

Support vector machine-based diagnostic classification of caries and healthy teeth in panoramic orthopantomography images

Süleyman Dal¹ , **Yasin Sönmez²** ¹ Batman University Rectorate, Energy Coordination Office, Batman, Türkiye² Batman University, Vocational School of Technical Sciences, Department of Computer Technologies, Batman, Türkiye**Received:** 12 November 2025**Accepted:** 24 December 2025**Correspondence:****Dr. Süleyman DAL**Batman University Rectorate, Energy
Coordination Office, Batman, Türkiye.E-mail: suleyman.dal@batman.edu.tr**How to cite this article:**

Dal S, Sönmez Y. Support vector machine-based diagnostic classification of caries and healthy teeth in panoramic orthopantomography images. Int Dent Res. 2025;15(3):136-147.

<https://doi.org/10.5577/intdentres.696>

Abstract

Aim: This study proposes a support vector machine (SVM)-based classification model for detecting dental caries in panoramic dental orthopantomography (OPG) images. The primary objective is to develop a computationally efficient, reproducible, and clinically adaptable method that enables the automatic differentiation between carious and healthy tooth structures.

Methodology: The images used in the study were first subjected to standard resizing and normalization steps. The extracted features were then prepared for classification through feature scaling, feature selection, and dimensionality reduction processes. Hyperparameter grid search, stratified k-fold cross-validation, and decision threshold optimization were applied during the model development process. Medically meaningful metrics, including accuracy, precision, sensitivity, F1 score, area under the receiver operating characteristic (ROC) curve (AUC), and the confusion matrix, were used in the performance evaluation.

Results: The SVM model demonstrated balanced and stable performance throughout cross-validation. ROC and Precision-Recall analyses showed that the model has a high capacity to distinguish between decayed and healthy classes. Decision threshold scanning revealed that false-positive and false-negative rates can be optimized according to clinical requirements.

Conclusion: The results show that the proposed SVM-based approach offers a reliable and feasible method for automatic caries detection from panoramic OPG images. This study highlights that traditional machine learning methods, when combined with structural feature engineering, are suitable for supporting dental diagnosis processes. Future studies should test the model on larger datasets and perform multi-classification of different dental pathologies.

Keywords: Support vector machine, machine learning, deep learning, hyperparameter, dentistry, panoramic dental radiography, caries

Introduction

Dental caries is recognized as one of the most common chronic diseases worldwide, affecting both children and adults (1). Global epidemiological studies show that untreated decay in permanent teeth affects billions of individuals and negatively impacts oral health, general health, quality of life, and the economic burden on healthcare systems (1, 2). Untreated caries can lead to pain, infection, tooth loss, impaired chewing function, and aesthetic problems; these effects are particularly pronounced in socioeconomically disadvantaged groups (2, 3). In this context, the early detection of dental caries with the highest possible sensitivity and specificity has become a fundamental requirement for both preventive dentistry and treatment planning. Traditional caries diagnosis is based on the combined interpretation of clinical examination and radiographic evaluation (4). However, this process is highly sensitive to human factors, such as the dentist's experience, perspective, and fatigue. Both clinical examination and conventional radiography have limited sensitivity, particularly in early-stage lesions; significant demineralization must occur in the enamel tissue before a lesion becomes visible on radiographs (4, 5). In addition, radiographic interpretations for caries detection can have high inter-observer and intra-observer variability, meaning that the same image may be assessed differently by different dentists (6). These limitations clearly highlight the need for more objective and reproducible methods that go beyond human vision in caries diagnosis.

In recent years, the growing interest in artificial intelligence and machine learning applications in healthcare has also been reflected in dental diagnosis processes; in particular, the automatic detection of dental caries has become a key topic on the research agenda (6, 7). Machine learning can process numerous clinical, behavioral, or image-based variables simultaneously to generate predictions regarding the presence of decay or the risk of decay, thereby providing decision support to clinicians (8-11). In contrast, deep learning-based image processing approaches can analyze complex structures at a level undetectable by the human eye, owing to their ability to learn high-dimensional features from raw radiographic images (12). However, in datasets with a limited number of labelled images, training deep networks directly from end to end can lead to overfitting and generalization problems. Therefore, hybrid approaches that extract features from pre-trained networks and combine these features with 'lighter' machine learning classifiers are gaining importance (13). In the dental imaging literature, a significant portion of studies conducted for caries detection focus on localized images, such as periapical or bitewing radiographs (14). However, panoramic orthopantomography (OPG) images enable a comprehensive assessment of the patient's intraoral condition, as they present all jaw bones and teeth in a single image (14,15). In this respect, OPG is considered an indispensable tool, particularly for initial consultation and general screening evaluations (15, 16).

However, panoramic images present a more challenging data structure for automatic analysis due to factors such as lower resolution compared to intraoral films, overlapping anatomical structures, geometric distortion, and noise (17). In this complex image environment, where teeth and surrounding structures are intermingled, the detection of small and early-stage caries lesions is difficult both to the human eye and with classical image processing methods (17, 18). This situation requires artificial intelligence models that work on panoramic radiographs to be specifically designed for both feature extraction and classification stages.

Despite this, the literature shows that the number of studies using open-access datasets and reporting results in a reproducible manner for caries detection from panoramic OPG images is quite limited (19). Most studies are conducted on datasets obtained from closed clinical archives that cannot be shared with external researchers, which limits the comparability and independent verifiability of these methods. In this context, the Dental OPG X-ray Dataset offered on the Kaggle platform stands out as an important data source for researchers, consisting of panoramic dental images and containing classes such as 'caries' and 'healthy' (20). The labelled caries and healthy images in the 'Dental OPG (Classification)' subfolder of this dataset are suitable for providing a common reference point for both caries detection-focused studies and new algorithms working on panoramic images. Working on an open and standard dataset is of particular scientific value, as it enables the methods developed to be replicated by independent researchers and different models to be compared under equal conditions (14, 20).

