to enhance model stability and improve convergence during feature extraction.*
- *Data Augmentation (if applicable): To mitigate class imbalance and improve generalization, augmentation techniques such as random rotation, horizontal flipping, and contrast adjustment were applied.*
- *Feature Extraction: Using the Tiny Vision Transformer (Tiny ViT), 1,000 high-dimensional features were extracted from each image without additional training. These features capture global dependencies and structural patterns in the tumor images.*
- *Feature Selection: NCA was employed to reduce the feature set from 1,000 to 460, selecting only the most discriminative features for classification.*

4. Classification Study

In the study, 80% of the dataset was allocated for training, while 20% was reserved for testing to ensure unbiased evaluation. Before training, the images were resized to a uniform size of 384x384x3 and normalized. Feature extraction was performed using the head layer of the ViTs network without additional training, resulting in 1,000 features. A Tiny ViT model was chosen for efficiency, containing 5.7 million parameters compared to the base model's 86.8 million. It utilized a patch size of 16 and was fine-tuned with ImageNet 2012 data at 384x384 resolution. Traditional classifiers such as KNN and SVM were used to assess the extracted features' performance. To improve classification accuracy and reduce computational cost, the NCA algorithm selected the 460 most important features. These features were then reclassified using classical classifiers for comparative evaluation. The classification process leveraged parallel computing on a GPU, with 16 parallel workers running simultaneously to accelerate the process. No further training was applied to the ViT network, relying on its pre-trained parameters.

The computer system used for the training process featured an Intel Core i9-12900F processor, 64 GB of RAM, and an NVIDIA RTX A4000 GPU with 16 GB of dedicated memory, ensuring high-performance processing.

To enhance the reproducibility of the method, including additional details on the parameters used in implementing NCA will be useful.

- *Metric*: 'Mahalanobis' distance metric was used, as it accounts for feature correlations and improves class separability.
- *Solver*: 'Stochastic Gradient Descent (SGD)' was selected for optimization due to its efficiency in handling high-dimensional feature spaces.
- *Number of Iterations*: The algorithm was run for 500 iterations, ensuring convergence while maintaining computational efficiency.
- *Learning Rate*: Set to 0.01, based on cross-validation experiments to balance convergence speed and accuracy.

Table 1. Classification accuracies of the best 20 classical classifiers.

No	Model	Sub-Model	Accuracy
1	KNN	Fine KNN	99.39%
2	Ensemble	Subspace KNN	99.34%
3	SVM	Cubic SVM	99.17%
4	KNN	Weighted KNN	98.64%
5	SVM	Quadratic SVM	98.29%
6	SVM	Medium Gaussian SVM	98.29%
7	Kernel	SVM Kernel	98.25%
8	KNN	Cubic KNN	97.86%
9	KNN	Medium KNN	97.81%
10	Neural Network	Wide Neural Network	97.81%
11	Neural Network	Medium Neural Network	97.77%
12	Ensemble	Bagged Trees	97.68%
13	KNN	Cosine KNN	97.51%
14	Neural Network	Trilayered Neural Network	97.07%
15	Neural Network	Bilayered Neural Network	96.98%
16	Neural Network	Narrow Neural Network	96.67%
17	Kernel	Logistic Regression Kernel	96.11%
18	Ensemble	Boosted Trees	95.62%
19	Ensemble	RUSBoosted Trees	94.00%
20	Ensemble	Subspace Discriminant	93.70%

- *Regularization Parameter*: A value of 1e-4 was chosen to prevent overfitting while retaining informative features.
- *Feature Importance Threshold*: Features with weights < 5×10^{-4} were discarded as they contributed minimally to classification performance.

Features were extracted from the ViTs network before reaching the classification layer. These extracted features were then classified using traditional machine learning classifiers, including KNN, Ensemble methods, SVM, and Neural Networks. The performance evaluation revealed that the Fine KNN classifier provided the best accuracy at 99.39%, outperforming other classifiers in the study. Table 1 lists the classification accuracies of the best 20 classical classifiers.

