

Article

Optimal Feature Selection based Dental Caries Prediction Model Using Machine Learning for Decision Support System

In-Ae Kang¹ and Jeong-Dong Kim^{2*}

¹ Department of Computer and Electronics Convergence Engineering, Sun Moon University, Asan-si 31460, Korea; inae5004@sunmoon.ac.kr

² Department of Computer and Electronics Convergence Engineering, Sun Moon University, Asan-si 31460, Korea; kjd4u@sunmoon.ac.kr

* Correspondence: kjd4u@sunmoon.ac.kr

Abstract: Caries is a prevalent oral disease that primarily affects children and teenagers. Advances in machine learning have caught the attention of scientists working with decision support systems to predict early tooth decay. Current research has developed machine learning algorithm for caries classification and reached high accuracy especially in ML for image data. Unfortunately, most studies on dental caries only focus on classification and prediction tasks, meanwhile dental caries prevention is more important. Therefore, this study aims to design an efficient feature for decision support system machine learning based that can identify various risk factors that cause dental caries and its prevention. The data used in the research work was obtained from the 2018 Korean Children's Oral Health Survey, which totaled nine datasets. The experimental results show that combining the mRMR and GINI Feature Importance methods when training with the GBDT model achieved the optimum performance of 95%, 93%, 99%, and 88% for accuracy, F1 score, precision, and recall, respectively. So, the proposed method has provided effective predictive model for dental caries prediction.

Keywords: Disease Dental Caries, Gradient Boosting Decision Tree, Feature Selection, Machine Learning, Feature importance

1. Introduction

Oral health is important for general health and quality of life. Oral health means being free from throat cancer, infections and sores in the mouth, gum disease, tooth loss, dental caries, and other diseases so that no disturbances limit biting, chewing, smiling, speaking, and psychosocial well-being. One that is related to oral health is dental health. Dental caries is a dental health disorder. Dental caries is formed because there is leftover food that sticks to the teeth, which in turn causes calcification of the teeth. As a result, teeth become porous, hollow, and even broken. Dental caries is a disease associated with the hard tissues of the teeth, namely enamel, dentin, and cementum, in the form of decayed areas on the teeth, occurring as a result of the process of gradually dissolving the mineral surface of the teeth and continuing to grow to the inside of the teeth.

Global Burden research reported at 2019 dental caries is the most common oral disease affecting approximately 3.5 billion people, which are 2 billion suffer from permanent dental caries [1]. Moreover, in 2020 more than 6 million patients in the Republic of Korea visited dentists of which 1.45 million were children (0~9 years old) [2]. So dental caries disease is a challenge for scientists to solve these problems.

Recent research on dental caries was conducted by Rimi et al. [3], who discussed the prediction of dental caries using Machine Learning(ML). In their experiments, they used nine algorithms, namely K-Nearest Neighbors (KNN), logistic regression (LR), support vector machine (SVM), random forest (RF), naïve Bayes (NB), classification and regression trees, multilayer perception (MLP), linear discriminant analysis (LDA), and adaptive

boosting (AdaBoost). The best results were obtained through logistic regression (LR) with an accuracy of 95.89%.

Meanwhile, to detect dental caries, researchers use data images such as those done by Zhang et al [4], Lee et al [5], Estai et al [6], and Leo et al [7]. Zhang et al. [4] used the convolutional neural network model method to classify 3,932 images of dental caries. They managed to get a value of 85.65% for the area under the curve (AUC). In their research, Lee et al. [5] proposed a U-shaped Deep Convolution Neural Network (U-Net). The highest sensitivity value in his research was 93.72%. Estai et al. [6] proposed the Inception-ResNet-v2 model to classify as many as 2,468 images. This number was less than the study by Leo et al. [7], which used 3,000 images to be classified using the transfer learning method based on the InceptionV3 model.

This study aims to investigate important features of dental caries data to make a DSS that can more accurately identify the causal factors of dental caries. In addition, DSS results will suggest strategies for early diagnosis and prevention of dental caries. To achieve this goal, the Decayed-Missing-Filled-Teeth (DMFT) questionnaire of the 2018 Oral Health Survey was used. Based on this data, five feature selection methods were applied to obtain an optimized data set. The presented model uses Gradient Boosted Decision Tree (GBDT), RF, LR, SVM, and Long and Short-Term Memory (LSTM) models. In addition, cross-validation and hyperparameter tuning were used to improve the performance of the classification model. Through this, it is possible to accurately diagnose dental caries prediction. The ML models described in this study efficiently provides patients with superior quality dental services including diagnosis and treatment.

