



Keros-Net: a convolutional block attention module–squeeze-and-excitation-integrated hybrid learning framework for olfactory fossa depth classification

Zülküf Akdemir¹
 Murat Canayaz²
 Abdulaziz Yalınkılıç³
 Suat Orak⁴

¹Van Yüzüncü Yıl University Faculty of Medicine,
Department of Radiology, Van, Türkiye

²Van Yüzüncü Yıl University Faculty of Engineering,
Department of Computer Engineering, Van, Türkiye

³Van Yüzüncü Yıl University Faculty of Medicine,
Department of Otorhinolaryngology, Van, Türkiye

⁴Lokman Hekim Van Hospital, Clinic of Radiology,
Van, Türkiye

PURPOSE

This study aimed to develop a hybrid decision support framework combining deep learning (DL) and machine learning (ML) to automatically classify olfactory fossa depth on paranasal computed tomography (CT) images according to the Keros classification. The goal was to enhance accuracy, reduce observer variability, and support safer endoscopic sinus surgery.

METHODS

A retrospective dataset of 481 individuals (1,549 cropped coronal CT slices) was analyzed and labeled as Keros types I–III by an experienced radiologist. Deep features were extracted using DenseNet121, DenseNet169, and DenseNet201 architectures enhanced with convolutional block attention modules (CBAMs) and squeeze-and-excitation (SE) blocks. Four feature selection techniques—SHapley Additive exPlanations, recursive feature elimination (RFE), principal component analysis, and SelectKBest—were applied to reduce dimensionality. Selected features were classified using support vector machines (SVMs), random forest, XGBoost, logistic regression, and Naive Bayes. Five-fold cross-validation was used to assess accuracy, precision, recall, and F1-score metrics.

RESULTS

Among the baseline models, DenseNet169 achieved the highest accuracy (88.37%). After feature selection, the RFE + SVM and RFE + logistic regression combinations yielded the best performance with an accuracy of 97.90%, demonstrating a substantial improvement over DL models alone. The most effective feature selection technique was RFE, and SVMs consistently produced well-balanced classification results.

CONCLUSION

Integrating CBAM–SE-enhanced DenseNet architectures with optimized feature selection and classic ML classifiers enables highly accurate and reliable automatic classification of Keros types. The proposed hybrid approach outperforms conventional DL models and provides a robust framework for objective radiological assessment.

CLINICAL SIGNIFICANCE

Accurate preoperative identification of olfactory fossa depth is essential for preventing complications such as cribriform plate injury and cerebrospinal fluid leakage during endoscopic sinus surgery. The proposed system offers an efficient, reproducible, and objective tool that may enhance surgical planning, reduce operator dependency, and increase patient safety.

KEYWORDS

Keros classification, olfactory fossa, computed tomography, deep learning, convolutional block attention modules, squeeze-and-excitation blocks, feature selection, machine learning

Handling editor: Tuğba Akıncı D'Antonoli

Corresponding author: Zülküf Akdemir

E-mail: za.radyoloji@hotmail.com

Received 20 November 2025; revision requested 16
January 2026; accepted 25 February 2026.



Epub: 18.03.2026

Publication date: 01.09.2026

DOI: 10.4274/dir.2026.263749

You may cite this article as: Akdemir Z, Canayaz M, Yalınkılıç A, Orak S. Keros-Net: a convolutional block attention module–squeeze-and-excitation-integrated hybrid learning framework for olfactory fossa depth classification. *Diagn Interv Radiol.* 2026;32(5):600-607.

Endoscopic surgery of the sinonasal region has become increasingly common in recent years. Because of its close relationship with the skull base and neighboring structures, this area carries considerable risk in terms of anatomical variations and complications. Surrounding landmarks include the anterior skull base, cribriform plate, lamina papyracea, optic nerve, and internal carotid artery. To reduce complications, surgeons must have detailed anatomical knowledge and make effective use of advanced imaging techniques.^{1,2} Thus, radiological assessment before and during surgery plays a key role in identifying complex variations and localizing pathology with precision.

Among the important structures in this region, the anterior ethmoidal artery (AEA) is particularly vulnerable during anterior skull base procedures.³ Injury to the AEA may lead to severe outcomes, including epistaxis, vision loss, intracranial hemorrhage, or cerebrospinal fluid (CSF) leakage.⁴ Therefore, careful evaluation of anatomical landmarks and classification of the AEA's course are crucial for improving surgical safety.⁵⁻⁷

The Keros classification is especially relevant in this context because it describes the depth between the ethmoid roof and the cribriform plate: type I: 1–3 mm, type II: 4–7 mm, and type III: 8–16 mm. Type III is considered the most critical variation, as a deeper cribriform plate position increases the risk of serious complications.⁸ Preoperative assessment using this classification helps reduce the likelihood of intracranial penetration and supports safer, more controlled interventions.^{3,9-13}

Main points

- This study introduces the first hybrid framework that integrates convolutional block attention module- and squeeze-and-excitation block-enhanced DenseNet architectures with multiple feature selection strategies for automatic Keros classification from paranasal computed tomography scans.
- The proposed model achieved a substantial performance improvement, increasing baseline DenseNet accuracy from 83%–88% to up to 97.9% using recursive feature elimination combined with support vector machines or logistic regression.
- By reducing inter-observer variability and providing objective, automated assessment of olfactory fossa depth, the framework offers strong potential as a clinical decision support tool for safer endoscopic sinus surgery.