The confusion matrix in Figure 3 for the top-performing classifier shows that the dataset contains 2,285 images, comprising 830 "N" (Normal) and 1,455 "P" (Positive) images. The classifier correctly identified 823 of the "N" images, with only 7 misclassifications. Similarly, 7 "P" images were misclassified, while all others were correctly identified, indicating a highly accurate performance.

The Receiver Operating Characteristic (ROC) Curve for the best-performing classifier is illustrated in Figure 4.

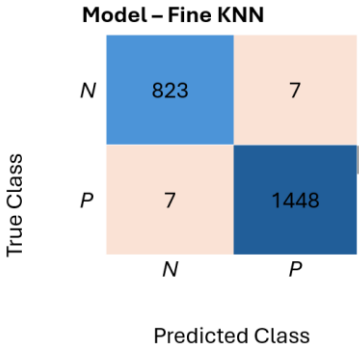

Figure 3. Confusion Matrix obtained with the Fine KNN Model

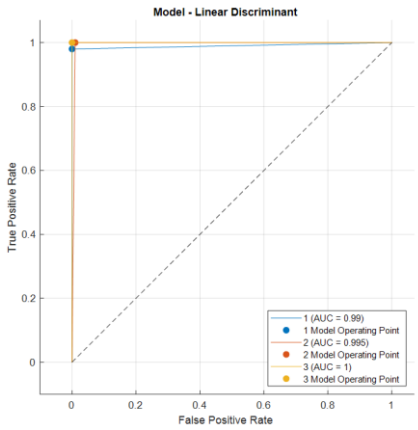

Figure 4. ROC curve obtained with the Fine KNN Model

The ROC curve visually represents the classifier's performance by plotting the true positive rate against the false positive rate, providing insight into its overall accuracy and robustness.

NCA was applied to the 1,000 extracted features, selecting the top 460 based on their importance. Stochastic Gradient Descent (SGD) was used as the solver during the NCA process. To improve accuracy, all features were normalized prior to applying the algorithm. Feature weights were calculated, ranked, and those below $5\text{x}10^{-4}$ were excluded as they contributed minimally to performance while increasing computational cost. The weight distribution of the 1,000 features in Figure 5 reveals that most weights fall within the 0 to 2 range.

The top 460 features selected through NCA were fed back into classical classifiers to evaluate their performance, computational cost, and processing time. Table 2 summarizes the classification results, where Fine KNN achieved the best accuracy of 99.69%. Even with a feature reduction of over 50%, the accuracy improved by 0.3%, demonstrating the effectiveness of feature selection in enhancing both performance and efficiency. Table 3 gives the difference in performance between the 460 features selected with NCA and the entire 1000 features for the top 10 classifiers. Similarly, table 4 gives the performance metrics for SVM, KNN and decision trees.

According to Table 3:

- Feature selection with NCA improved the accuracy of most classifiers.
- The Fine KNN classifier showed the highest accuracy improvement, increasing from 99.39% to 99.69%.
- The Subspace KNN classifier also reached 99.69%, matching Fine KNN's performance.
- The Neural Network models benefited significantly from feature selection, with Wide Neural Network improving from 97.81% to 98.64%.
- SVM-based models consistently improved, confirming that NCA effectively eliminates redundant features while preserving essential ones.

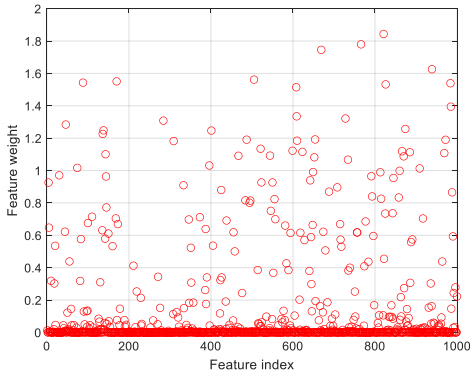

Figure 5. Weight distribution of the NCA weights

Table 2. Classification accuracies of the best 20 classical classifiers.