The rest of this paper is structured as follows: Section 2 presents the Related works. The related studies were written classified into two, dental caries research using image data and dental caries research using survey data. In Section 3, the caries prediction model proposed in this study was described. The experimentation and results are described in Section 4. Section 5 presents discussion, while Section 6 concludes the paper.

2. Related works

Recently, various research on DSS for diagnosing dental caries based on ML and AI have been conducted. Among the dental caries identification cases, that were investigated are selected works that utilize ML and AI methods only based on survey data.

2.1. Case of caries classification using image data

Mohammad-Rahimi Hossein et al. [9] compared the accuracy of deep learning models from various image data used in dentistry (intraoral photography, apical radiation, bitewing radiation, near-infrared transmitted illumination, optical coherence tomography, and panoramic radiation). The results of performance in accuracy of 78.0% for the near-infrared transmitted illumination picture and 96% for the rest of the image dataset [9]. Dental caries discrimination using the Histogram of Oriented Gradient (HOG) imaging technique and SVM [10] that performs training with 60 X-ray images, and validation was performed with 15 images with a result of testing with 25 images, the precision was 92.4%.

In the study of Lian Luya et al. [11] a deep learning model and a professional dentist evaluated dental caries discrimination using 1,160 panoramic films. The deep learning model applied nnU-net and DenseNet121 to classify carious lesions and lesion depths. The accuracy of nnU-Net is 0.986, and the highest accuracy of DenseNet121 achieve 0.957. The results were similar to the values of dental caries identification accuracy of 6 experienced dental specialist. Similar research cases include a convolutional neural network model created on 3,686 X-ray images and the study evaluated the presence or absence of dental caries by 4 dentists [12]. Intersection-over-Union was used as a validation metric by applying a CNN (U-net). As a result of comparing the performance of 7 independent neural networks, the average accuracy of the neural network was 80%, and that of the dentist was 71% on average. The study of Muktabh Ma-yank Srivastava et al. [13] is a case of discriminating dental caries learned with a neural network consisting of more than 100

Fully Convolutional Neural Network CNN layers. In addition, 150 out of 3,000 X-ray images learned were used as a validation set. The proposed system showed an F1-Score of 0.7, which was 0.14 - 0.2 higher than dental experts.

ML methods such as Soft-max classifier and Linearly Adaptive Particle Swarm Optimization (LA-PSO) and then using a Feed-Forward Neural Network (FFNN) are also used to discriminate dental caries. The results of the studies achieved of each study obtained very good accuracies of 97% and 99% [14,15]. In addition, training was conducted in GoogLeNet Inception V3 using X-ray images consisting of 1,500 images of damaged teeth and 1,500 images of normal teeth with image size changed to 299 * 299 pixels and performed in 82% accuracy. Vinayahalingam et al. [16] A classification study was conducted to classify carious lesions on molars with a dataset of 400 panoramic images was trained with CNNMobileNetV2, achieving accuracy 0.87 and AUC 0.90 for classifying carious lesions of the third molar. Berdouses et al. [18] classify dental caries by applying the feature selection technique to image data by radon transformation and discrete cosine transformation. The highest accuracy among these eight classifiers is 86% for the RF algorithm. Continuing with Wang et al. [18] were select 36 features with the K-means algorithm, also applies J48, Random Tree(RT), RF, SVM, Naive Bayes algorithms to achieve highest precision and recall of 86% in the RF. Liu et al. [19] focus on prevention strategies for dental caries in the elderly, by oral assessments, including surveys, were conducted on 1,144 elderly Chinese. The selected variables were trained with GRNN and an unconditional logistic regression algorithm. As a result of the study, the area under the ROC curve of the GRNN algorithm was 0.626, and the value was 0.002. The caries risk variables of residents were toothache and smoking habits and can be utilized for early diagnosis and treatment planning of dental caries in the elderly.

Wang et al. [20] designed a parent and child toolkit to predict Children's Oral Health Status Index (COHSI) scores and Recommendations For Treatment Needs (RFTN). The COHSI program was developed from survey responses of 545 families with children between the ages of 2 and 17. The survey response data were trained with Gradient Boosting and Naive Bayesian algorithms. The results expressed the probability of needing treatment and the performance of the toolkit in terms of correlation, residual mean square error (RMSE), sensitivity, and specificity. The correlation of COHSI was 0.88 (percentile 0.91), and the RMSE for COHSI was 4.2 (percentile 1.3).