From a machine learning perspective, the complex structure of panoramic OPG images and the relatively limited number of examples make hybrid approaches more appealing than classical single-stage deep network architectures. Feature extraction from deep architectures previously trained on large datasets and processing these features with relatively smaller parameterized classifiers, such as support vector machines, has been proposed as a robust solution, particularly for problems with limited data (21, 22). Support vector machines (SVM) stand out as a method capable of achieving good generalization in scenarios with few classes but many features, owing to their ability to create decision boundaries that maximize the margin in high-dimensional feature spaces and their regularization mechanisms (22, 23). In noisy and complex data structures, such as radiographic dental images, combining SVM with appropriate preprocessing and feature selection steps can increase both the stability and performance of the model (23).

This study was designed based on the requirements and gaps mentioned above. The primary objective of the study is to develop a classification model that can distinguish between decayed (caries) and healthy teeth classes with high performance, using images obtained from panoramic radiographs located in the 'Dental OPG (Classification)' subfolder within the Kaggle 'Dental OPG X-ray Dataset'. To this end, region-of-interest (ROI)

cropping focused on the tooth region in panoramic OPG images and limited data augmentation steps consistent with biological reality were first applied. Deep feature vectors were then extracted from the images using two pretrained deep networks (EfficientNet-B0 and ResNet-50), and these vectors were combined to create a high-dimensional but rich feature space. In the final stage, dimension reduction was performed using mutual information-based feature selection and principal component analysis, and the remainder was fed into an SVM classifier with a radial basis function kernel. The original scientific contributions of this study to the literature can be summarized as follows:

- Demonstrating that high performance can be achieved even under limited data conditions by using a classification framework based on an SVM with deep feature extraction.
- Establishing a foundation for a reliable and explainable artificial intelligence-based caries screening infrastructure that can be integrated into clinical decision support systems in the future, based on the findings obtained.

This study aimed to automatically classify dental caries and healthy examples from panoramic dental orthopantomography (OPG) images.

Materials and Methods

Dataset

The images used were obtained from the ‘Dental OPG (Classification)’ subsection of the open-access Dental OPG X-ray Dataset database. As shown in Fig. 1, the dataset is organized into two classes and contains a total of 342 images; 119 of these belong to the caries category and 223 to the healthy category.

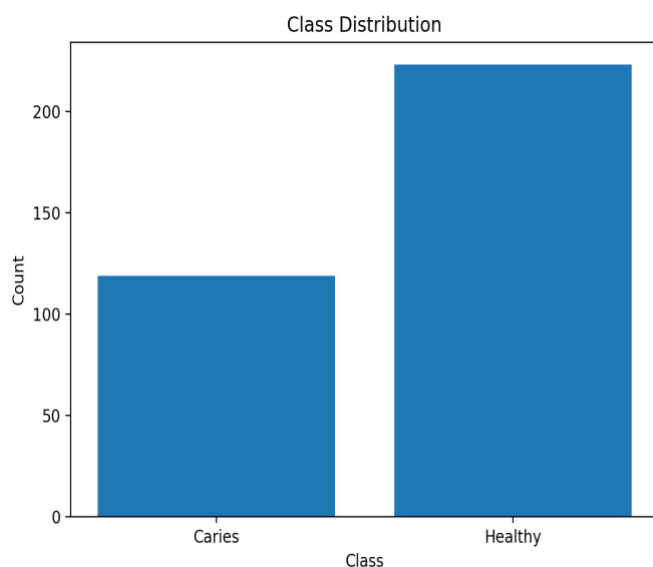


Figure 1. Class distribution

The dataset was organized into two classes and comprised 342 images: 119 in the caries category and 223 in the healthy category. This asymmetry in class distribution reflects the imbalance in the actual clinical population, ensuring that the model is evaluated with higher representativeness for practical scenarios. Fig. 2 shows sample images from both classes and provides a visual reference for the dataset's overall appearance.

General methodology framework

The method proposed in this study is based on the classification of deep feature representations obtained from panoramic dental images using support vector machines (SVM). The model development process has been refined within a three-dimensional methodological framework. In the first stage, a basic performance evaluation was conducted using SVM on high-dimensional features extracted through deep networks, revealing the classification potential of the model. In the second stage, different types of deep features were combined to create a richer and more discriminative representation space, thereby enabling the SVM to effectively learn non-linear decision boundaries. In the third stage, the SVM pipeline was optimized using a systematic hyperparameter search, feature selection, and dimensionality reduction strategies to control the model complexity and improve the generalization performance. By comprehensively addressing these three stages, this study robustly and systematically establishes the methodological foundations of the final SVM-based classification model.

Image pre-processing

The panoramic OPG images used within the scope of the applied method were subjected to a systematic pre-processing procedure to ensure that the classification process was performed reliably and reproducibly. Only files with the .jpg extension in the dataset were evaluated within the scope of the study, thus preserving data integrity. All images were converted to a standard resolution and rescaled to 224×224 pixels to ensure consistency during the classification stage. This process eliminated the potential effects of size differences arising from different devices, imaging parameters, and clinical conditions on the learning process. In the next step, pixel intensities were normalized by reducing them to a specific range, preventing variability in brightness, contrast, radiographic exposure, and image quality from negatively affecting the model performance. This pre-processing strategy standardized the numerical representation of all examples in the dataset, contributing to more stable results in the feature extraction and classification steps. This enhanced the reliability and clinical evaluability of the performance metrics achieved by the classifier algorithm.

Sample Images per Class

Caries



Healthy



Figure 2. Sample OPG images from caries and healthy classes

Data augmentation

Data augmentation methods were applied to mitigate the risk of overfitting due to the limited number of images used during data processing and the high-dimensional structure of the representation space. Data augmentation procedures were designed in a controlled manner to enhance the model generalization ability. First, data augmentation was not applied during the cross-validation phase, thereby preventing information leakage between different layers and preserving the statistical integrity of the performance measurements. In the final training process, the diversity of the dataset was increased by deriving a limited number of additional examples from each of the original image. The transformations applied were limited to operations that preserve the anatomical structure of the image and do not alter the label information, such as brightness and contrast changes, the addition of low-level Gaussian noise, and limited-scale zooming. This approach resulted in a training set that was more resilient to variations arising from radiographic imaging conditions, thereby contributing to a more robust and generalizable performance of the classification model across different image quality levels.