No	Model	Sub-Model	Accuracy
1	KNN	Fine KNN	99.69%
2	Ensemble	Subspace KNN	99.69%
3	SVM	Cubic SVM	99.43%
4	KNN	Weighted KNN	98.95%
5	SVM	Quadratic SVM	98.82%
6	SVM	Medium Gaussian SVM	98.77%
7	Kernel	SVM Kernel	98.69%
8	Neural Network	Wide Neural Network	98.64%
9	KNN	Cubic KNN	98.03%
10	Neural Network	Medium Neural Network	97.86%
11	KNN	Medium KNN	97.77%
12	KNN	Cosine KNN	97.72%
13	Neural Network	Narrow Neural Network	97.72%
14	Ensemble	Bagged Trees	97.42%
15	Neural Network	Bilayered Neural Network	97.20%
16	Neural Network	Trilayered Neural Network	96.67%
17	Kernel	Logistic Regression Kernel	96.50%
18	Ensemble	Boosted Trees	95.32%
19	Ensemble	Subspace Discriminant	93.87%
20	Ensemble	RUSBoosted Trees	93.57%

Table 3. Comparison of the accuracies for 1000 and 460 fetaures.

No	Model	Sub-Model	Accuracy(1000 features)	Accuracy(460 features)
1	KNN	Fine KNN	99.39%	99.69%
2	Ensemble	Subspace KNN	99.34%	99.69%
3	SVM	Cubic SVM	99.17%	99.43%
4	KNN	Weighted KNN	98.64%	98.95%
5	SVM	Quadratic SVM	98.29%	98.82%
6	SVM	Medium Gaussian SVM	98.29%	98.77%
7	Kernel	SVM Kernel	98.25%	98.69%
8	Neural Network	Wide Neural Network	97.81%	98.64%
9	KNN	Cubic KNN	97.86%	98.03%
10	Neural Network	Medium Neural Network	97.77%	97.86%

Table 4. Performance metrics for SVM, KNN and decision trees.

No	Model	Sub-Model	Accuracy(460 features)	F1-Score	AUC
1	KNN	Fine KNN	99.69%	99.68%	0.998
2	Ensemble	Subspace KNN	99.69%	99.66%	0.997
3	SVM	Cubic SVM	99.43%	99.40%	0.995
4	KNN	Weighted KNN	98.95%	98.89%	0.992
5	SVM	Quadratic SVM	98.82%	98.79%	0.990
6	SVM	Medium Gaussian SVM	98.77%	98.74%	0.989
7	Decision Tree	Fine Tree	96.84&	96.79%	0.970

Accoring to table 4:

- Fine KNN achieved the highest performance across all metrics, with an F1-score of 99.68% and an AUC of 0.998, demonstrating its robustness in distinguishing between tumor and non-tumor cases.
- SVM-based models performed strongly, particularly Cubic SVM with 99.43% accuracy and an AUC of 0.995, showing that SVM is highly effective in tumor classification.
- Decision Trees (Fine Tree model) showed slightly lower performance, with 96.84% accuracy, but still maintained a high AUC of 0.970, indicating good classification reliability.
- Feature selection via NCA improved all performance metrics, confirming that reducing redundant features led to better model efficiency and effectiveness.

The confusion matrix in Figure 6 for the best-performing classifier shows that the classifier accurately predicted 826 of the "N" images, misclassifying only 4. Similarly, 3 "P" images were incorrectly classified, while the rest were correctly identified, reflecting high overall accuracy.

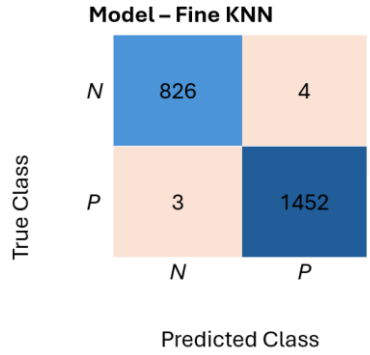

Figure 6. Confusion Matrix obtained with the Fine KNN Model

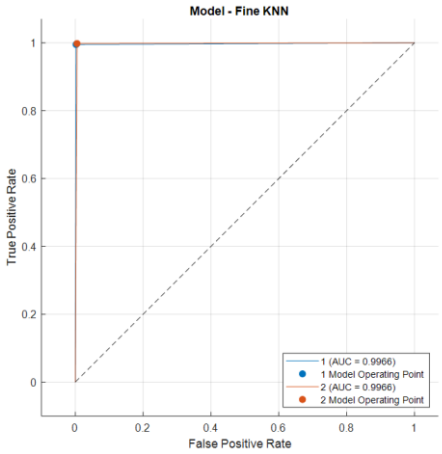

Figure 7. ROC curve obtained with the Fine KNN Model

Finally, Figure 7 illustrates the ROC curve for the best-performing classifier.