2.2. Case of caries classification using survey data

Hung et. al. [21] identifies important factors in the identification of root caries from US National Health and Nutrition Examination Data Survey. The study's most effective method, SVM showed 97.1% of root caries specification with the main cause of root caries is age and identifies problems with longitudinal data, and the study's predictions were severely limited. The study of Karhade et. al [22] predicts the prevalence of Early Children Caries (ECC) in infants through clinical, demographic, behavioral, and parent reported oral health status. A set of 10 ECC predictors closely related to ECC induction were deployed with AutoML on Google Cloud, and a survey was conducted on the presence or absence of dental caries. Using the results of this investigation, ECC classification accuracy (Area Under the ROC curve [AUROC], sensitivity [Se], and positive predictive value [PPV]) was evaluated. As a result of the study, the model performance of single-item reporting was the highest, with AUROC 0.74, Se 0.67, and PPV 0.64. In a similar case study, Ramos-Gomez et al. [23] identifies variables for induced dental caries in infants 2-7 years of age living in Los Angeles. A random forest algorithm is trained to identify dental caries predictors. The most influential variable was the parent's age (MDG = 2.97, MDA = 4.74), the presence or absence of dental health problems in infants within 12 months (MDG = 2.20, MDA = 4.04).

You-Hyun Park et. al [24] researched a predictive model for caries in infancy utilizing logistic regression, XGBoost (version 1.3.1), RF, and LightGBM (version 3.1.1) algorithms. Feature selection is performed through regression-based reverse removal and

random forest-based permutation importance classifier. The results of this study had AU-ROC values of LR 0.784, XGBoost 0.785, RF 0.780, and LightGBM 0.780.

Table 1. The related works on the classification of dental caries.

References	Dataset	Models	Performance
Ainas A. ALbahbah et. al [10]	60 X-ray pictures	SVM	92.4%
Luya Lian et. al [11]	1,160 panoramic films	nnU-Net DenseNet121	98.6% 95.7%
Anselmo GarciaCantu et. al [12]	3,686 bitewing radiographs	CNN	80%
Shankeeth Vinayahalingam et. al [16]	400 cropped panoramic	MobileNet V2	87%
Elias D. Berdouses et. al [18]	91 posterior extractions and 12 in vivo human teeth image	J48	78%
		Random Tree	73%
		Random Forest	86%
		SVM	63%
		Naive Bayes	49%
Lu Liu et. al [19]	1,144 elderly questionnaires	LR	84%
		GRNN	85%
Y. Wang et. al [20]	Survey responses from 545 families	Gradient Boosting Naive Bayesian	Correlation 0.88
Karhade, Deepti S. et. al [22]	6,404 children aged 3-5 (average age 54 months)	AutoML ECC classifier	80%
Francisco Ramos-Gomez et. al [23]	182 guardians with children ages 2 to 7	RF	70%
		LR	78.4%
You-Hyun Park et. al [24]	4,195 Survey data	XGBoost	78.5%
		RF	78.0%
		LightGBM	78.0%

3. Dental Caries Prediction

In this study, we present an optimal dataset and prediction model for predicting dental caries. In order to select the optimal dataset, Chi-square, Relief F, minimum Redundancy Maximum Relevancy(mRMR), Correlation and one feature importance (GINI) method were applied. Five prediction models were trained using the reduced feature subsets, and the performance of each model was compared. Among them, the data set and model that showed the best performance are presented. Figure 1 depicts our proposed dental caries prediction model.

