Original Research Article

ANALYSIS OF DENTAL CARIES FROM INTRA-ORAL PERIAPICAL RADIOGRAPHS USING MACHINE LEARNING MODELS

Mathuvanti K.K[1], Prabu D[2], Sindhu R[3], Dinesh Dhamodhar[4], Rajmohan M[4], Bharathwaj V.V[3], Sathiyapriya S[3], Vishali M[5]

[1]Bachelor of dental surgery, Department of Public Health Dentistry, S.R.M. Dental College, Ramapuram, Chennai, India.
[2]Professor and Head of Department, Department of Public Health Dentistry, S.R.M. Dental College, Ramapuram, Chennai, India.
[3]Senior lecturer, Department of Public Health Dentistry, S.R.M. Dental College, Ramapuram, Chennai, India.
[4]Reader, Department of Public Health Dentistry, S.R.M. Dental College, Ramapuram, Chennai, India.
[5]Postgraduate student, Department of Public Health Dentistry, S.R.M. Dental College, Ramapuram, Chennai, India.

Received: 10-05-2022 / Revised: 13-06-2022 / Accepted: 15-07-2022
Corresponding author: Prabu D
Conflict of interest: Nil

Abstract

Background:
With the advancement of technology, dentists can improve their performance in identifying various dental caries. In addition, Artificial Intelligence techniques such as Machine learning methods provide a second opinion for dentists on the task of detecting caries.

A.I.M.:
The study aimed to apply caries prediction through machine learning was analyzed in this paper through an experimental analysis.

Materials And Method:
For this study, there was a data set of 700 X-rays gathered from the different age groups .500 X-rays for training the model and 200 X-rays for testing the model. In addition, seven other machine learning techniques: SVM, non-linearSVM+pca, K.N.N., KNN+pca, Decision tree, Decision tree+depth and two deep learning techniques: Custom CNN, Inception net, were applied to this data set.

Result:
We have compared the performance of our proposed system with certified dentists for marking dental caries with working examples. After analogizing the performance of nine different classifiers, the results show that for carries detection CNN method performs better than other methods.

Conclusion:
To date, for the detection of dental caries, visual and clinical examination and diagnostic aids, which are the IOPA's commonly called the X-rays are in practice. This sequela of procedures helps the dentist to identify the cause and treat accordingly. But it is said that in certain cases, dentists miss the appropriate diagnosis if presented with the bitewing radiographs. So in that cases, when the CNN method of machine learning is used in dentistry, it would be a great help to

the dentist in case of appropriate diagnosis, and in normal circumstances also would always help the dentist to recheck the diagnosis whether the diagnosis made is fair and accurate.

Keywords:
Dental caries, artificial intelligence, machine learning, and deep learning technique.

Keywords: mandibular first molar, middle mesial canal, root canal treatment.

Introduction

Dental caries is one of the most prevalent diseases affecting the oral cavity. It is a major oral health problem in most industrialized countries; consequently, dental services related to the oral cavity, including diagnosis, prevention, and treatment of the carries, are the fastest growing sectors in the healthcare industry. Dental caries is a dynamic process where micro-organisms play a major role as the causative factor of disease. The primary and salient micro-organisms causing the infection are Streptococci and Lactobacilli. The salivary flow, ph of saliva before and after sugar intake, sugar-rich diets and carbohydrates, acid-producing bacteria, dental plaque etc., aids in cavity formation.[1] Dental caries is a progressive process starting with affecting the enamel; when not recognized by the patient, neither the dentist will progress into dentin. At the dentin-enamel junction, the patient experiences sensitivity to hot and cold food items. If not treated, it decays the dentin where the patient experiences pain- different types of pain such as throbbing pain, stabbing pain, and nocturnal pain; based on the particular condition, if still not concerned by the patient might even lead to vitality loss of the pulp [2], ie. Pulpal loss where the tooth is considered non-functional. When left untreated, it might even progress to the periapical region, causing periapical conditions or even death in the upper third molar thorough penetration of sinuses when not treated even at the late stage [3]. Diagnosis begins with the patient's medical history, clinical examination, and radiographic examination. The radiographic examination helps the dentists to visualize the progress of the disease. Dentists often use X-rays such as IOPA (Intra-oral periapical radiograph) to assist in locating the extent of dental caries. The dentist's clinical experience is used as additional information to validate their caries' findings on X-rays. It is stated that even experienced dentists miss cavities in the range (of 20%-40%) of the probability if presented with the bitewing radiographs. Automating the dental caries detection process has huge potential to raise standards of medical care by providing increased efficiency and reliability, which would result in appropriate diagnosis and treatment planning.