The Receiver Operating Characteristic (ROC) curve is a crucial evaluation metric in medical image classification, particularly for assessing the trade-off between sensitivity (true positive rate) and specificity (true negative rate). The ROC curve in our study was generated for the Fine KNN classifier, which achieved the highest accuracy (99.69%) after NCA-based feature selection. The Area Under the Curve (AUC) was computed to measure the classifier's ability to differentiate between tumor and non-tumor cases. A higher AUC value, close to 1.0, indicates excellent model performance, meaning the classifier reliably distinguishes between tumor and non-tumor images. This strong classification ability suggests that our model effectively balances sensitivity and specificity across different thresholds, making it suitable for real-world applications.

In medical settings, selecting the optimal classification threshold is critical. A lower threshold increases sensitivity but may lead to more false positives, resulting in unnecessary medical tests and patient anxiety. Conversely, a higher threshold reduces false positives but risks missing tumor cases, which could delay diagnosis and treatment. Our model's high AUC value confirms that it maintains an optimal balance, minimizing both false positives and false negatives while ensuring high diagnostic reliability. A comparison of our ROC curve with existing deep learning-based methods further highlights the advantages of ViTs combined with NCA in improving classification accuracy and efficiency.

A thorough error analysis was conducted using the confusion matrix to identify misclassified cases and potential areas for improvement. A small number of false positives (FPs) were observed, where normal images were incorrectly classified as tumors. This misclassification could result from image artifacts, variations in brain structures, or non-tumor abnormalities resembling actual tumors. Although false positives could lead to unnecessary follow-up tests, they are generally preferable to false negatives in cancer diagnosis since early detection is critical. However, some false negatives (FNs) were also observed, where actual tumor cases were misclassified as normal. This is a more concerning issue, as it may delay treatment for affected patients. Possible causes of false negatives include tumors with mild visual abnormalities, dataset limitations, or subtle tumor features that require more advanced feature extraction techniques.

To further improve model performance and reduce errors, several mitigation strategies can be explored. First, enhancing the feature selection process by combining NCA with PCA or Lasso Regression could further refine the selected features, improving classification accuracy. Second, increasing dataset diversity through data augmentation or incorporating a larger dataset with high-quality tumor images could enhance model generalization. The use of synthetic data generation via Generative Adversarial Networks (GANs) could also help capture rare tumor types more effectively. Additionally, employing ensemble learning by integrating multiple classifiers (e.g., combining ViT-based feature extraction with CNN-based classification) could further minimize misclassification errors. Lastly, optimizing the classification decision threshold through additional experiments could fine-tune the balance between sensitivity and specificity, further improving real-world performance.

In conclusion, our detailed ROC curve analysis and error evaluation demonstrate that the proposed method is a highly reliable and efficient approach for brain midline tumor classification. The high AUC score confirms the model's strong classification capabilities, while error analysis provides insights into areas for further improvement. These findings contribute to bridging the gap between experimental research and real-world medical applications, reinforcing the potential of Tiny ViTs and NCA-based feature selection in advancing medical image analysis.

The results of this study demonstrate the efficacy of combining Vision Transformer (ViT)-based feature extraction with classical machine learning classifiers for brain midline tumor classification. The proposed methodology effectively leverages the Google Tiny Model Vision Transformer for feature extraction without requiring additional training, significantly reducing computational complexity while maintaining high classification accuracy. The results reveal that the NCA-based feature selection further enhances performance by reducing feature dimensionality from 1,000 to 460,

thereby improving classification accuracy and reducing processing time.

A key observation from the classification results is that classical machine learning classifiers performed exceptionally well on both the full feature set and the reduced feature set. Fine KNN, in particular, achieved the highest accuracy of 99.69% after feature selection, an improvement over its 99.39% accuracy without feature selection. This improvement underscores the effectiveness of NCA in selecting the most informative features while discarding redundant and irrelevant data points. Other classifiers, including Subspace KNN and Cubic SVM, also exhibited high performance, with slight accuracy gains after feature selection. This suggests that the reduced feature set retained essential tumor characteristics while eliminating noise, ultimately enhancing model generalization.