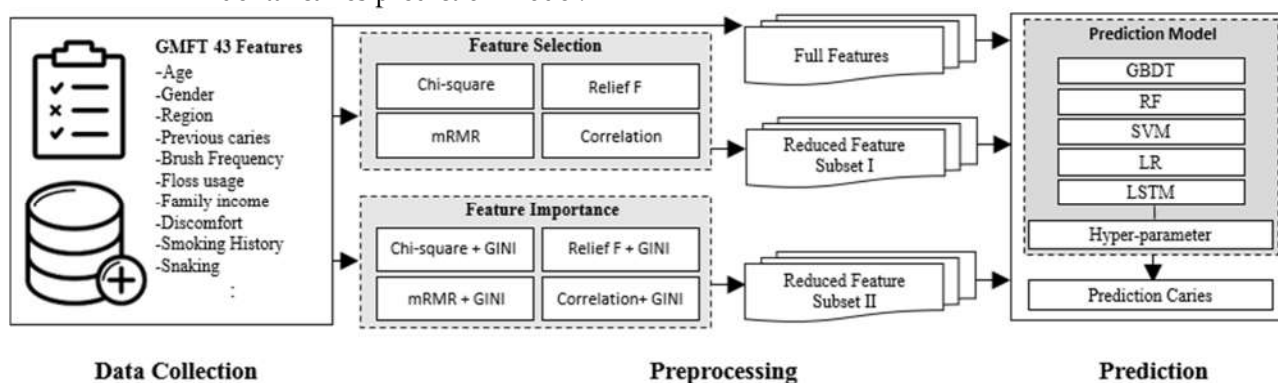


Figure 1. Proposed model for dental caries prediction.

3.1. Data Collection

The dataset used in this study is the 2018 Children's Oral Health Survey conducted by the Korea Centers for Disease Control and Prevention. Data of dentists visiting each institution for oral examination and survey. A total of 22,287 respondents were surveyed

and the oral health awareness survey was selected and used in the survey. The oral health awareness survey consisted of a total of 43 items and 1 label, including age, gender, place of residence, snack frequency, tooth brushing frequency, oral care use, smoking experience, oral health awareness and behavior, and act_caries as a label. This data is not subject to Institutional Review Board (IRB) approval as it does not record patient personal information. The composition of the detailed data set is attached as Appendix A.

3.2. Data Preprocessing

Data cleaning of ML models is an essential step to increase the efficiency and accuracy of the models [25]. Data cleaning of ML models is an essential step to increase the efficiency and accuracy of the models. Empty or useless samples were removed from the raw data. Then, it was scaled to have a value between 0 and 1 using the Min-Max Scaling technique.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

In equation (1), x' is the predicted value, x is the original value, $\min(x)$ is the minimum value of the column, and $\max(x)$ is the maximum value. Min Max Scaler is a way to rescale all characteristic values between 0 and 1. This method reacts fairly quickly in the presence of outliers and does not change the original content of the data. The study adopted the Synthetic Minority Oversampling Technique (SMOTE). The SMOTE algorithm is the most widely used model among oversampling techniques as a method of generating synthetic data. SMOTE is a synthetic minority sampling technique that samples a majority class and interpolates existing minority samples to synthesize a new minority instance [26].

3.3. Feature Selection

Feature selection is a technique to reduce the number of input variables for many dataset features. The computational complexity of the model increases as the input variables increase. Optimized subsets through feature selection can increase the accuracy of the classifier model and reduce the amount of computation required to train the model [27]. Section 3.3 describes the Chi-square, Relief F, mRMR, Correlation, and feature importance techniques(GINI) applied to this experiment.

3.3.1. Chi-Square

The Chi-Square method is a nonparametric test in which the dependence between the independent features and the target variable is expressed numerically. All predictive features and statistics between classes are measured, and the feature with the highest chi-square score is selected [28, 29]. The mathematical expression for the chi-square is :

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (2)$$

where, c is the degree of freedom, O is the observed value, and E is the expected value.

3.3.2. Relief F

The Relief F algorithm does not restrict data types as a filter-based feature selection. Effective handling of nominal or continuity features, missing data, and noise tolerance [30]. This algorithm distinguishes whether the classifications are strongly or weakly correlated. If the classifications are strongly correlated, treat them as similar samples and keep those samples close together. On the contrary, samples with weakly correlated classifications are kept away. The feature weights are calculated by computing the nearest neighbor samples' within-class and between-class distances. This operation is repeated in order to update the weight vectors of features, and the weights of all features are eventually yielded [31].

The formula used in updating the weight value of features by the Relief F algorithm is given as, [32]

$$W[A] = W[A_0] - \frac{\sum_{j=1}^k \text{diff}(A, x_i, H)}{mk} + \sum_{C \neq \text{class}(x_i)} \frac{p(C)}{1-p(\text{class}(x_i))} \cdot \frac{\sum_{j=1}^k \text{diff}(A, x_i, M(C))}{mk} \quad (3)$$

where, the weight coefficient determined at A_0 is $W[A]$. The original dataset feature set is represented by $W[A_0]$. x_i is sample, and H is the sample of the closest class to which x_i belongs. The difference between x_i and H for each attribute of A is represented by the formula $\text{diff}(A, x_i, H)$. The Manhattan distance between the values of the features is calculated by the $\text{diff}(A, x_i, H)$. A for two boundary conditions x_i and H , where k is the number of closest neighbors and m is the total number of iterations. The ratio of sample C to all samples is known as $p(C)$. The percentage of samples in the class to which sample x_i belongs to the entire sample is expressed as $p(\text{class}(x_i))$. The difference between x_i and $M(C)$ for each feature of A is represented by the expression $\text{diff}(A, x_i, M(C))$.