Deep learning-based feature extraction

In the feature extraction process, pre-trained deep learning models commonly used for medical images were utilized. The final classification layers of these models were disabled, and global average pooling was applied to ensure that each image was represented by a fixed-size feature vector. Thus, spatial information where the positional relationships in panoramic OPG images are concentrated was gathered under a single descriptor structure. The features obtained in the first stage were tested directly on classical machine learning classifiers, while in the subsequent stage, a more comprehensive representation space was created by combining different model outputs. This approach enabled the capture of both the fine textural patterns between dental tissues and the general structural components of the image, providing a richer data foundation to enhance the accuracy of the final classification process.

Label coding and class weights

The classes in the dataset were initially defined with textual labels, which were converted into numerical values to make them compatible with the classification algorithms. Accordingly, healthy tooth images were coded as '0' and decayed tooth images as '1'. Owing to

the imbalance in class distribution, that is, the number of healthy examples being significantly higher than the number of decayed examples, class weights were used during the training process. The class weights were determined inversely proportional to the class ratios in the dataset, thereby preventing the model from overfitting to the majority class. Such an approach contributes to increasing the recall rates of the decay class, in particular, and enables the classifier to improve its ability to detect the minority class and evaluate the overall performance metrics in a more balanced manner.

Feature scaling and dimension reduction

The high dimensionality and varying scales of the vectors obtained after deep feature extraction can lead to various numerical imbalances in the classification process. In particular, in algorithms with distance-based decision mechanisms, large-scale features can disproportionately influence the model's decision; therefore, the aim is to ensure that all features have a similar distribution structure. To this end, each feature dimension was scaled to have a mean of zero and a variance of one, thereby balancing the dominance of the features on margin optimization. In the subsequent stage, based on the assumption that the entire high-dimensional feature space does not carry equal levels of information for classification, a feature selection process was applied to highlight the most informative components. This process helped eliminate components that had a low direct contribution to classification, were noisy, or were redundant. Following feature selection, dimensionality reduction was performed on the remaining dimensions, and the data space was transformed into a new set of components explaining the maximum variance. This made the classifier learning process more stable and reduced the computational cost. This sequential process, consisting of scaling, feature selection, and dimensionality reduction, improved the overall performance of the model, reduced the risk of overfitting, and ensured that classification decisions were made in a more balanced manner.

Support Vector Machines (SVM) and classification framework

Support Vector Machines are a powerful classification method based on supervised learning that aims to determine the separating decision boundary between two classes with the widest margin. The fundamental approach of the model is to find the optimal hyperplane that separates the classes in the feature space and allows the data points closest to this plane (support vectors) to determine the structure (25, 26). In this context, the goal is not only to achieve correct classification but also to establish a decision boundary with the safest possible margin. This feature provides a significant advantage, particularly for datasets consisting of high-dimensional feature vectors. The theoretical foundation of SVM is highly suitable for problems with relatively few examples

but high feature dimensions, and vectors obtained through deep learning-based representation extraction on dental panoramic images also possess a similar quality. Therefore, the preference for SVM in the classification stage is based on both the stability provided by the margin-based decision mechanism and its high-dimensional generalizability (27).

The performance of the SVM classifier used in this study was optimized by evaluating different kernel functions and tuning parameters. Within the scope of kernel selection, linear and radial basis functions were tested, systematically observing how the decision surface behaved in both linear and nonlinear scenarios. The regularization parameter was tested with different values to control the balance between the margin width and the model error tolerance. In addition, various parameter combinations were tested to evaluate the effect of feature selection and dimension reduction steps on the SVM decision mechanism. Ten-fold balanced cross-validation was performed for different scenarios, and the best structure was determined as the final model based on the average F1 score, which represents the success of the decay class. This approach ensured that the classification performance was optimized through a multidimensional evaluation rather than being dependent on a single parameter, thereby increasing the reproducibility and statistical reliability of the results.

Cross-validation strategy

To reliably assess model generalizability, a balanced ten-fold cross-validation method was applied. Under this strategy, the dataset was divided into ten equal sub-sections, but the ratio of defective to healthy classes in each section was kept consistent with the overall distribution. This approach reduced the biases that class imbalance could create in model performance measurements and allowed for an average performance estimate across different sampling scenarios without relying on a single training-test split. Furthermore, this method enabled the assessment of the statistical stability of the classifier performance by providing the ability calculating the variance of the results obtained from different splits. Reapplying the scaling, feature selection, and dimensionality reduction steps for each fold prevented information leakage between the training and validation processes and ensured the methodological integrity.

Hyperparameter optimization

Hyperparameter optimization, a key component of the classification process, was performed using a systematic search approach. To this end, a parameter space encompassing different kernel functions, regularization parameters, and feature dimensions was evaluated. Ten-fold cross-validation was applied for each parameter combination, and the optimal structure was determined based on the average F1 scores, which better reflected the representativeness of the defective class. Thus,

without using an overly extensive search space on the limited-sized dataset, the critical components affecting the model's decision boundary were optimized with an appropriate balance. This process contributed to obtaining classification performance in a more stable and iterative manner by reducing parametric uncertainties.

Decision threshold optimization

Most classifiers use a fixed threshold value to determine class labels from probability outputs. However, in medical image analysis, adjusting the decision threshold is crucial due to the potential clinical consequences of missing positive cases. In this study, potential threshold values were evaluated across a wide range, and the performance profile was examined by calculating the F1 score for each of the thresholds. The threshold value corresponding to the highest F1 score was defined as the optimal decision level. While basic performance reporting was conducted using the traditional 0.5 threshold, findings related to different threshold levels were supported by additional visuals, and a discussion was held on which threshold values provide a more balanced error distribution depending on the clinical situations.