Artificial intelligence in medicine and related areas has significantly increased in recent years. The creation of computer software algorithms that enable computer programmes to improve via experience automatically is known as machine learning (ML), one of the many types of A.I.

Machine learning is like a format built within the computer; with the help of the pattern inbuilt, the machine can detect the condition.[4]. Machine learning also aids in decision-making based on observation.[5] The application of machine learning techniques is already being used in the field of dentistry for such things as dental radiographic analysis and interpretation. Oliveira et al. used machine learning techniques [6][7]to predict the success of a dental implant. In the framework of this paper, we discuss nine different techniques in machine learning with the data set of 700 X-rays where 500 X-rays for the training model and 200 X-rays for the testing model, to conclude with the best technique among the nine techniques in machine learning which gives the maximum accuracy and assessment for the dentists. So in the future, applying this particular technique in machine learning, which has the maximum F1 Score among the nine methods, would greatly help dentists in many ways. Such as, it is stated as even experienced dentists miss (20%-40%) of carries detection in bitewing radiograph; when carries are in the starting stage and not clinically visible and present with other conditions, there is a high probability of missing caries in such situations this technique would be of great use for the dentist as a second opinion and in reducing the chances of idiopathic injury caused by the dentist. This technique will assist the dentist with the diagnosis and treatment planning.

Methodology:

A Data set characteristic:

Though there are various machine learning methods, they all require a comprehensive dataset for the learning process. We obtained 700 X-rays from different clinics in Chennai, and all X-rays were certified by three dentists after clinical examination for the existence of dental caries. The dentist's reports were the ground truth for training and testing. We used 500X-rays for training our model. The remaining 200 X-rays were used for testing the model. For testing, gave 200 X-rays for marking caries to the dentists, who were unaware of the patient's clinical history. Compared the marking of testing dentists and our model against the ground truth.

Inclusion criteria: all X-rays used for testing and training are present with dental carriers.

Exclusion criteria: all X-rays used for testing and training are not only specific to dental carries but also consist of other minor anomalies like attrition, etc.

The classifiers evaluated:

In this section, we briefly review the four classification techniques used in this work, namely, (1) decision trees, (2) convolution neural networks, (3) kNN, and (4) support vector machines.

Decision Trees are statistical models for classification and data forecast. These models employ the "divide-and-conquer" strategy, which involves breaking down a difficult issue into smaller, simpler sub-models. Applied to each sub-problem [8], We selected the C4.5 algorithm, one of the most widely used methods for creating decision trees, for our project. [9] Quinlan created C4.5 as a software enhancement to the fundamental ID3 algorithm to address various concerns that ID3 does not handle, namely avoiding overfitting. Fitting the data, determining how deeply to grow a decision tree, improving computational efficiency, etc. Quinlan's C4.5 has a factor named confidence that represents a pruning-related component. In general, pruning occurs more when C is smaller. To conduct the studies, we altered the value of the confidence factor to obtain a more accurate classification model. The M.L.P. neural network (Multi-Layer Perceptron) derives from the neural network's Perceptron model. Unlike the fundamental perceptron, M.L.P.s can solve non-linearly separable problems. We have chosen the backpropagation learning algorithm for training M.L.P. neural networks for this work. The weights are modified to train the M.L.P.

network. The output of the network is compared with the desired outcome. The weights are adjusted using the error or the difference between these two. The pace of learning determines the rate of adaptation. The network will swiftly modify its weights if its learning rate is high, but it could become unstable. In practical applications, it is advised to employ small learning rates.

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning method that can take in an input image, give various elements and objects in the image importance (learnable weights and biases), and distinguish between them. ConvNets require substantially less pre-processing than other classification techniques do. Contrary to the hand-engineered filters used in basic approaches, ConvNets are capable of learning these filters. Their attributes ConvNets' structure was influenced by the Visual Cortex's organizational scheme and resembled the neuronal connectivity network in the human brain. Individual neurons react to stimuli only in this constrained area of the visual field, known as the Receptive Field. The entire visual field is covered by a series of such fields that overlap.

The weights are modified to train the CNN. Then, the network output is compared to the desired outcome during training. The error, which is the difference between these two signals, is used to alter the weights of kernels. The pace of learning determines the rate of adaptation. The network will swiftly modify its weights if its learning rate is high, but it could become unstable. Therefore, it is advised to employ small learning rates in practical applications.