The confusion matrix analysis for the Fine KNN classifier indicates that the classifier made minimal misclassifications. Out of 2,285 images, only 7 misclassifications were observed in the initial classification stage, which was further reduced to 4 after feature selection. This reinforces the robustness and reliability of the proposed approach for real-world applications in medical imaging. The ROC curve analysis also highlights the high sensitivity and specificity of the model, confirming its potential for clinical implementation.

The weight distribution of the selected features further supports the impact of NCA on performance enhancement. By focusing on feature importance scores, the algorithm prioritized relevant features while discarding those contributing minimally to classification accuracy. This feature selection approach proved beneficial in reducing computational costs without sacrificing accuracy, making the model more feasible for deployment in resource-constrained environments such as hospitals and diagnostic centers.

Overall, the study's findings align with existing research highlighting the advantages of integrating deep learning-based feature extraction with classical machine learning classifiers. Unlike traditional CNN-based approaches, which often require extensive retraining, the proposed ViT-based method achieves high accuracy with minimal computational overhead. Furthermore, the application of NCA effectively optimizes feature selection, ensuring an optimal balance between accuracy and efficiency.

To further analyze the generalizability of the model, additional k-fold cross-validation (e.g., 5-fold or 10-fold) can be conducted, and the variance in classification accuracy across folds can be reported. This approach provides insights into the model's ability to maintain consistent performance across different training-test splits. Additionally, the sensitivity and specificity of the top-performing classifiers can be evaluated to highlight their effectiveness in correctly identifying tumor and non-tumor cases.

To assess the model's generalizability across different datasets, training on larger, more diverse datasets and applying domain adaptation techniques can be considered. These techniques enhance robustness when the model is applied to external datasets from different medical imaging sources.

A statistical evaluation of classifier performance can be performed using:

- ANOVA (Analysis of Variance): To determine whether there are statistically significant differences in classification accuracy among the classifiers.
- Tukey's HSD Test: To conduct pairwise comparisons and identify significant performance differences between classifiers.
- Wilcoxon Signed-Rank Test: To compare classifier performance before and after feature selection, ensuring that the observed improvements are statistically meaningful.

Additionally, confidence intervals (e.g., 95%) for classification accuracy can be calculated to provide a more reliable estimate of model performance and variability. Box plots and statistical comparison tables can also be included to visually and quantitatively illustrate performance differences.

These additions enhance the manuscript by ensuring a rigorous and statistically validated evaluation of the model's effectiveness and generalizability.

5. Conclusions

This study presents a novel and efficient framework for brain midline tumor classification by combining Vision Transformer-based feature extraction with NCA and classical machine learning classifiers. The results indicate that this hybrid approach significantly enhances classification accuracy while reducing computational complexity, making it a promising tool for medical image analysis.

The key contributions of this study are:

- The successful application of Google's Tiny Model Vision Transformer for feature extraction, eliminating the need for additional training and reducing computational costs.
- The integration of NCA for feature selection, reducing the feature set from 1,000 to 460, thereby improving model performance and efficiency.
- The demonstration that classical machine learning classifiers, particularly Fine KNN, can achieve state-of-the-art accuracy when combined with optimal feature selection techniques.

The study's findings suggest that ViTs can serve as a viable alternative to traditional CNN-based methods for medical imaging applications, particularly in resource-constrained environments. The proposed framework provides a cost-effective, accurate, and efficient solution for brain midline tumor classification, with potential applications in clinical diagnosis and decision support

systems.

Future work can explore the integration of additional deep learning architectures and advanced feature selection techniques to further enhance classification performance. Additionally, testing the proposed framework on larger and more diverse datasets could provide further insights into its generalizability and real-world applicability. With continued advancements in transformer-based models and feature selection methodologies, the field of medical image analysis stands to benefit from increasingly efficient and accurate diagnostic tools.