Computed tomography (CT) of the paranasal sinuses is routinely used not only for evaluating sinonasal disease before surgery but also for mapping anatomical structures and their relationships, thereby improving surgical planning and reducing complications.⁹ Accurate classification of olfactory fossa depth remains essential for preventing complications such as cribriform plate damage and CSF leakage; however, manual interpretation of CT images can vary between radiologists and over time, introducing subjectivity into the assessment.

In parallel with these developments, artificial intelligence (AI) has rapidly gained importance in medicine. AI-based decision support systems are increasingly used in medical imaging to improve diagnostic accuracy, reduce intraoperative risks, and support physicians by easing workloads. Machine learning (ML) and deep learning (DL), particularly deep convolutional neural networks (CNNs), enable automatic feature extraction and robust classification from CT, magnetic resonance imaging, and ultrasound, facilitating more objective and reproducible imaging-based assessments.

Motivations

Given that manual Keros assessment is time-consuming and subject to inter-observer variability—especially in cases with subtle anatomical differences—there is a clear need for automatic, objective, and reproducible classification methods. In this context, AI offers an opportunity to reduce subjectivity, increase efficiency, and, ultimately, improve surgical outcomes. Our motivation was to address these challenges by developing a reliable decision support framework that can integrate seamlessly into clinical practice.

Contributions

To the best of our knowledge, this is the first study to apply a hybrid DL and ML approach for automatic olfactory fossa depth classification based on the Keros system. We propose DenseNet architectures enhanced with convolutional block attention modules (CBAMs) and squeeze-and-excitation (SE) blocks for deep feature extraction, followed by feature optimization using SHapley Additive exPlanations (SHAP), recursive feature elimination (RFE), principal component analysis (PCA), and SelectKBest and classification using supervised algorithms including support vector machines (SVMs), random forest, logistic regression, Naive Bayes, and XGBoost. This integrated pipeline achieved an

accuracy of 97.9% and demonstrated strong performance in anatomically critical regions, supporting its robustness and potential for clinical application as an explainable and efficient decision support tool for surgical planning in sinonasal disease.

Related research

In their review, Alsalama et al.¹⁴ provided an overview of imaging and modeling techniques based on paranasal sinus anatomy, emphasizing automatic segmentation and classification approaches. They highlighted that sinus morphology can support individual identification and demographic profiling (e.g., age and sex) when conventional methods are unavailable and discussed the role of three-dimensional (3D) modeling in predicting demographic traits such as age, sex, and ethnicity, underlining the reliability of sinus structures for forensic identification and demographic inference.¹⁴

Wang et al.¹⁵ created a CT dataset of 242 patients with chronic rhinosinusitis with nasal polyps (CRSwNP) and applied a customized 3D nnU-Net v2 (Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany model for segmentation and deep feature extraction. They reported strong performance (Dice 92.8%, Intersection over Union 86.64%, accuracy 99.69%, precision 92.63%, recall 93.22%) while noting that the dataset consisted only of surgical CRSwNP cases, which may limit generalizability to broader clinical populations and across centers.¹⁵

In our study, we used high-resolution paranasal CT scans from a larger and more diverse population, with careful cropping to standardize inputs. To improve sensitivity to subtle or low-contrast regions, DenseNet was augmented with CBAMs and SE attention blocks, enabling more effective focus on anatomically subtle structures and improving classification performance across patient subgroups.

Similarly, Lee et al.¹⁶ proposed a CNN-based model to automatically compute the Lund–Mackay score (LMS) by segmenting sinus regions on CT and directly generating LMS values. Using 1,399 patients and manual segmentations of 77 scans for training, they achieved an average Dice score of 0.85, with LMS prediction accuracy across regions ranging from 86% to 99%, suggesting that automation can standardize scoring and support clinical decision-making.¹⁶

Beyond paranasal sinus analysis, Gogus et al.¹⁷ introduced a hybrid model combining CBAMs with Swin Transformer blocks for femoral stem implant classification, outperforming architectures such as DenseNet201, VGG19, and InceptionV3. Their MSFT-Net addressed inter-class morphological similarity through attention mechanisms and transformer layers, paralleling our approach in leveraging attention to improve feature extraction from complex medical images. In our framework, CBAMs and SE blocks similarly enhance focus on subtle regions such as the lateral lamella in paranasal CT, improving sensitivity and classification accuracy; together, these findings support the value of attention-enhanced designs for reliable performance in anatomically challenging scenarios.

Methods

We retrospectively reviewed the CT images of 4,427 patients who underwent non-contrast paranasal sinus scans between April 2020 and January 2024, retrieved from the Picture Archiving and Communication System. Based on the study criteria, a total of 481 individuals aged 18–67 years (mean age: 29.1 ± 10.5 years) were included in the analysis. Of these, 255 (53%) were men and 226 (47%) women.

Routine paranasal sinus CT examinations were performed with patients in the supine position, the head placed in hyperextension. Axial images were obtained using 64-slice SOMATOM go.Up and 128-slice SOMATOM Definition AS multidetector CT scanners (Siemens Healthineers, Erlangen, Germany). The imaging protocol was as follows; tube voltage: 100 kV, effective mAs: 67, slice thickness: 3.0 mm, field of view: 170 mm, pitch: 0.8–1 mm, rotation time: 1 s, and collimation: 128×0.6 mm. The bone window settings were WW: 2,200 and WL: 475. Reconstructions were performed in the coronal plane perpendicular to the hard palate, with 0.75-mm slice thickness, using the J70h very sharp bone filter as standard.