Performance metrics

The performance metrics used to evaluate the classification model have been selected with particular consideration given to their clinical relevance and interpretability, especially in the context of medical image analysis. Although accuracy can provide the proportion of correctly classified examples within the entire observation set, it can be misleading when the class distribution is imbalanced. Therefore, in this study, the accuracy value has been evaluated only as a complementary indicator. Balanced accuracy is based on the average of the sensitivity values calculated for each class and provides a more balanced evaluation when the number of healthy images in a dataset is high.

Accuracy measures the effect of false positive predictions by showing how many of the samples predicted to be positive are actually positive (28). It is particularly important in medical decision-making processes in terms of preventing unnecessary further investigations or treatments (29). In contrast, sensitivity indicates how many of the true positive samples are correctly detected and provides a critical criterion for ensuring that cases of decay are not overlooked (29, 30). Since these two metrics are insufficient on their own in medical classification problems, the F1 score, which represents the harmonic mean of precision and sensitivity, has been used. The F1 score was preferred as the primary performance indicator for model selection and decision threshold optimization because it provides a more meaningful summary evaluation for imbalanced datasets. In addition to these metrics, a confusion matrix was also used in the study. The confusion matrix allows for a detailed examination of the model error types by

displaying the number of correct and incorrect classifications for each class at the row and column levels (31, 32). In particular, the distribution of false negatives (samples predicted to be healthy but actually decayed) and false positives (samples predicted to be decayed but actually healthy) contributes to a more qualitative interpretation of the performance of this matrix model, as it yields clinically distinct outcomes in the context of dental imaging. The findings obtained from the confusion matrix provide an opportunity to discuss, for example, whether errors are systematically concentrated in specific anatomical regions or whether the model has the capacity to distinguish early-stage caries images.

The model's discriminatory capacity was also evaluated using the ROC curve and the area under the curve (ROC-AUC). The ROC curve graphically illustrates the relationship between the true positive rate and the false positive rate at various decision thresholds; the AUC value summarises the discrimination power between classes in a single value. AUC values approaching 1 indicate that the model has a high ability to distinguish between images of decayed and healthy teeth. In this study, all metrics were calculated separately for each fold within a ten-fold cross-validation protocol, and variance values were presented alongside the average performance. Thus, not only the model performance but also the consistency and statistical stability of the predictions were systematically evaluated [31-33]. This comprehensive performance evaluation approach reveals not only the model's overall accuracy level but also different error profiles and clinically critical misclassification types, thus providing a comprehensive analysis of the model's applicability, reliability, and areas for improvement.

Results

This study proposes an integrated classification framework that combines traditional machine learning algorithms with a deep feature extraction-based approach to distinguish between decayed and healthy teeth using panoramic dental images. The developed methodology systematically incorporates multi-layered processes, such as scaling, feature selection, dimensionality reduction, and decision threshold optimization, leveraging the high discrimination power of support vector machines (SVM) in the final classification stage. The decision mechanism of SVM, based on the principle of margin maximization in high-dimensional feature spaces, offers a significant advantage in producing stable decision boundaries, even under limited and imbalanced data conditions. The cross-validation and hyperparameter optimization steps applied throughout the study enabled a reliable evaluation of the developed model's performance. This approach points to a potential decision support structure that can be integrated into computer-aided clinical evaluation systems in the future, offering an explainable

and reproducible artificial intelligence-based decay detection infrastructure.

Table 1 shows the optimal hyperparameter configuration determined for the SVM classifier. The basic parameters include feature selection with SelectKBest for 300 features, variance preservation with PCA at 90%, a regularization coefficient of $C=10$, and an RBF kernel function. Furthermore, the model performance was improved through decision threshold optimization, achieving consistent classification behavior across different layers. The average best decision threshold was calculated as 0.613 ± 0.093 , indicating the model's tendency to produce balanced decisions on ambiguous examples.

Tables 2 and 3 present a summary of the 10-fold cross-validation results and layer-based performance metrics. The average accuracy value was 87%, the F1 score was 0.82, and the ROC-AUC was 0.92, with precision and recall metrics hovering around 0.79 and 0.85, respectively. The variation in accuracy values between 70% and 97% in the layer-based results indicates performance differences depending on the data subsets. However, the fact that the ROC-AUC values are above 0.90 in most layers indicates that the discrimination power between classes can be maintained independently of the data divisions. Overall, the table reveals that despite individual variations, the model exhibits a statistically consistent and generalizable performance.

Table 1. Optimal SVM hyperparameters

Hyperparameter	Value
Feature selection (SelectKBest) - k	300
Variance ratio preserved by PCA	0.90
Regularisation coefficient (C)	10
Kernel function	RBF
Decision threshold optimization	Applied
Mean_BestThresh±Std_BestThresh	0.613 ± 0.093

Table 2. Performance metrics (10-fold cross-validation)

Metric	mean ± std
Accuracy	0.8709 ± 0.079
Precision (Caries)	0.7949 ± 0.1088
Recall (Caries)	0.8576 ± 0.1039
F1 (Caries)	0.824 ± 0.102
ROC-AUC (Caries)	0.926 ± 0.0624

Table 3. 10- Detailed metrics

Fold	Best	Accuracy	Precision	Recall	F1	ROC-AUC
1	0.57	0.9714	0.9231	1.0	0.96	1.0
2	0.71	0.9429	0.8571	1.0	0.9231	1.0
3	0.67	0.9118	0.8462	0.9167	0.88	0.9678
4	0.63	0.8824	0.8333	0.8333	0.8333	0.9697
5	0.61	0.8529	0.7692	0.8333	0.8	0.9015
6	0.51	0.8235	0.7143	0.8333	0.7692	0.8712
7	0.75	0.7059	0.5714	0.6667	0.6154	0.8277
8	0.67	0.7941	0.6923	0.75	0.72	0.8542
9	0.44	0.9118	0.9091	0.8333	0.8696	0.9034
10	0.57	0.9118	0.8333	0.9091	0.8696	0.9644

Figure 3 shows the performance of the developed classification model in distinguishing positive and negative classes under different decision threshold values. The high area under the curve (AUC = 0.933) indicates that the model maintains its ability to detect true positives even when the false-positive rate increases and that the discrimination power between classes is significant. This metric serves as a fundamental reference for evaluating the overall accuracy of the classification performance.