K.N.N. is a classical prototype-based (or memory-based) classifier, which is often used in real-world applications due to its simplicity [10]. Despite its simplicity, it has demonstrated impressive classification accuracy in various applications and is frequently used as a benchmark when comparing traditional classifiers. Support vector machine (SVM) is a relatively new classification and regression algorithm that has demonstrated impressive performance in a variety of significant issues [11], [12], [13], [14]. While the *structural risk* considers the complexity of the class of functions used to fit the data and the error in the training set, the empirical risk only examines the mistake in the training set. Although SVMs are widely used in machine learning and pattern recognition, a recent study has demonstrated that they can perform as well as or better than simpler techniques like kNN and neural networks in several classification and regression situations [15].

Experiments:

To evaluate the generalization performance and contrast the classifiers considered for this article, we used 10-fold cross-validation. During 10-fold cross-validation (CV), ten subsets of a given dataset are used. Nine subsets are joined to generate a subset for training a classifier, while the remaining subset is used for testing. [16]. We have also done P.C.A. on images. The test set error, Ei, is calculated ten times, using a different subset as the test set. The mean of the ten mistakes Ei, 1 I 10, is used to calculate the cross-validation error. It is crucial to note that all of the simulations described here used *stratified CV*, in which the subsets were created using the same original patterns' frequency distributions [17]. The principal component analysis is a method for lowering the dimensionality of such datasets and improving interpretability while minimizing information loss (P.C.A.). It accomplishes this by producing fresh, uncorrelated variables that maximize variance one after the other.

Results:

The classification error and the area under the

R.O.C. curve (A.U.C.) are the performance metrics used to compare the classifiers [8], [18], and [19]

R.O.C. curve, also known as a receiver operating characteristic curve, A two-dimensional figure known as a R.O.C. curve, shows how effectively a classifier system performs as the cut-off discrimination value is varied over the predictor variable's range.

In the R.O.C. curve, the x-axis represents, or independent variable is the false positive rate, which identifies normal patterns wrongly classified as novelties; the y-axis or dependent variable is the true positive momentum for the predictive test, which determines the likelihood of prints of the novelty class being recognized correctly true positive/false positive data pair for a cut-off discrimination value of the prediction test is represented by each point in the R.O.C. space. Can produce a R.O.C. curve from the cumulative distribution function if both the true positive and false positive probability distributions are known to need a gold-standard approach for distinguishing genuine positive and true negative situations to get the true positive and false positive rates. We must review the confusion matrix or contingency table to comprehend R.O.C. curves more clearly.

The confusion matrix:

When the truth is known, a confusion matrix—also known as an error matrix—is used to describe the performance of a classifier or classification system. Each row (or column) in a confusion matrix reports the numbers in a true class, such as the number of true dental caries or true normal. In contrast, each column (or row) gives the numbers in a predicted class, such as the number of predicted dental caries or predicted normal. For example, a standard 22 contingency table reports four numbers: There are four types of positive outcomes: true positive (T.P.; also known as sensitivity); false negative (F.N.; a measurement of the proportion of predicted negatives given that it is truly positive); false positive (F.P.; a measure of the balance of predicted positives given that it is truly negative); and true negative (T.N.) (T.N., also known as specificity, is a measurement of the proportion of truly negative predictions to actual negative predictions.). Naturally, a better classifier ought to have greater specificity and sensitivity. Naturally, a better classifier ought to have greater specificity and sensitivity. Remember that specificity equals F.P. - 1.

Table 1: A confusion matrix

		Predicted Condition	
		Carries	Normal
True condition	Carries	True positive (TP) (sensitivity)	False negative (FN)
	Normal	False positive (FP)	True negative (TN) (specificity)

If the discriminating cut-off value for the predictive variable is smaller than the lowest value observed, the (0, 0) point is established in the R.O.C. space. We produce a series of attributes within the R.O.C. space that can be connected by a curve when we increase the discriminating cut-off value to incorporate more and more data points. The (1, 1) point is produced by a discriminative cut-off value greater than the maximum value seen.

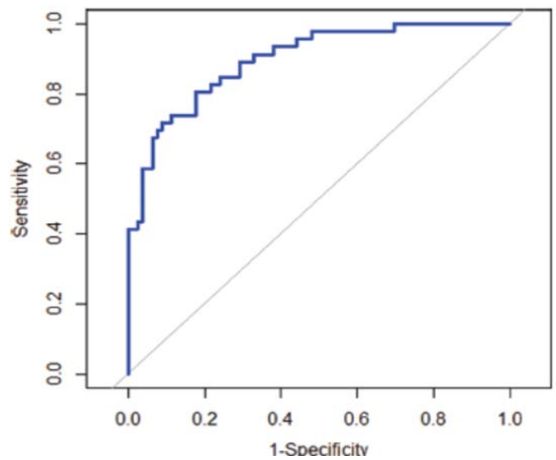

Figure 1:

A hypothetical R.O.C. curve showing the trade-off between sensitivity and specificity is shown in Figure 1. In particular, the relationship between sensitivity and specificity is inverse, meaning that as sensitivity rises, specificity falls, and vice versa.