For evaluation, only coronal reformatted images were used. Classification followed the Keros system, which defines the depth of the olfactory fossa as type I (1–3 mm), type II (4–7 mm), and type III (8–16 mm). In cases of asymmetry, the side with greater depth was considered. Ultimately, 1,549 images were prepared, consisting of 330 for type I, 677 for type II, and 542 for type III. Images were saved in JPEG format, averaging 3.2 slices per patient.

All images were uniformly cropped according to predefined anatomical reference points. The superior border included the crista galli, the lateral borders encompassed the medial walls of the maxillary sinuses, and the inferior limit was set at the superior surface of the palatine process of the maxilla.

Inclusion criteria: Patients over 18 years of age who underwent paranasal sinus CT imaging.

Exclusion criteria: Cases with ethmoidal or frontal mucosal thickening, sinonasal polyps, poor positioning, image artifacts, inverted papilloma, septal or cribriform fractures, nasolacrimal duct pathologies, osteoma, foreign bodies, intranasal masses, crista galli lesions, septal perforation, prior surgery, or low-quality scans.

All image selection and cropping were performed by a radiologist with 14 years of experience.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

The de-identified participant data and analysis outputs are available from the corresponding author upon reasonable request and subject to institutional/ethical approvals. This study was approved by Van Yüzüncü Yıl University Non-Interventional Clinical Research Ethics Committee (protocol number: 2023/05-03, date: 12.05.2023).

Proposed approach

We developed a three-stage hybrid framework for the automatic classification of olfactory fossa depth. First, a labeled dataset of CT images was created. These images were then processed using a DenseNet-based fea-

ture extractor augmented with CBAMs and SE blocks to extract high-dimensional deep feature vectors. The CBAMs improved the model's ability to attend to spatial and channel-wise information, whereas the SE blocks rescaled feature channels to strengthen discriminative representations. The overall workflow of the proposed method is illustrated in Figure 1.

In the second stage, the extracted deep features were optimized using four different feature selection methods: SHAP, RFE, PCA, and SelectKBest. These methods identified the most meaningful subsets of features for classification and reduced model complexity. In the third stage, the selected features were fed into five different traditional classifiers (SVMs, random forest, XGBoost, logistic regression, and Naive Bayes). Each combination was tested using cross-validation and evaluated with metrics such as accuracy, precision, recall, and F1-score. Using this framework, high accuracy in Keros classification was achieved by combining the powerful feature extraction capabilities of DL with the flexibility of traditional ML methods.

Dataset and preprocessing

The image data used in this study consist of high-resolution CT scans acquired in the coronal plane of the paranasal sinuses, labeled as Keros types 1, 2, and 3 by an experienced radiologist. The data were obtained in anonymized form and used within the scope of the study after receiving ethical approval. Basic preprocessing steps such as normalization and resizing (e.g., to 256×256 pixels) were applied to prepare the images for classification. Some examples from the dataset are shown in Figure 2.

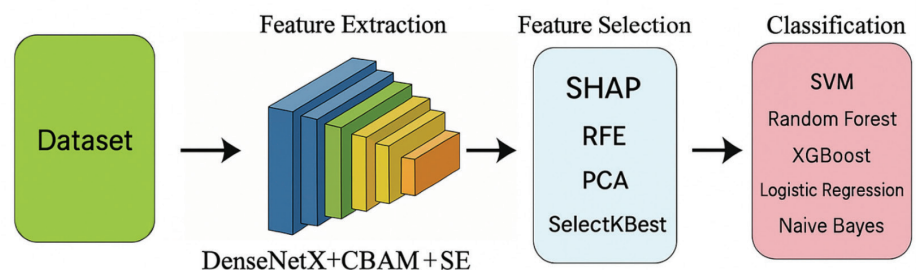


Figure 1. Structure of proposed model. Schematic overview of the proposed three-stage hybrid framework for automatic classification of olfactory fossa depth. A labeled computed tomography image dataset is first constructed and used as input to a DenseNet-based feature extractor augmented with CBAMs and SE blocks to obtain high-dimensional deep feature vectors. These features are then subjected to feature selection (SHAP, RFE, PCA, or SelectKBest), and the selected features are finally fed into machine learning classifiers (SVMs, random forest, XGBoost, logistic regression, or Naive Bayes) for final category prediction. CBAM, convolutional block attention module; SE block, squeeze-and-excitation block; RFE, recursive feature elimination; PCA, principal component analysis; SVM, support vector machine; SHAP, SHapley Additive exPlanations.

Deep feature extraction

In the feature extraction stage, the DenseNet architecture was used as the backbone. However, this base architecture was enhanced with attention mechanisms. The CBAMs optimized the attention distribution of the input images based on both spatial and channel features, whereas the SE block reinforced channel-wise key information. This design enabled the model to focus more on anatomically critical regions (e.g., the lateral lamella area). As output, a high-dimensional deep feature vector was obtained from each image.

Statistical analysis

Statistical analyses were conducted to evaluate the performance, reliability, and generalizability of the proposed hybrid classification framework. To prevent data leakage and avoid artificially inflated performance due to patient-specific anatomical similarity, all data partitioning procedures were performed at the patient level rather than at the slice level. All CT slices belonging to a given patient were assigned exclusively to either the training or testing set. An 80–20 stratified split was applied at the patient level to

preserve the distribution of Keros types, and model evaluation was performed using five-fold stratified cross-validation, ensuring strict patient-level independence across all folds. Hyperparameters for the classic ML classifiers were selected based on standard configurations and maintained across experiments. Model performance was evaluated using cross-validation on the training data. Due to the relatively limited dataset size and the primary focus on comparative performance evaluation, exhaustive hyperparameter optimization and nested cross-validation were not applied. Class imbalance was addressed at multiple stages of the pipeline. During DL model training, stratified data splitting was applied to preserve class proportions, and class weights were computed from the training labels using a balanced weighting scheme and incorporated into the loss function. For the classic ML classifiers trained on selected deep features, five-fold stratified cross-validation was employed to maintain class distribution across folds. In addition, weighted-averaged precision, recall, and F1-score metrics were reported to ensure balanced performance evaluation across all Keros types. For each experiment, accuracy, precision, recall, and F1-score were computed

and averaged across folds to obtain stable performance estimates.