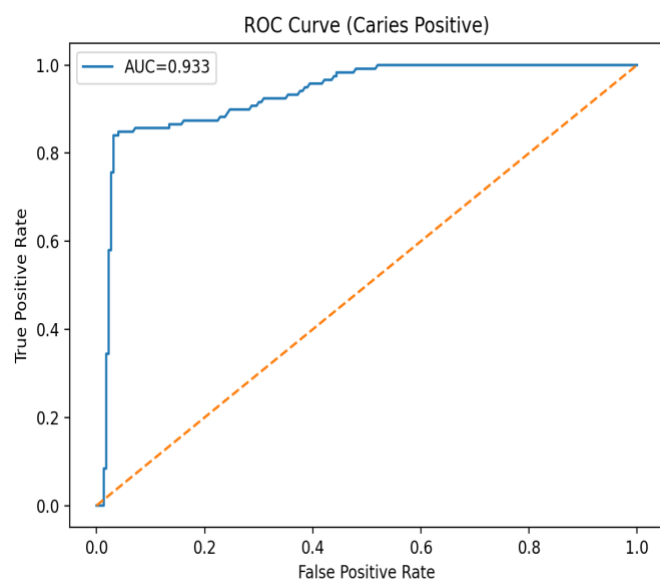


Figure 3. ROC curve

Figure 4 shows the relationship between precision and recall for the decay class across different decision threshold values. The high area under the curve (PR-AUC = 0.828) indicates that the model maintains a low probability of producing false positives, particularly when detecting examples related to the positive class. The precision-recall curve provides a complementary perspective to the ROC curve for evaluating the model's positive class performance in imbalanced classification problems.

Fig. 5 comprehensively presents the performance characteristics of the developed classification model during the 10-fold cross-validation process. The upper panel presents the fold-based distribution of the accuracy and F1 scores. The variability observed across layers revealed the structural heterogeneity of the data subsets and the direct sensitivity of the classification problem to the data distribution. The lower panel shows that ROC-AUC values mostly remain at high levels across folds, indicating that the discrimination power between classes is sustainable, independent of data splits. The decision threshold-F1 relationship graph in the lower right panel reveals the sensitivity of the classification performance to threshold selection, indicating that the optimal F1 score is achieved at threshold values of approximately 0.60. This visual whole provides an

interpretable analytical framework for evaluating the statistical stability of the model, the dependence of the decision mechanism on the threshold parameters, and the effects of the cross-validation process on the classification performance.

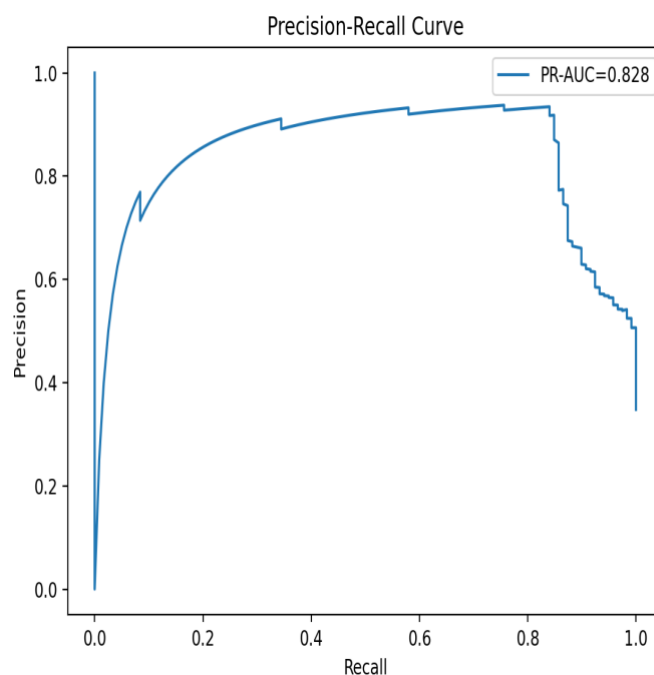


Figure 4. Precision-recall curve

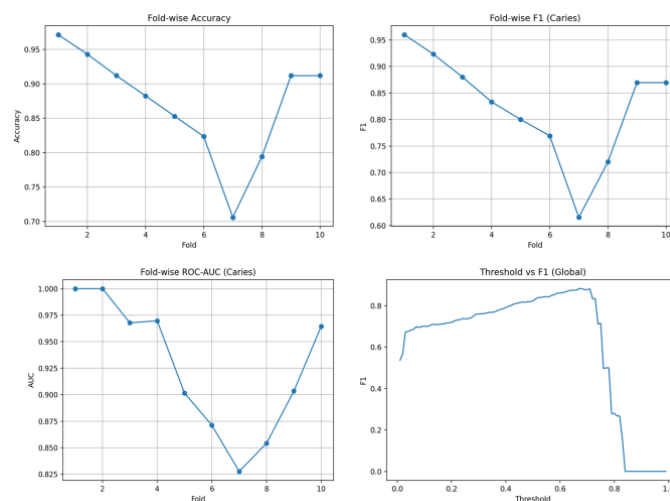


Figure 5. Cross-validation performance and threshold optimization

Fig. 6 illustrates the prediction behavior of the classification model on the samples, both in absolute values and normalized ratios. The confusion matrix presented in the left panel shows the numerical distribution of the correct and incorrect classifications. The normalized matrix in the right panel provides a

proportional assessment of the classification performance by the class. The limited level of false positive rates for the healthy classes and the high true positive rates for decayed classes indicate that the model can consistently distinguish between the two classes. This visualization provides a fundamental reference for evaluating the structure of classification errors and interpreting the model performance at the class level.