Figure 1 test predictions don't perform much better than random guesses, according to the diagonal line connecting the (0, 0) and (1, 1) points. The test's predictive value increases with distance from the diagonal line in the R.O.C. space.

Discussion

R.O.C. for ML algorithms

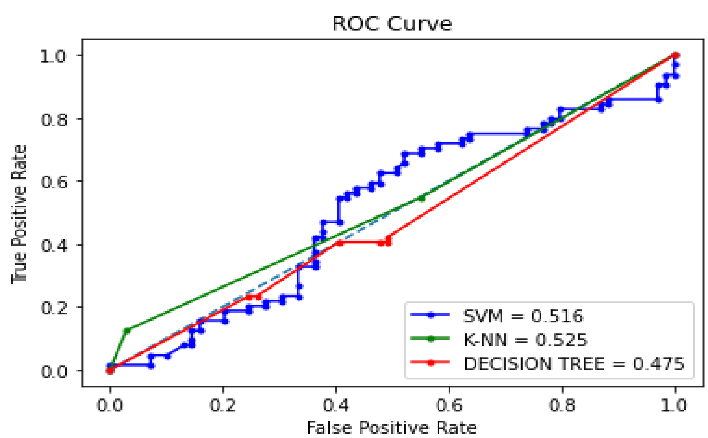

Figure 2:

Figure 2 of the R.O.C. curve X-axis depicts the true positive rate. Y axis represents the false positive rate. According to Huang and Ling, the area under the R.O.C. curve (A.U.C.) summarises the R.O.C. curve and offers an alternative to comparing classifiers based on accuracy. In comparison with other classifiers, the best classifier is the one that obtains an A.U.C. close to 1. For example, figure 2 depicts A.U.C. for SVM is 0.51, A.U.C. for K.N.N. is 0.52 and A.U.C. for Decision tree is 0.47.

Custom CNN

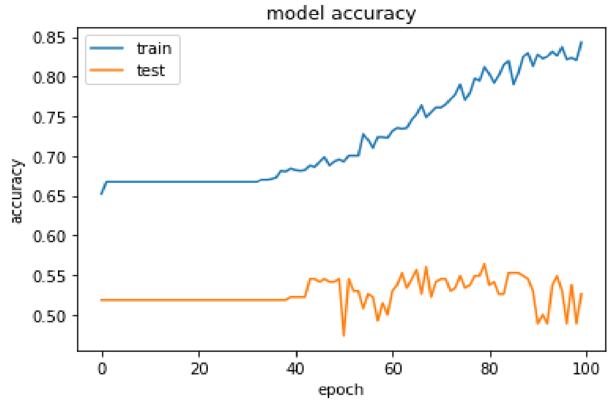

Figure 3:

Inception Net

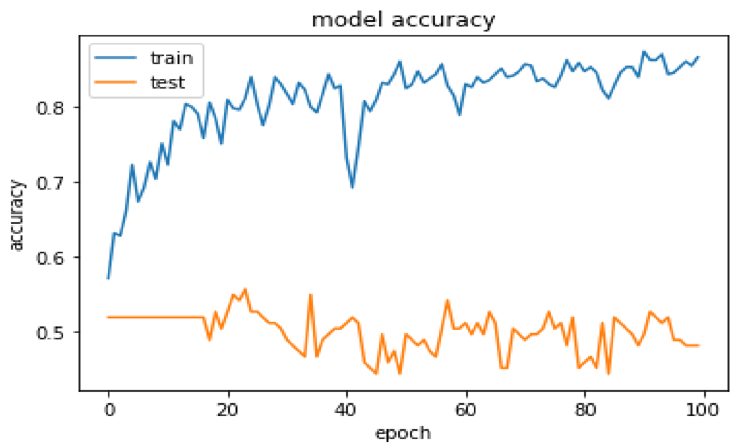

Figure 4:

For both the models custom CNN and inception net, model accuracy is computed and shown in figure3and figure 4, respectively. In the figures X-axis denotes precision, and Y-axis indicates epoch. For training, 5oo x-rays are used, and 200 x-rays are used here for testing. The model accuracy is better for train data.