To reduce the dimensionality of the deep feature vectors and enhance classifier performance, four statistical feature selection methods were applied: RFE, PCA, SelectKBest, and SHAP. RFE was used to iteratively eliminate less informative features, PCA reduced multicollinearity by transforming features into orthogonal principal components, SelectKBest ranked features based on statistical significance, and SHAP values were used to quantify feature importance and enhance model interpretability. The dimensionality of the deep feature vectors before and after each feature selection method is summarized in Table 1.

Comparative statistical evaluation was performed between baseline DL models and feature-selected ML classifiers by examining improvements in accuracy and metric consistency. Since the objective was predictive performance rather than inferential hypothesis testing, no parametric or non-parametric statistical tests were applied. All analyses were conducted using Python (v3.x), scikit-learn, SHAP library, XGBoost, NumPy, and related DL frameworks.

Machine learning classifiers

The selected features were evaluated using five classic classifiers with different characteristics:

- **Support vector machines:** Effective in constructing decision boundaries in high-dimensional spaces.
- **Random forest:** An ensemble model composed of decision trees, often successful in handling class imbalance.
- **XGBoost:** A tree-based gradient boosting method.
- **Logistic regression:** Creates linear decision boundaries and is commonly used as a baseline comparison model.
- **Naive Bayes:** A probabilistic model that assumes features are independent of each other.

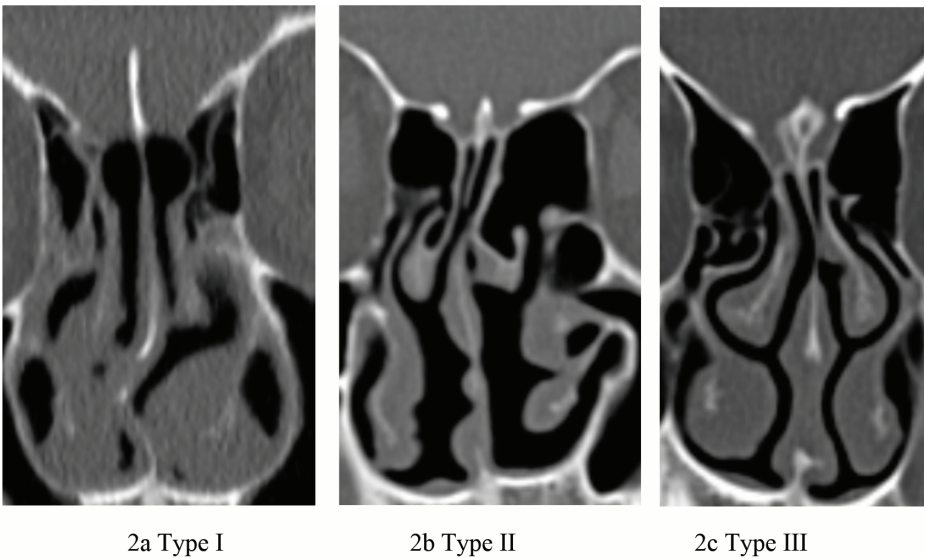


Figure 2. Representative examples of high-resolution coronal computed tomography images of the paranasal sinuses used in this study, labeled according to the Keros classification: (2a) type 1, (2b) type 2, and (2c) type 3.

Table 1. Feature dimensionality before and after feature selection						
Backbone (CBAMs + SE DenseNet)	Deep feature source layer	Before selection (n × d)	SelectKBest (k = 100)	PCA (n = 100)	RFE (n = 100)	SHAP (top 100)
DenseNet121	Dense(128) (penultimate layer)	n × 128	n × 100	N × 100	n × 100	n × 100
DenseNet169	Dense(128) (penultimate layer)	n × 128	n × 100	n × 100	n × 100	n × 100
DenseNet201	Dense(128) (penultimate layer)	n × 128	n × 100	n × 100	n × 100	n × 100
CBAM, convolutional block attention module; SE, squeeze-and-excitation; PCA, principal component analysis; RFE, recursive feature elimination; SHAP, SHapley Additive						
exPlanations.						

Findings and results

In the experimental findings of the study, classification was first performed directly using the DenseNet121, DenseNet169, and DenseNet201 architectures without applying any feature extraction or selection methods. The models were trained and tested for Keros classification (types 1, 2, and 3). Table 2 presents the comparative results of the baseline models in terms of accuracy, precision, recall, and F1-score metrics.

DenseNet169 achieved the highest performance across all metrics, making it the best baseline model. With an accuracy of 88.37%, it demonstrated a strong generalization ability in classifying olfactory fossa depths.

DenseNet201, with an accuracy of 86.05%, ranked second and exhibited competitive performance, particularly in terms of precision (0.8645).

DenseNet121 showed the lowest performance, with an accuracy of 83.33%. This result indicates that its shallower architecture limited its classification capacity.