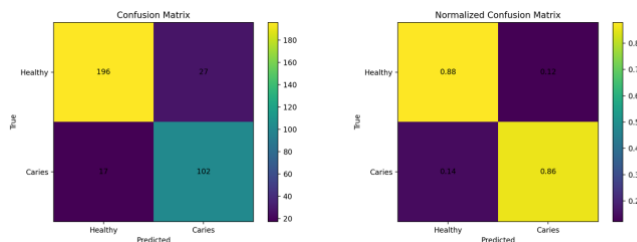


Figure 6. Confusion matrix and normalized confusion matrix

Figure 7 shows the distribution of the model's predicted probabilities for the caries class across classes. The clustering of samples in the caries class at high probability values indicates that the model has a strong tendency to distinguish the positive class; the concentration of samples belonging to the healthy class at low probability values indicates high distinguishability in the negative class. The clear separation of the distribution at both ends allows flexibility in threshold selection and provides a reliable basis for classification in clinical applications.

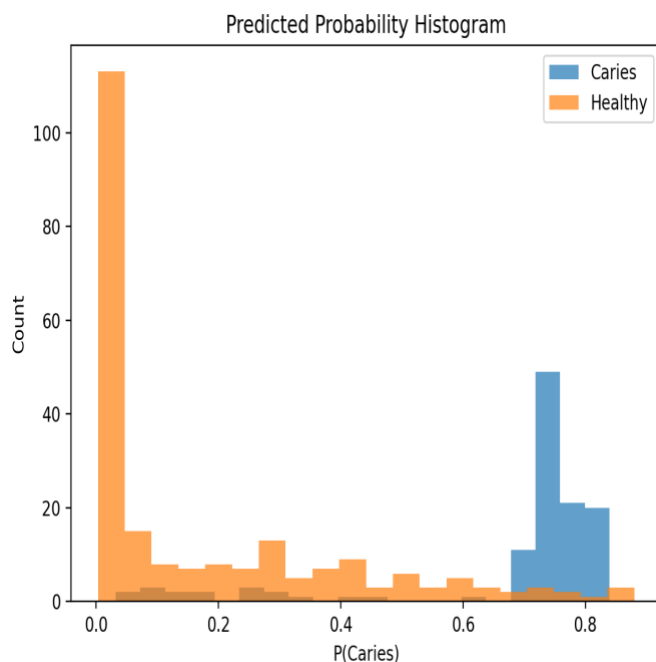


Figure 7. Predicted probability histogram

Figure 8 presents a comparative analysis of the distribution of the model probability estimates for the 'decayed' class across the two classes in the form of a box and violin plot. The concentration of the probability distribution in the lower region and the proximity of the median to zero in healthy samples indicate that the model represents the negative class with low estimate values. In contrast, in rotten samples, the median is located at high values, and the distribution is concentrated in the upper region, revealing that the positive class is distinctly separated. When the box and violin plots are evaluated together, the overlap between class distributions is limited; therefore, the model exhibits a consistent probability structure in terms of determining the decision threshold and analyzing misclassifications.

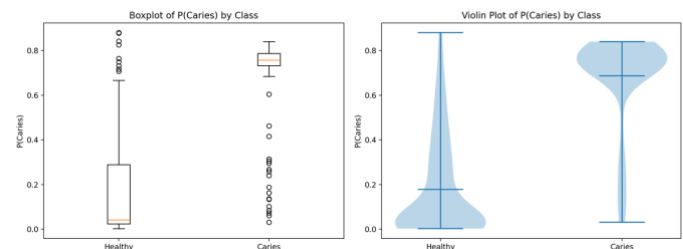


Figure 8. Distribution of Predicted Probabilities Across Classes

Figure 9 presents a two-dimensional representation of the feature space obtained by the classification model using Principal Component Analysis (PCA) and t-distributed stochastic neighborhood embedding (t-SNE) methods. In the PCA graph (for the 'Healthy' and 'Caries' classes), it can be seen that the samples generally exhibit an overlapping distribution, although there are clustering tendencies between classes in certain regions. This indicates that the linear components alone cannot fully explain the class separation. In contrast, the t-SNE representation better reflects the non-linear relationships between the samples, revealing class density differences, particularly in high-probability regions. When both visualizations are evaluated together, it is concluded that structural patterns enabling differentiation between classes exist in the data manifold; however, the distinction is not entirely clear, thus justifying the use of margin-based methods such as SVM.

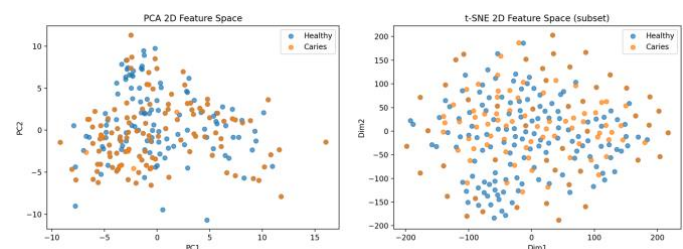


Figure 9. Two-dimensional feature space visualization (PCA and t-SNE)

Discussion

This study evaluated the performance of an SVM-based classification model developed to automate the detection of dental caries from panoramic orthopantomography (OPG) images within a multidimensional evaluation framework. The findings indicate that the combined application of appropriate feature selection, dimensionality reduction, and decision threshold optimization significantly improves classification success. In particular, the hyperparameter search results revealed that parameters such as the regularization coefficient (C) and RBF kernel are effective in dealing with class imbalance and image variation in the dataset. In this context, the high average ROC-AUC and F1 scores achieved by the model demonstrate not only its strong discrimination ability but also that false positive and false negative rates can be reduced in a balanced manner. The combined evaluation of the tables and figures confirms that the variation between the models remained limited throughout the 10-fold cross-validation and that the results are not based on randomness. In particular, the ROC and Precision-recall curves demonstrated that the sensitivity-specificity balance could be significantly optimized by visualizing the model's prediction behavior at different decision thresholds. However, probability distribution analyses, such as prediction probability histograms and box and violin plots, revealed that the model could produce well-separated probability regions for both classes. This indicates that, in classification studies where decision threshold optimization is applied, multiple decision functions should be considered rather than a single metric. Furthermore, it was found that the majority of misclassifications observed in the confusion matrix results originated from examples close to the boundary. These finding highlights that the model's decision margin is the most critical determinant in clinical image analyses where intra-class variation is pronounced.