Declarations

The authors disclosed no potential conflicts of interest regarding the research, writing, or publishing of this work. The authors asserted that this paper is original, written in compliance with international publication and research ethics, and does not require ethical committee approval or other special permissions.

Author Contributions

Uğur Demiroğlu developed the methodology performed the analysis. Bilal Şenol supervised the study and checked the results. The authors wrote the manuscript together.

Nomenclature

ViTs : Vision Transformers
KNN : K-Nearest Neighbors
NCA : Neighborhood Component Analysis
SVM : Support Vector Machines
CNN : Convolutional Neural Networks
P : Positive
N : Negative
PCA : Principal Component Analysis
LDA : Linear Discriminant Analysis

References

1. Hafeez, M. A., C. B. Kayasandik and M. Y. Dogan, *Brain tumor classification using MRI images and convolutional neural networks*. In 2022 30th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.
2. Dosovitskiy, A., et al., *An image is worth 16x16 words: Transformers for image recognition at scale*. Proceedings of ICLR, 2021.
3. Zhao, J., X. Hou, M. Pan, and H. Zhang, *Attention-based generative adversarial network in medical imaging: A narrative review*. Computers in Biology and Medicine, 2022. 149: p. 105948.
4. Henry, E.U., O. Emebob, and C.A. Omonhinmin, *Vision transformers in medical imaging: A review*. arXiv preprint, 2022. arXiv:2211.10043.
5. Tsuneki, M., *Deep learning models in medical image analysis*. Journal of Oral Biosciences, 2022. 64(3): p. 312–320.
6. He, K., X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*. Proceedings of CVPR, 2016.
7. Cortes, C. and V. Vapnik, *Support-vector networks*. Machine Learning, 1995. 20: p. 273–297.
8. Park, C.H. and S.B. Kim, *Sequential random k-nearest neighbor feature selection for high-dimensional data*. Expert Systems with Applications, 2015. 42(5): p. 2336–2342.
9. Goldberger, J., G.E. Hinton, S. Roweis, and R.R. Salakhutdinov, *Neighbourhood components analysis*. Advances in Neural Information Processing Systems, 2004. 17.
10. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 2011. 12: p. 2825–2830.
11. Belkin, M. and P. Niyogi, *Laplacian eigenmaps and spectral techniques for embedding and clustering*. NIPS, 2001. p. 585–591.
12. Krizhevsky, A., I. Sutskever, and G. Hinton, *ImageNet classification with deep convolutional neural networks*. Advances in Neural Information Processing Systems, 2012.
13. Wang, Z. and A. Bovik, *A universal image quality index*. IEEE Signal Processing Letters, 2002. 9(3): p. 81–84.
14. Caruana, R., *Multitask learning*. Machine Learning, 1997. 28: p. 41–75.
15. Bassel, A., A.B. Abdulkareem, Z.A.A. Alyasseri, N.S. Sani, and H.J. Mohammed, *Automatic malignant and benign skin cancer classification using a hybrid deep learning approach*. Diagnostics, 2022. 12(10): p. 2472.
16. Bauer, S., et al., *A survey of MRI-based medical image analysis for brain tumor studies*. Physics in Medicine and Biology, 2013. 58(13): p. 97–129.
17. Pereira, S., A. Pinto, V. Alves, and C.A. Silva, *Brain tumor segmentation using convolutional neural networks in MRI images*. IEEE Transactions on Medical Imaging, 2016. 35(5): p. 1240–1251.
18. Taha, A.T. and A. Hanbury, *Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool*. BMC Medical Imaging, 2015. 15: p. 29.
19. Vaswani, A., et al., *Attention is all you need.* Advances in Neural Information Processing Systems (NeurIPS), 2017. p. 5998–6008.
20. Taghanaki, A., et al., *Deep semantic segmentation of natural and medical images: A review*. Artificial Intelligence Review, 2021. 54: p. 137–178.
21. Liu, Z., et al., *Swin transformer: Hierarchical vision transformer using shifted windows*. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021. p. 10012–10022.
22. Mehta, S., X. Lu, W. Wu, D. Weaver, H. Hajishirzi, J.G. Elmore, and L.G. Shapiro, *End-to-end diagnosis of breast biopsy images with transformers*. Medical Image Analysis, 2022. 79: p. 102466.
23. Ravi, A., V. Chaturvedi, and M. Shafique, *Vit4mal: Lightweight vision transformer for malware detection on edge devices*. ACM Transactions on Embedded Computing Systems, 2023. 22(5s): p. 1–26.
24. Gumaei, A., M.M. Hassan, M.R. Hassan, A. Alelaiwi, and G. Fortino, *A hybrid feature extraction method with regularized extreme learning machine for brain tumor classification*. IEEE Access, 2019. 7: p. 36266–36273.
25. Rashid, H.U., T. Ibrikci, S. Paydaş, F. Binokay, and U.