The accuracy, precision, recall, and F1-score values of the models are similar. This indicates that the class distribution is balanced and that the models generally demonstrate consistent performance across classes. In particular, the minimal difference between metrics in the DenseNet169 model shows that it produces balanced and reliable outputs.

Second, deep features extracted using DenseNet architectures enhanced with CBAMs and SE blocks were optimized through different feature selection algorithms (RFE, PCA, SHAP, SelectKBest) and used for Keros classification with traditional ML classifiers (SVMs, logistic regression, Naive Bayes, etc.). Table 3 presents the comparative results of the models in terms of accuracy, precision, recall, and F1-score metrics.

DenseNet169 emerged as the highest-performing model. In particular, the combinations of RFE + SVMs and RFE + logistic regression achieved an accuracy of 97.90%, indicating a very high generalization capacity.

RFE proved to be the most effective feature selection method overall, achieving the highest accuracies with almost all models.

SVMs and logistic regression produced high and consistent results. The fact that the accuracy, precision, recall, and F1-Score values were similar indicates that the models performed balanced classification.

Comparative evaluation

Table 4 compares the accuracy values of only the baseline models (without feature selection or different classifiers) with those of the best feature selection + ML classifier combinations.

Models supported with feature selection and traditional classifiers achieved an accuracy improvement of 10%–14% compared with the baseline models.

This increase highlights that feature engineering and the choice of an appropriate classifier are critical, especially for models operating on complex and imbalanced classes.

The simplest architecture, DenseNet121, achieved 83.33% accuracy as a baseline model while reaching 97.05% accuracy with feature selection and SVMs. This demonstrates that even simpler models can deliver strong performance when supported with appropriate enhancement methods. The graphical representation of the results is presented in Figure 3.

These findings demonstrate that the integration of CBAMs + SE-enhanced deep feature extraction, feature selection, and ML classifiers is a powerful and effective approach for anatomically critical tasks that influence surgical decisions, such as Keros classification. In particular, the DenseNet169 + RFE + SVMs combination yielded the best results in terms of both accuracy and metric consistency.

Clinical usability and interface integration

To facilitate real-world clinical adoption, the proposed Keros classification model was deployed within a dedicated web-based interface designed for radiologists and endoscopic sinus surgeons. The interface, developed using the Flask framework, enables clinicians to upload cropped coronal CT slices of the olfactory fossa and instantly obtain Keros type I–III predictions with their corresponding probability estimates. The clean, task-oriented design supports multi-image uploading, automatic preprocessing, model selection, and real-time visualization of results.

The interface displays explanatory information regarding the clinical relevance of each Keros type, allowing the prediction to be contextualized alongside anatomical detail and potential surgical risk. This design contributes to decision support by present-

Table 2. Performance comparison of the baseline models

Model	Accuracy	Precision	Recall	F1 score
DenseNet121	0.8333	0.8342	0.8333	0.8326
DenseNet169	0.8837	0.8902	0.8837	0.8839
DenseNet201	0.8605	0.8645	0.8605	0.8606

The table summarizes the performance of the baseline DenseNet architectures for the classification task. DenseNet169 achieved the best overall performance among the baseline models (accuracy, 0.8837; precision, 0.8902; recall, 0.8837; F1 score, 0.8839). Accuracy, recall, precision, and F1-score are reported as proportions, with F1-score representing the harmonic mean of precision and recall.

Table 3. Performance comparison (feature-selected models)

Model	Feature selection	Classifier	Accuracy	Precision	Recall	F1 score
DenseNet169	RFE	SVM	0.9790	0.9792	0.9790	0.9791
DenseNet169	RFE	Logistic Regression	0.9790	0.9791	0.9790	0.9791
DenseNet169	PCA	SVM	0.9783	0.9784	0.9783	0.9783
DenseNet201	RFE	Naive Bayes	0.9775	0.9775	0.9775	0.9775
DenseNet201	SHAP	SVM	0.9767	0.9768	0.9767	0.9767

This table summarizes the performance of feature-selected models based on DenseNet169 and DenseNet201 backbones combined with different feature selection methods (RFE, PCA, SHAP) and classifiers (SVMs, logistic regression, Naive Bayes). The highest overall performance was achieved by RFE-based DenseNet169 models with SVMs and logistic regression (accuracy, 0.9790; precision, 0.9792 and 0.9791; recall, 0.9790; F1-score, 0.9791). Accuracy, precision, recall, and F1-score are reported as proportions, with F1-score representing the harmonic mean of precision and recall. RFE, recursive feature elimination; PCA, principal component analysis; SHAP, SHapley Additive exPlanations; SVM, support vector machine.

Table 4. Comparison of base and optimized classification accuracies for DenseNet models

Model	Base model accuracy	Best combination accuracy	Increase (%)
DenseNet121	0.8333	0.9705 (RFE + SVM)	+13.7
DenseNet169	0.8837	0.9790 (RFE + SVM)	+10.8
DenseNet201	0.8605	0.9775 (RFE + Naive Bayes)	+13.4

This table summarizes the effect of combining each DenseNet backbone with the best-performing feature selection and classifier configuration on classification accuracy. “Base Model Accuracy” indicates the performance of the convolutional neural network alone, without feature selection or an external classifier, whereas “Best Combination Accuracy” denotes the highest-performing feature selection–classifier pipeline for that backbone. “Increase (%)” indicates the percentage improvement in accuracy relative to the corresponding base model. Accuracy values are reported as proportions.

RFE, recursive feature elimination; SVM, support vector machine.