3.3.3. minimum Redundancy - Maximum Relevance (mRMR)

The features are rated according to how relevant they are to the target variable in the mRMR feature selection approach. The redundancy of the features is taken into account when ranking them. The feature with the highest rank in mRMR is the one with the most relevance to the target variable and the least redundancy among the characteristics. Using Mutual Information, redundancy and relevance are measured (MI) [33, 34].

$$j(X_k) = I(X_k; y) - \frac{1}{|s|} \sum_{x_i \in s} I(X_i; X_k) \quad (4)$$

where, y represents the target variable, while s represents the set of selected features. x_s is one of the characteristics of set s that has been selected, whereas X_i is a feature that has not yet been selected [35].

3.3.4. Correlation

Correlation analysis is a method that analyzes the linear relationship between two variables measured as curb variables. It analyzes whether variable B increases or decreases as variable A increases. Correlation analysis has various analysis methods, such as Pearson correlation analysis and Spearman correlation analysis, and this study conducted experiments using Pearson correlation analysis. The closer the coefficient is to 1, the more significant the correlation, and the closer to -1, the inversely proportional. Each coefficient has a value of +1 if it is precisely the same, 0 if it is completely different, and -1 if it is precisely the same in the opposite direction [36].

3.4. Feature Importance

3.4.1. GINI

GINI impureness (or GINI index) is a statistic utilized in decision trees to determine how to divide the data into smaller groups. It estimates the frequency with which a randomly selected element from a set would be erroneously classified if randomly labeled according to the distribution of labels in a subset [37]. GINI importance (also known as mean decrease in impurity) is one of the most often used processes for estimating feature importance. It is simple to implement and readily available in most tree-based algorithms, such as Random Forests and gradient boosting. The GINI significance of a feature gauges its effectiveness in minimizing uncertainty or variance in decision tree creation. Thus, each time a split occurs in a tree regarding a particular characteristic, the GINI impurity is added to its total importance [38].

3.5. Prediction Models

The predictive model of this study was designed to predict the presence or absence of caries as per the dataset of the 2018 child oral health examination results. Preprocessed

data from children's oral health examinations are subjected to Chi-square, Relief F, mRMR, Correlation, and GINI method to identify valuable features. Selected subsets are trained with GBDT, RF, LR, SVM and LSTM algorithms. Then, the caries prediction performance of the model was evaluated by comparing the predicted output values of the algorithm with the actual values. The following describes the characteristics of the algorithm tested in this study.

3.5.1. GBDT (Gradient Boosting Decision Tree)

Gradient boosting is a type of ML technique used for regression and classification problems, with its weak prediction model (typically the decision tree) generating a forecast model in the form of a collection. It, like other strengthening approaches, constructs the model in stages and allows the optimization of the loss function of any separable variables to a generalized model [39,40].

Step 1 : The initial the model with constant γ , Where $L(y_i, \gamma)$ is the loss function.

$$F_0(X) = \operatorname{argmin} \gamma \sum_{i=1}^m L(y_i, \gamma) \quad (5)$$

Step 2 : The residual along the gradient direction is written as,

$$\hat{y}_i = - \left[\frac{\partial L(y_i, F(X_i))}{\partial F(X_i)} \right]_{f(x)=f_{n-1}(x)} \quad (6)$$

where, n indicates the number of iterations, and $n = 1, 2, \dots, m$.

Step 3 : The initial model $T(x_i, \alpha_n)$ is obtained by fitting sample data, and the parameter α_n is calculated based on the least square method as,

$$\alpha_n = \operatorname{argmin} \alpha, \beta \sum_{i=1}^m (\hat{y}_i - \beta T(x_i; \alpha))^2 \quad (7)$$

Step 4 : By minimizing the loss function, the current model weight is expressed as,

$$\gamma_n = \operatorname{argmin} \gamma \sum_{i=1}^m L(y_i, F_{n-1}(x) + \gamma T(x_i; \alpha_n)) \quad (8)$$

Step 5 : The model is updated as,

$$F_n(X) = F_{n-1}(x) + \gamma T(x_i; \alpha_n) \quad (9)$$

This loop is performed until the specified iterations times or the convergence conditions are met.