Dimension reduction visualizations (PCA and t-SNE) were evaluated together to reveal linear and non-linear separation structures in the data manifold. PCA results show that linear components cannot provide a complete separation between classes; in contrast, t-SNE produces distinct clusters, particularly in marginal regions. This supports the notion that the margin-based separation of SVM in high-dimensional space is an appropriate choice for complex, noisy, and anatomically variable datasets, such as dental images. Although deep learning-based methods are frequently recommended in the literature for similar problems, it is noteworthy that the standard machine learning pipeline used in this study (feature selection + PCA + SVM) yielded satisfactory results with relatively fewer data requirements.

When the results are considered holistically, the developed model demonstrates consistent performance in terms of generalizability, decision threshold optimization, and ability to produce precise probability estimates. However, it has been observed that

misclassifications, particularly in the caries class, are concentrated in certain sub-image groups. This situation may necessitate integrating additional data augmentation techniques, local lesion region segmentation, or attention-based weighting approaches into the model in the future. Furthermore, addressing class imbalance with more advanced balancing methods (e.g., SMOTE variants and dynamic updating of class weights) has the potential to increase both sensitivity and specificity. Overall, the findings demonstrate that SVM-based decision models can be an effective tool for caries detection in panoramic dental images, with clinical applicability possible when evaluated using multiple performance metrics. This study demonstrates that classical machine learning methods, when combined with appropriate systematic pre-processing steps, enable the development of classification systems that are more cost-effective and computationally faster than deep learning networks.

Conclusion

This study has developed an SVM-based classification approach for detecting dental caries from panoramic dental orthopantomography (OPG) images. The systematic combination of feature selection, dimensionality reduction, and decision threshold optimization ensured that the model produced balanced and stable results across different performance metrics. The findings obtained through cross-validation revealed that the model effectively balanced sensitivity and specificity, achieving high values for metrics such as ROC-AUC and F1. Furthermore, visualizations based on the distribution of prediction probabilities demonstrated that the model successfully separated areas of uncertainty for distinct classes. When the results are evaluated holistically, it can be stated that the proposed system offers a reliable, computationally efficient, and practically adaptable structure for caries detection from panoramic dental images.

This study has achieved high classification accuracy but has certain limitations. First, although the dataset used represents real clinical records, the imbalance in class distribution and variability in image quality may lead to decision errors in certain subcases. The anatomical diversity of carious lesions and artifacts in the images increase the model's margin of error in marginal cases. Additionally, the model only produces decisions at the global image level and does not utilize segmentation information regarding the location of the lesion. This may limit the clinician feedback in terms of explainability. Finally, the model has only been evaluated with a single type of radiographic image (panoramic OPG) and has not been tested with different modalities, such as periapical, bite-wing, or CBCT; therefore, its generalizability requires validation with different data sources.

Future studies should focus on improving the current approach in terms of both performance and clinical applicability. First, addressing class imbalance using advanced sampling techniques (e.g., SMOTE variants, adaptive synthetic sampling) and applying cost-sensitive learning strategies in decision threshold optimization would be beneficial. Automatic segmentation of the lesion region or region-based deep feature extraction could increase the lesion-focused interpretability of the model. Furthermore, the inclusion of different radiographic modalities would enable comparative performance analyses using multiview learning strategies. Finally, integrating model outputs into clinical decision support systems, applying semi-supervised learning techniques with expert feedback, and performing performance validation through prospective real-time experiments can strengthen the clinical validity of the results obtained.

Disclosures

Peer-review: Externally peer-reviewed.

Author Contributions: Conception - S.D.; Design - S.D., Y.S.; Supervision - Y.S.; Materials - Y.S., S.D.; Data Collection and/or Processing - Y.S.; Analysis and/or Interpretation - S.D., Y.S.; Literature Review - Y.S.; Writer - S.D.; Critical Review - S.D., Y.S.

Conflict of Interest: No conflict of interest was declared by the authors.

Funding: The authors declared that this study has received no financial support.