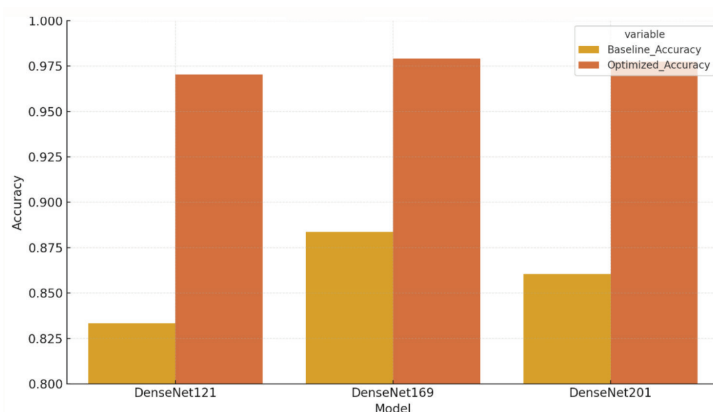


Figure 3. Comparison of baseline and optimized accuracy values for three different DenseNet architectures. The yellow bars indicate the baseline accuracy of the deep networks used without additional processing, and the orange bars represent the optimized accuracy obtained after applying feature selection and a machine learning classifier. The most prominent improvement is observed for DenseNet121, where the accuracy increases from 83.33% in the baseline model to 97.05% in the optimized configuration using feature selection and a support vector machine, indicating that even the simplest architecture can achieve high performance when appropriate enhancement methods are applied.

ing the model output together with radiological slice quality, patient history, and relevant anatomical variants. The system stores previously analyzed scans in a “History” module, enabling clinicians to revisit earlier evaluations for comparison or follow-up planning.

Importantly, the interface is optimized for routine clinical workflow without requiring graphics processing unit (GPU) hardware, as the deep feature extraction and model optimization steps are embedded within the pretrained hybrid model. This ensures rapid inference and seamless integration into standard radiology workstations. By providing reproducible, objective Keros grading with an intuitive user experience, the interface demonstrates strong potential for use in pre-operative planning, reducing observer variability, and improving surgical risk assessment. Taken together, this system represents a practical and clinician-friendly implementation of the hybrid classification framework, supporting its feasibility for real-world deployment in otolaryngology and radiology settings. A graphical representation of the interface is presented in Figure 4.

Discussion

In this study, we proposed a hybrid model to automate Keros classification of olfactory fossa depth, combining DenseNet with attention mechanisms and multiple feature selection techniques. Unlike earlier studies, our approach integrated CBAMs and SE blocks into the DenseNet framework, improving the model’s focus on critical anatomical regions such as the lateral lamella and strengthening the representational quality of the extracted features. By applying feature selection methods such as SHAP, RFE, PCA, and SelectKBest, we were able to filter deep features before feeding them into conventional ML algorithms. This not only boosted classification accuracy but also enhanced interpretability and computational efficiency. The strongest outcomes were achieved with the RFE + SVMs and RFE + logistic regression pipelines, reaching an accuracy of 97.9%, underlining the robustness of the chosen strategy and its clinical potential.

Beyond the accuracy values, the study contributes to the field by showing how au-

tomated approaches can support surgical planning. Traditional methods, which rely on manual interpretation, are time-consuming and subject to variation between observers, making the assessment of risky anatomical variations difficult. Our findings demonstrate that automated feature extraction and classification can help reduce surgical risks and improve safety. Moreover, the performance gain—rising from approximately 83%–88% in baseline DL models to nearly 98% when combined with feature selection and classic classifiers—is a substantial step forward in clinical applicability. These results confirm that model success does not depend only on network depth but also on careful feature engineering and algorithm choice.

In addition to the technical contributions of the proposed framework, we developed a web-based clinical interface to ensure its applicability in real-world clinical environments. This interface allows clinicians to upload cropped coronal CT slices and instantly receive Keros type I–III predictions along with probability estimates, enabling rapid and objective assessment. Its intuitive design, support for multiple image uploads, automated preprocessing, and integrated history management make it well suited for routine radiology workflows. Importantly, the system operates without requiring GPU hardware, as all deep feature extraction and optimization steps are embedded within the pretrained hybrid model. By reducing observer variability and providing reproducible grading of olfactory fossa depth, the interface highlights the clinical relevance and practical translational potential of the proposed hybrid decision support system.

Nevertheless, several limitations should be acknowledged. First, image labeling and cropping were performed by a single radiologist with 14 years of experience. Although this ensured consistent anatomical localization and annotation, it may introduce potential observer bias, as inter-observer variability is a known issue in Keros classification. Due to the retrospective nature of the dataset and institutional constraints, the inclusion of a second independent rater and quantitative inter-rater agreement analysis was not feasible. Future studies will aim to incorporate multi-reader annotations and inter-observer agreement metrics, such as Cohen’s kappa, to further strengthen reliability. Second, all CT images were obtained from a single center using Siemens scanners with relatively uniform acquisition parameters. Although this consistency reduces variability related to imaging protocols, it may limit the general-