3.5.2. RF (Random Forest)

RF is another name for Random Decision Forest (RDF), and it is used for classification, regression, and other tasks that require the construction of multiple decision trees. This RF Algorithm is based on supervised learning, and it has the advantage of being used for both classification and regression. The RF Algorithm outperforms all other existing systems in terms of accuracy, and it is the most widely used algorithm [41]. After obtaining the model output value r_i of the Out Of Bag (OOB) dataset and the model output value e_i of the data set in which the variable is a random value, the variable importance \bar{d} can be calculated by calculating as follows.

$$d_i = e_i - r_i \quad (10)$$

$$\bar{d} = \frac{1}{N} \sum d_i \quad (11)$$

where, \bar{d} is a variable representing a change in performance when a specific variable is randomly transformed. Therefore, the higher \bar{d} the greater the importance of the variable [42].

3.5.3. SVM (Support Vector Machine)

SVM are kernel-based ML models that define decision boundaries. As the number of properties increases, the decision boundary becomes higher order, called hyperplane [43].

The fundamental reason for using SVM is to separate numerous classes in the training data using a surface that maximizes the margin between them. In other words, SVM allows a model's generalization ability to be maximized. This is the goal of the Structural Risk Minimization principle (SRM), which allows for the minimization of a bound on a model's generalization error rather than minimizing the mean squared error on the set of training data, which is the commonly used by empirical risk minimization methods [44].

Step 1 : The hyperplane is defined.

$$g(x) = w \cdot x + b = 0 \quad (12)$$

where, w is a vector, and b is an offset between the origin plane and the hyperplane.

Step 2 : Transform the objective function into a double optimization.

$$\min_{w,b} \max_{\alpha} L(w,b,\alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i [y_i (w \cdot x_i + b) - 1] \quad (13)$$

$$(0 \leq \alpha_i \leq c, i = 1, 2, \dots, n) \quad (14)$$

$$\text{s.t. loss function : } \sum_{i=1}^n [y_i (w \cdot x_i + b) - 1] \quad (15)$$

$L(w, b, a)$ is a Lagrangian function, i is a Lagrangian coefficient, c is the penalty coefficient, which is the upper bound of i [44].

3.5.4. LR (Logistic Regression)

LR is a mathematical model that estimates the likelihood of belonging to a specific class. The LR model is used for binary classification in this paper, but it can easily be extended for multi label classification in other cases. The formula for linear estimation is expressed as follows [45].

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (16)$$

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + \varepsilon_i \quad (17)$$

$$Y_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i \quad (18)$$

Where, β_0 is the intercept, β_1 is the slope, which is a regression coefficient in the form of a column vector, and ε_i is the error term [46]. The formula in (16) refers to the description of single linear regression, and the formula in (17) is for multiple linear regression with several independent variables. Equation (18) is an expression representing the least squares method. The above equation shows that a straight line with a small number of errors is an optimal straight line. To estimate the optimal straight line, finding a value that minimizes the sum of squared errors between the predicted and observed values is necessary. The measure of the degree of prediction error is called the coefficient of determination [47].

$$g(z) = \frac{1}{(1+e^{-z})} \quad (19)$$

Equation (19) is a sigmoid with an output value of 0 or 1. In this experiment, a value of 0 was expressed as a patient without dental caries, and a value of 1 was expressed as a patient with dental caries [47].

3.5.5. LSTM (Long Short-Term Memory)

Recurrent neural networks (RNN) with LSTM have emerged as an effective and scalable solution for various learning problems using sequential data [48]. Because they are broad and practical and excellent for capturing long-term temporal dependencies. The LSTM is an RNN-style architecture with gates that regulate information flow between cells. The input and forget gate structures can modify the information as it travels along the cell state, with the eventual output being a filtered version of the cell state based on the input context [49]. The mathematical expression for the LSTM algorithm is:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (20)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (21)$$

$$C_t = f_t * C_{t-1} + i_t * \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (22)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (23)$$

$$h_t = o_t * \tanh(C_t) \quad (24)$$

where i_t , O_t , and f_t are the activation vectors for the input, output, and forget gates, respectively; C_t and h_t are the activation vectors for each cell and memory block, respectively, and W and b are the weight matrix and bias vector, respectively. Furthermore, $\sigma(\cdot)$ signifies the sigmoid function [49].