References

- Kassebaum NJ, Bernabé E, Dahiya M, Bhandari B, Murray CJL, Marcenes W. Global burden of untreated caries: a systematic review and metaregression. *J Dent Res*. 2015;94(5):650-658. <https://doi.org/10.1177/0022034515579703>
- World Health Organization. Global Oral Health Status Report: Towards Universal Health Coverage for Oral Health by 2030. Geneva: World Health Organization; 2022.
- Bomfim RA. Machine learning to predict untreated dental caries in adolescents. *BMC Oral Health*. 2024;24(1):316. <https://doi.org/10.1186/s12903-024-04153-2>
- Litzenburger F, Schäfer G, Hickel R, Kühnisch J, Heck K. Comparison of novel and established caries diagnostic methods: a clinical study on occlusal surfaces. *BMC Oral Health*. 2021;21(1):97. <https://doi.org/10.1186/s12903-021-01453-7>
- Schwendicke F, Tzschoppe M, Paris S. Radiographic caries detection: a systematic review and meta-analysis. *J Dent*. 2015;43(8):924-933. <https://doi.org/10.1016/j.jdent.2015.06.019>
- Biltekin H, Geduk G, Altan A, Karasu S. Evaluation of deep learning systems in detection of dental caries on panoramic radiography. *Am J Dent*. 2025;38(4):163-168.
- Forouzeshefar P, Safaei AA, Ghaderi F, Hashemi Kamangar S, Kaviani H, Haghi S. Dental caries diagnosis using neural networks and deep learning: a systematic review. *Multimedia Tools Appl*. 2024;83(10):30423-30466. <https://doi.org/10.1007/s11042-023-17179-5>
- Lee JH, Kim DH, Jeong SN, Choi SH. Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *J Dent*. 2018;77:106-111. <https://doi.org/10.1016/j.jdent.2018.07.015>
- Maltz M, Jardim JJ, Alves LS. Health promotion and dental caries. *Braz Oral Res*. 2010;24(2):18-25. <https://doi.org/10.1590/s1806-83242010000200003>
- Wang X, Zhang P, Lu H, Luo D, Yang D, Li K, et al. Risk prediction models for dental caries in children and adolescents: a systematic review and meta-analysis. *BMJ Open*. 2025;15(3):e088253. <https://doi.org/10.1136/bmjopen-2024-088253>
- Mallineni SK, Sethi M, Punugoti D, Kotha SB, Alkhayal Z, Mubarak S, et al. Artificial intelligence in dentistry: a descriptive review. *Bioengineering (Basel)*. 2024;11(12):1267. <https://doi.org/10.3390/bioengineering11121267>
- Chauhan RB, Shah TV, Shah DH, Gohil TJ, Oza AD, Jajal B, et al. An overview of image processing for dental diagnosis. *Innovation Emerging Technol*. 2023;10:2330001. <https://doi.org/10.2174/26661610-v10i1-2330001>
- Tsuneki M. Deep learning models in medical image analysis. *J Oral Biosci*. 2022;64(3):312-320. <https://doi.org/10.3164/job.22-0005>
- Vashisht R, Sharma A, Kiran T, Jolly SS, Brar PK, Puri JV. Artificial intelligence in dentistry—a scoping review. *J Oral Maxillofac Surg Med Pathol*. 2024;36(4):579-592. <https://doi.org/10.1016/j.ajoms.2024.02.002>
- White DK, Street CC, Jenkins WS, Clark AR, Ford JE. Panoramic radiograph in pathology. *Atlas Oral Maxillofac Surg Clin North Am*. 2003;11(1):1-53. [https://doi.org/10.1016/s1061-3311\(03\)00024-3](https://doi.org/10.1016/s1061-3311(03)00024-3)
- Mayil M, Keser G, Pekiner FN. Clinical image quality assessment in panoramic radiography. *Clin Exp Health Sci*. 2014;4(3):126-131. <https://doi.org/10.5455/aim.20140326035250>
- Choi JW. Assessment of panoramic radiography as a national oral examination tool: review of the literature. *Imaging Sci Dent*. 2011;41(1):1-14. <https://doi.org/10.5624/isd.2011.41.1.1>
- Mosavat F, Ahmadi E, Amirfarhangi S, Rafeie N. Evaluation of diagnostic accuracy of CBCT and intraoral radiography for proximal caries detection in the presence of different dental restoration materials. *BMC Oral Health*. 2023;23(1):419. <https://doi.org/10.1186/s12903-023-03098-6>
- Albano D, Galiano V, Basile M, Di Luca F, Gitto S, Messina C, et al. Artificial intelligence for radiographic imaging detection of caries lesions: a systematic review. *BMC Oral Health*. 2024;24(1):274. <https://doi.org/10.1186/s12903-024-04050-8>
- Almalki YE, Din AI, Ramzan M, Irfan M, Aamir KM, Almalki A, et al. Deep learning models for classification of dental diseases using orthopantomography X-ray OPG images. *Sensors (Basel)*. 2022;22(19):7370. <https://doi.org/10.3390/s22197370>
- Zhang D. Support vector machine. In: *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval*. Cham: Springer; 2021:201-228. https://doi.org/10.1007/978-3-030-69281-0_8

22. Kang IA, Njimbouom SN, Kim JD. Optimal feature selection-based dental caries prediction model using machine learning for decision support system. *Bioengineering* (Basel). 2023;10(2):245. <https://doi.org/10.3390/bioengineering10020245>
23. Roy A, Chakraborty S. Support vector machine in structural reliability analysis: a review. *Reliab Eng Syst Saf*. 2023;233:109126. <https://doi.org/10.1016/j.res.2022.109126>
24. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297. <https://doi.org/10.1007/BF00994018>
25. Mammone A, Turchi M, Cristianini N. Support vector machines. *WIREs Comput Stat*. 2009;1(3):283-289. <https://doi.org/10.1002/wics.42>
26. Chapelle O, Haffner P, Vapnik VN. Support vector machines for histogram-based image classification. *IEEE Trans Neural Netw*. 1999;10(5):1055-1064. <https://doi.org/10.1109/72.788176>
27. Dal S. Eye state classification from electroencephalography (EEG) signals using the extra trees classifier algorithm. *Eur J Tech*. 2025;15(1):29-36. <https://doi.org/10.31528/ejt.2025.15.1.29>
28. Dal S, Sezgin N. Heart attack classification with a machine learning approach based on the random forest algorithm. *Balkan J Electr Comput Eng*. 2025;13(2):140-147. <https://doi.org/10.17694/bajece.1691905>
29. Géron A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 3rd ed. O'Reilly Media; 2022.
30. Yin M, Wortman Vaughan J, Wallach H. Understanding the effect of accuracy on trust in machine learning models. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019:1-12. <https://doi.org/10.1145/3290605.3300529>
31. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*. 2006:233-240. <https://doi.org/10.1145/1143844.1143874>
32. Markoulidakis I, Markoulidakis G. Probabilistic confusion matrix: a novel method for machine learning algorithm generalized performance analysis. *Technologies*. 2024;12(7):113. <https://doi.org/10.3390/technologies12070113>