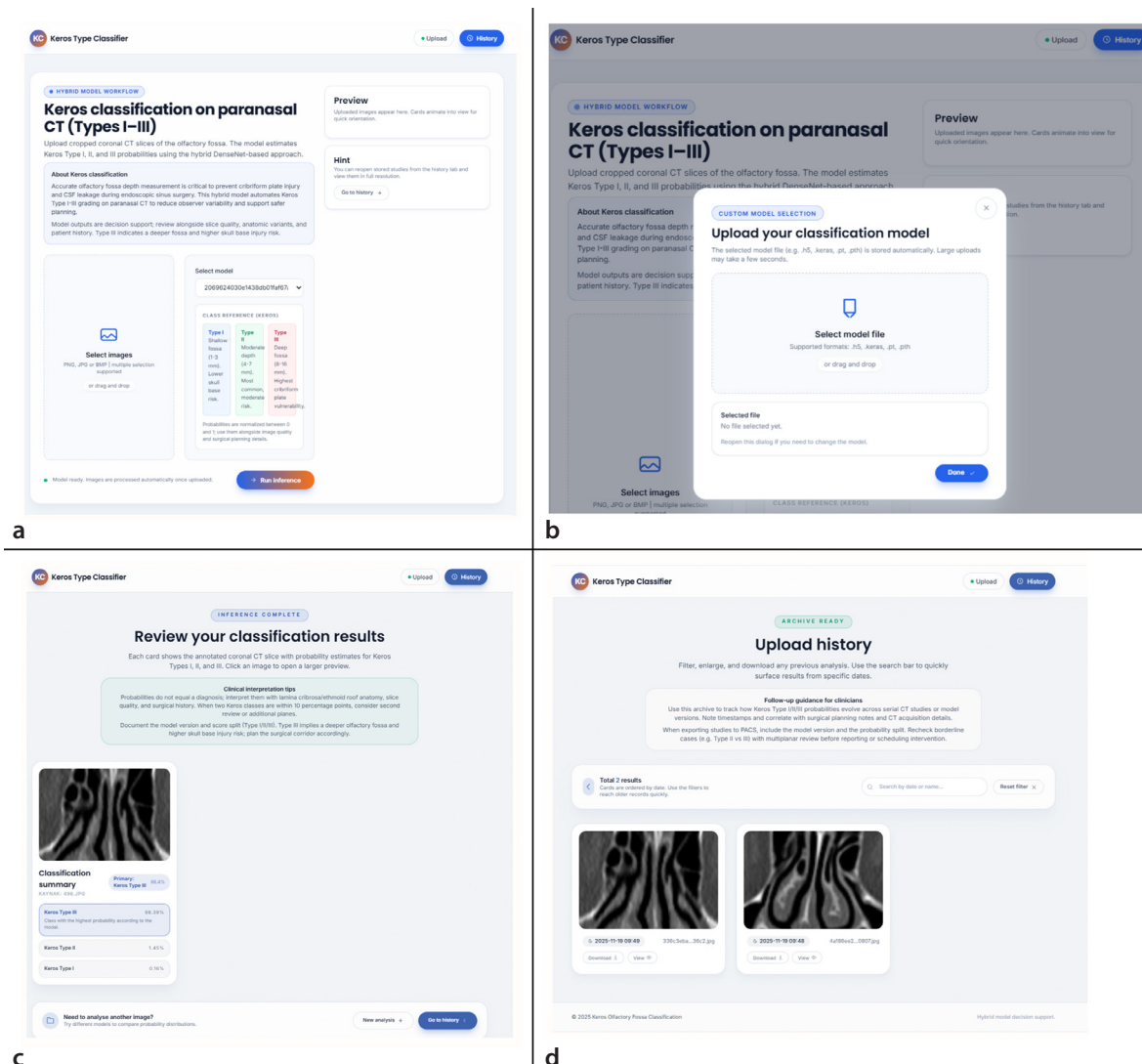


Figure 4. Clinical interface of the proposed decision support system. Overview of the clinical application interface developed for automated Keros classification. (a) Interface of application. The main interface of the application, allowing clinicians to upload coronal computed tomography slices for analysis. (b) Different models loading screen. The model selection screen, enabling users to choose among different pretrained hybrid models. (c) Show the results. The results display panel showing the predicted Keros type and corresponding probability values. (d) Show the history. The history section, where previously analyzed scans are stored for review and comparison.

izability of the proposed framework to data acquired from different institutions, scanner vendors, or acquisition settings. In addition, the dataset size and single-center design led to underrepresentation of some categories, particularly Keros type I. Future multicenter studies with larger and more heterogeneous datasets are therefore required to better assess model robustness under real-world clinical conditions. Third, nested cross-validation was not employed, which may limit the precision of generalization error estimation. This aspect should be addressed in future investigations with larger datasets that allow for more computationally intensive validation strategies. Furthermore, as a fixed DenseNet + CBAMs + SE architecture was used, the potential benefits of alternative attention mechanisms or transformer-based networks could not be explored. Future research may

investigate these architectures and incorporate interactive or explainable visualization tools to further improve interpretability and user confidence. Despite these limitations, the proposed framework demonstrates strong methodological rigor and clinical relevance and represents a meaningful contribution to the literature on AI-assisted radiological assessment of anatomically critical regions.

We presented a comprehensive modeling framework for Keros classification by integrating DL-based feature extraction, diverse feature selection methods, and classic ML algorithms. DenseNet, enhanced with CBAMs and SE blocks, was used to generate high-quality deep features, which were then optimized with selection techniques such as SHAP, RFE, PCA, and SelectKBest. These

features were classified using SVMs, random forest, logistic regression, Naive Bayes, and XGBoost.

Among the tested models, DenseNet169 combined with RFE and SVMs achieved the highest accuracy at 97.9%. This highlights the importance of both the architectural depth of the network and the choice of feature selection strategy. It is also significant that models relying solely on DL (approx. 88% accuracy) were markedly improved by the integration of feature selection and traditional classifiers. Other performance indicators, including precision, recall, and F1-score metrics, further confirmed the stability and reliability of the proposed method.