4. Experimentation

This section describes the experiments of the caries prediction model and their results. In this experiment consists of two parts. The first step was to generate various subsets by applying feature selection and feature importance methods to the data set. In the second step, five ML models are applied using subsets, the performance of each model is compared, and the model with the best performance is selected.

4.1. Dataset

The 2018 pediatric oral health examination data were used to train and evaluate the proposed method's effectiveness and performance. There is a total of forty-three features in this dataset. Chi-square, relief F and mRMR approaches were applied and set the number of feature (k) to forty-three, forty, thirty-five, thirty, twenty-five, twenty, fifteen, ten, and five, for experiment respectively. Correlation approaches also applied. In addition, GINI feature importance was applied to these four methods (Chi-square, Relief F, mRMR, and Correlation). A total of eight methods are applied in the dataset. In this experiment, the correlation technique set the threshold and fixed to 0.85 in every experiment. For correlation approach the model is trained by using forty-two features regardless of threshold. The table below shows the number of features to which the feature selection technique and feature importance technique were applied, excluding the set of all features. In the study, feature importance were used with GBDT, RF, and LR models except for LSTM and SVM. Because SVM lacks feature importance properties, they cannot be applied. For LSTM only the permutation importance algorithm can be applied [50]. The detailed data set is attached as Appendix B.

4.2. Hyper parameter of Differnet Machine Learning Models

In order to predict whenever a tooth has caries, an optimal dataset and ML training are required. In this study, five models were used, with 20% of the total data were put aside as test data and the remaining 80% as train data to determine the suitable model for caries prediction. As each dataset involves several validation steps, hyperparameter tweaking was carried out in this experiment utilizing random rather than grid search to maximize experimental efficiency.

Table 2. Optimizable parameters for different models.

Models	Dataset	Parameter	Range
GBDT	mRMR + GINI	subsample	0.80
		n_estimators	200
		min_samples_leaf	8
		max_features	5
		max_depth	320
		learning_rate	0.02
RF	Relief F + GINI	n_estimators	320
		min_samples_leaf	1
		max_features	5
		max_depth	None
LR	Chi-square + GINI	solver	sag
		penalty	l2
		C	5
SVM	Relief F	probability	True
		kernel	rbf
		gamma	0.01
		C	10
LSTM	Chi-square	learning rate	0.001
		beta_1	0.09
		beta_2	0.999
		epsilon	1e-2
		epochs	100

4.3. Results

In this experimentation divided the complete data into training and test data, which were 80% and 20% respectively to predict caries. Only the high accuracy of each experimental condition is described in the results presented below which are the outcomes of the test data.

4.3.1. Performance of the Classifiers without Feature Selection

This subsection presents the experimental results obtained when the Full feature set was used to train the GBDT, RF, LR, SVM, LSTM models. These results are tabulated in Table 3. Additionally, Figure 2 shows the AUC values and the ROC curves of the GBDT, RF, LR, SVM, LSTM models.

Table 3. Performance of the classifiers trained with the full features.

Models	#of features	F1-Score	Prediction	Recall	Accuracy
GBDT	43	0.8635	0.9490	0.7921	0.8966
RF		0.8868	0.9186	0.8572	0.9105
LR		0.7773	0.7959	0.7598	0.8203
SVM		0.7862	0.7434	0.8345	0.8128
LSTM		0.7575	0.7428	0.7436	0.7467