Table 2: Comparison between the models: SVM, K.N.N., Decision tree and CNN

Model	Accuracy	Precision	Recall	F1 Score	AUC
SVM	51.13%	23.44%	48.39%	31.58%	51.64%
non-linearSVM+ pca=2	55.64%	15.62%	66.67%	25.32%	47.71%
KNN	56.39%	12.50%	80.00%	21.62%	52.46%
KNN +pca=2	53.38%	10.94%	58.33%	18.42%	47.2%
Decision tree	49.62%	23.44%	45.45%	30.93%	49.1%
Decision tree+depth=5	51.13%	25.00%	48.48%	32.99%	53.02%
Decision tree+pca=2	51.13%	26.56%	48.57%	34.34%	48.14%
Custom CNN	53.38%	75.36%	53.61%	62.65%	48.1%
Inception Net	48.12%	78.26%	50.00%	61.02%	46.24%

There was a huge class imbalance, so we took the F1 Score to identify the best model. Concerning F1 Score CNN models, both inception net and custom CNN model produces the best F1 Score, followed by the decision tree model and SVM. This result is attributed to the automatic feature processing in the case of CNN models. However, in the case of decision trees, P.C.A. with two components proved to give better results, implying that the first two principal components were critical to classifying the image.

Conclusion:

We experimented with doing a comparative case study on how different machine learning models classify the X-ray data to detect caries. This study examined the effectiveness of nine classifiers used in caries detection analysis. Figure 2 depicts the R.O.C. for machine learning algorithms such as SVM, K.N.N. and decision tree. Figure 3 and Figure 4 show the model accuracy curves for neural network algorithms. Table 2 compares nine classifiers concerning the accuracy, precision, recall, F1 Score and A.U.C. The F1 Score identifies the best model

among the various measures. The results have shown that the best method for caries detection was obtained by the custom CNN method

with an F1Score of 62.65%

References:

1. Usha C, Sathyanarayanan R. Dental caries-A complete changeover (Part I). Journal of conservative dentistry. 2009;12(2):46-54
2. Diniz MB, Rodrigues JA, Lussi A. Traditional and novel caries detection methods. Contemporary approach to dental caries. 2012; 14:105-28.
3. Srivastava MM, Kumar P, Pradhan L, Varadarajan S. Detection of tooth caries in bitewing radiographs using deep learning. arXiv preprint arXiv.1711.07312. 2017
4. Mayank Srivastava M, Kumar P, Pradhan L, Varadarajan S. Detection of Tooth caries in Bitewing Radiographs using Deep Learning. arXiv.2017 ;arXiv-1711.
5. El-Damanhoury HM, Fakhruddin KS, Awad MA. Effectiveness of teaching International Caries Detection and Assessment System II and its e-learning program to freshman dental students on occlusal caries detection. European journal of dentistry. 2014;8(4):493-497.
6. Oliveira AL, Baldisserotto C, Baldisserotto J. A comparative study on machine learning techniques for prediction of success of dental implants. Mexican International Conference on Artificial Intelligence.Springer 2005;3789: 939-948.
7. Oliveira AL, Baldisserotto C, Baldisserotto J. A comparative study on support vector machine and constructive RBF neural network for prediction of success of dental implants. Iberoamerican Congress on Pattern Recognition.Springer.2005;3773: 1015-1026
8. Gama J. Functional trees. Machine learning. Springer. 2004;1;55(3):219-250.
9. Quinlan J. C4. 5: programs for machine learning. Elsevier; 2014
10. Webb AR. Statistical Pattern Recognition, 2nd edn Wiley.
11. Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995 20(3):273-297.
12. Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis. Cambridge university press; 2004
13. Cristianini N, Shawe-Taylor J. An introduction to support vector machines

and other kernel-based learning methods. Cambridge university press; 2000

14. Sánchez A VD. Advanced support vector machines and kernel methods. Neurocomputing. 2003;55(1-2):5-20.

15. Meyer D, Leisch F, Hornik K. The support vector machine under test. Neurocomputing. 2003 1;55(1-2):169-186.

16. Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. Artificial intelligence in medicine. 2005 1;34(2):113-127.

17. Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations Morgan Kaufmann. San Francisco. 2000.

18. Huang J, Ling CX. Using A.U.C. and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering. 2005 31;17(3):299-310.

19. Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. Journal of biomedical informatics. 2005 Oct 1;38(5):404-415.

20. Farsi NM, Salama FS, Pedo C. Sucking habits in Saudi children: prevalence, contributing factors and effects on the primary dentition. Pediatric dentistry. 1997 ;19:28-33.

21. Fawcett T. An introduction to R.O.C. analysis. Pattern recognition letters. 2006 1;27(8):861-874.