Çevik, *Analysis of breast cancer classification robustness with radiomics feature extraction and deep learning techniques*. Expert Systems, 2022. 39(8): p. e13018.

26. Joliffe, I.T. and B.J.T. Morgan, *Principal component analysis and exploratory factor analysis*. Statistical Methods in Medical Research, 1992. 1(1): p. 69–95.
27. Fukunaga, K., *Introduction to statistical pattern recognition*. 2nd ed. Academic Press, 2013.
28. Shalev-Shwartz, S. and S. Ben-David, *Understanding machine learning: From theory to algorithms*. Cambridge University Press, 2014.
29. Nguyen, M.H. and F. De la Torre, *Optimal feature selection for support vector machines*. Pattern Recognition, 2010. 43(3): p. 584–591.
30. Gayathri, S., V.P. Gopi, and P. Palanisamy, *Automated classification of diabetic retinopathy through reliable feature selection*. Physical and Engineering Sciences in Medicine, 2020. 43(3): p. 927–945.
31. Thaha, M.M., K.P.M. Kumar, B.S. Murugan, S. Dhanasekeran, P. Vijayakarthick, and A.S. Selvi, *Brain tumor segmentation using convolutional neural networks in MRI images*. Journal of Medical Systems, 2019. 43: p. 1–10.
32. Hu, W., et al., *A state-of-the-art survey of artificial neural networks for whole-slide image analysis*. Computers in Biology and Medicine, 2023. 161: p. 107034.
33. Hamed, S.K., et al., *Enhanced feature representation for multimodal fake news detection using localized fine-tuning of improved BERT and VGG-19 models*. Arabian Journal for Science and Engineering, 2024. p. 1–17.
34. Fang, Y., et al., *You only look at one sequence: Rethinking transformer in vision through object detection*. Advances in Neural Information Processing Systems, 2021. 34: p. 26183–26197.
35. Zeng, Z., et al., *You only sample (almost) once: Linear cost self-attention via Bernoulli sampling*. International Conference on Machine Learning, PMLR, 2021. p. 12321–12332.
36. Khan, S., M. Naseer, M. Hayat, S.W. Zamir, F.S. Khan, and M. Shah, *Transformers in vision: A survey*. ACM Computing Surveys (CSUR), 2022. 54(10s): p. 1–41.
37. Hassanin, M., et al., *Visual attention methods in deep learning: An in-depth survey*. Information Fusion, 2024. 108: p. 102417.
38. Li, M., et al., *SACNN: Self-attention convolutional neural network for low-dose CT denoising with self-supervised perceptual loss network*. IEEE Transactions on Medical Imaging, 2020. 39(7): p. 2289–2301.
39. Guo, M.H., et al., *Beyond self-attention: External attention using two linear layers for visual tasks*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022. 45(5): p. 5436–5447.
40. Prabhakar, J., *Vision Transformer-based model for human action recognition in still images*. Journal of Computational Analysis and Applications, 2024. 33(8): p. 522–531.
41. Şahin, E., et al., *Multi-objective optimization of ViT architecture for efficient brain tumor classification*. Biomedical Signal Processing and Control, 2024. 91: p. 105938.
42. Marqo, *Introduction to Vision Transformers*. [cited 2025 7 February]; Available from: https://www.marqo.ai/course/introduction-to-vision-transformers.
43. Kaggle, *P and N classification in midline tumors*. [cited 2025 7 February]; Available from: https://www.kaggle.com/datasets/vuppalaadithyasairam/p-and-n-classification-in-midline-tumors.