These findings suggest that combining deep feature extraction with effective fea-

ture selection and appropriate ML classifiers can enable the development of highly accurate clinical decision support systems for medical image analysis.

Although the results are promising, some areas remain for further improvement. First, the dataset was relatively small and drawn from a single center, limiting generalization. Larger, multicenter studies are needed to confirm robustness. Second, this work relied on a fixed DenseNet + CBAMs + SE design. Future research may explore other attention mechanisms, transformer-based architectures, or multi-scale feature extraction. Hybrid feature selection methods, such as SHAP combined with RFE or genetic algorithm-based selection, could also be investigated.

Another important direction concerns explainability. Although SHAP was employed, further efforts are needed to enhance interpretability and develop interfaces that clearly communicate the model's decision process to clinicians. Testing the system within real-time decision support platforms and assessing its usability in surgical planning will also be essential for successful clinical translation. Furthermore, the clinical interface developed in this study demonstrates that the proposed hybrid model can be effectively integrated into routine radiology workflows, enabling rapid, objective, and reproducible Keros classification to support surgeons in preoperative planning and risk assessment.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Snyderman CH, Pant H, Carrau RL, Prevedello D, Gardner P, Kassam AB. What are the limits of endoscopic sinus surgery?: the expanded endonasal approach to the skull base. *Keio J Med*. 2009;58(3):152-160. [\[Crossref\]](#)
2. O'Brien WT Sr, Hamelin S, Weitzel EK. The preoperative sinus CT: avoiding a "CLOSE" call with surgical complications. *Radiology*. 2016;281(1):10-21. [\[Crossref\]](#)
3. Özdemir A, Bayar Muluk N. The important adjacent structures for anterior ethmoidal artery in FESS: anterior ethmoidal artery canal angle, supraorbital ethmoid cells and Keros classification. *J Clin Neurosci*. 2022;98:207-212. [\[Crossref\]](#)
4. Ding J, Sun G, Lu Y, et al. Evaluation of anterior ethmoidal artery by 320-slice CT angiography with comparison to three-dimensional spin digital subtraction angiography: initial experiences. *Korean J Radiol*. 2012;13(6):667-673. [\[Crossref\]](#)
5. Souza SA, Souza MM, Gregório LC, Ajzen S. Anterior ethmoidal artery evaluation on coronal CT scans. *Braz J Otorhinolaryngol*. 2009;75(1):101-106. [\[Crossref\]](#)
6. Başak S, Karaman CZ, Akdilli A, Mutlu C, Odabaşı O, Erpek G. Evaluation of some important anatomical variations and dangerous areas of the paranasal sinuses by CT for safer endonasal surgery. *Rhinology*. 1998;36(4):162-167. [\[Crossref\]](#)
7. Pandolfo I, Vinci S, Salamone I, Granata F, Mazziotti S. Evaluation of the anterior ethmoidal artery by 3D dual volume rotational digital subtraction angiography and native multidetector CT with multiplanar reformations. Initial findings. *Eur Radiol*. 2007;17(6):1584-1590. [\[Crossref\]](#)
8. Abdullah B, Chew SC, Aziz ME, et al. A new radiological classification for the risk assessment of anterior skull base injury in endoscopic sinus surgery. *Sci Rep*. 2020;10:4600. [\[Crossref\]](#)
9. VAM, Santosh B. A study of clinical significance of the depth of olfactory fossa in patients undergoing endoscopic sinus surgery. *Indian J Otolaryngol Head Neck Surg*. 2017;69(4):514-522. [\[Crossref\]](#)
10. Yağmur AR, Çıvgın E, Özcan KM, Yurtsever Kum N, Karakuş MF, Dere HH. Analysis of the correlation of the lamina papyracea-to-midline distance with the location of anterior ethmoidal artery and Keros classification. *Indian J Otolaryngol Head Neck Surg*. 2023;75(4):3146-3151. [\[Crossref\]](#)
11. Şahan MH, Inal M, Muluk NB, Şimşek G. Cribriform plate, crista galli, olfactory fossa and septal deviation. *Curr Med Imaging Rev*. 2019;15(3):319-325. [\[Crossref\]](#)
12. Paudel S, Sah R, Budhathoki T, Pandey G. Evaluation of olfactory fossa depth using computed tomography. *J Nepal Health Res Counc*. 2025;22(4):707-711. [\[Crossref\]](#)
13. Babu AC, Nair MRPB, Kuriakose AM. Olfactory fossa depth: CT analysis of 1200 patients. *Indian J Radiol Imaging*. 2018;28(4):395-400. [\[Crossref\]](#)
14. Alsalama A, Harous S, Elnagar A. Paranasal sinus analysis based on deep learning and machine learning techniques: a comprehensive survey. *Intelligent Systems with Applications*. 2025;27:200559. [\[Crossref\]](#)
15. Wang Y, Zhang X, Du W, Dai N. Deep learning-based fully automatic segmentation of the paranasal sinuses in chronic rhinosinusitis patients using computed tomographic images. *IEEE Access*. 2025;13:16444-16454. [\[Crossref\]](#)
16. Lee DJ, Hamghalam M, Wang L, et al. The use of a convolutional neural network to automate radiologic scoring of computed tomography of paranasal sinuses. *Biomed Eng Online*. 2025;24(1):49. [\[Crossref\]](#)
17. Gogus E, Yılmaz A, Enercan M. A novel deep hybrid model for automatic femoral stem classification in hip arthroplasty from radiographs: MSFT-Net with CBAM and transformer modules. *IEEE Access*. 2025;13:102564-102577. [\[Crossref\]](#)