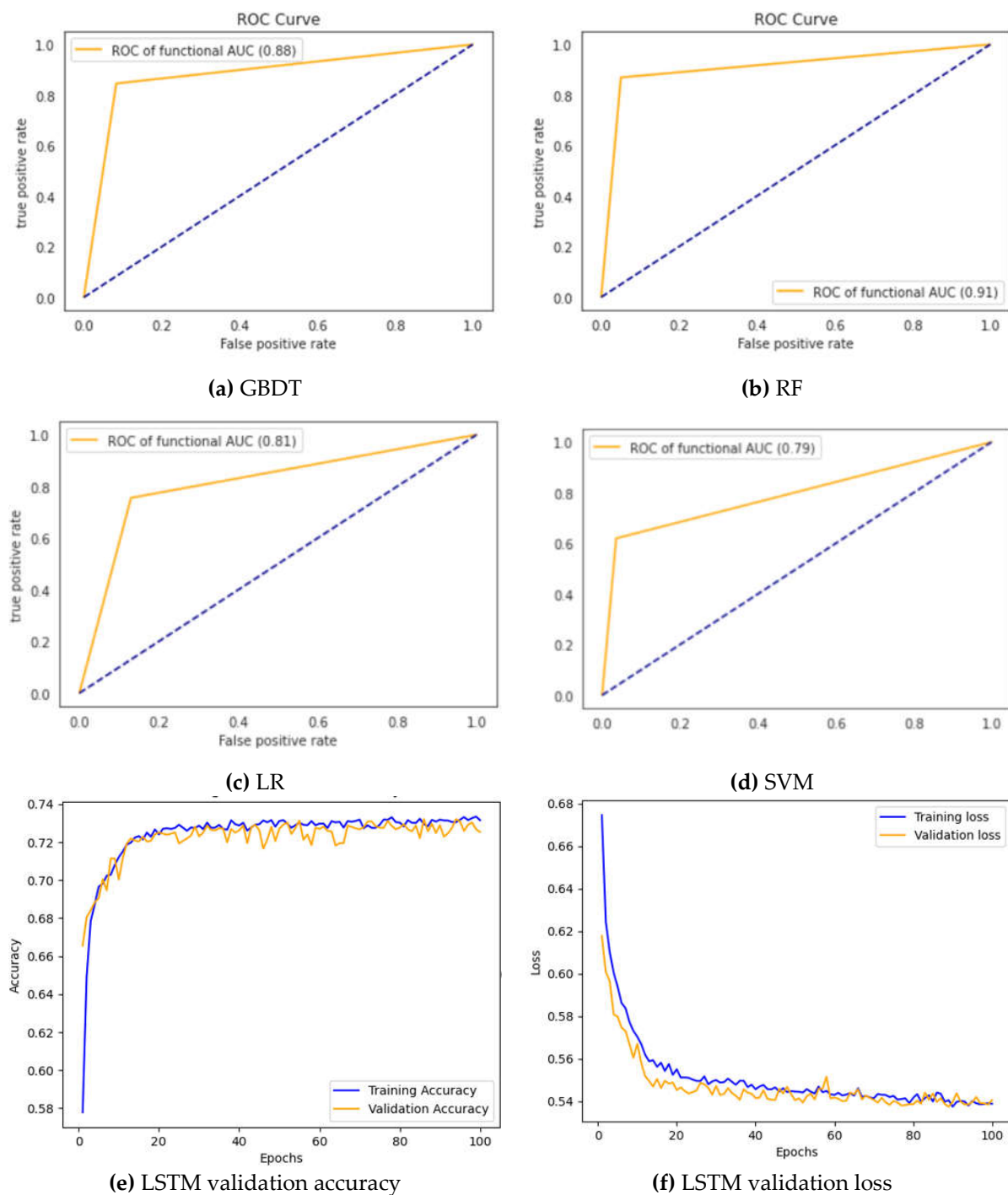


Figure 2. Evaluation of different models used (a), (b), (c), (d) ROC curve of the classifiers trained using the full features. (e), (f) Training vs. validation loss LSTM.

4.3.2. Performance of the Classifiers after Feature Selection

Various reduced feature subsets were produced in this study using Chi-square, Relief F, mRMR, Correlation, and GINI methods. The reduced feature subset is used to train the classifier model. The experimental results are shown in table 4 and figure 3 that the suggest GBDT model obtained a f1-score, precision, recall, and accuracy of 0.9379, 0.9984, 0.8844, and 0.9519, respectively, which outperformed the RF, SVM, LSTM, LR conventional GBDT. Also it is observed that the performance of the various classifiers in Table 4 is better than their corresponding performance in Table 3. This improvement demonstrates the effectiveness of the feature selection. Therefore, the mRMR + GINI feature selection and GBDT model is an effective method for predicting dental caries.

Table 4. Performance of the classifiers trained with the subset.

Models	Feature selection	#of features	F1-Score	Prediction	Recall	Accuracy
GBDT	mRMR + GINI	18	0.9379	0.9984	0.8844	0.9519
RF	Relief F + GINI	20	0.9372	0.9978	0.8835	0.9513
LR	Chi-square + GINI	40	0.7814	0.8012	0.7625	0.8256
SVM	Relief F	43	0.8806	0.9028	0.8596	0.9039
LSTM	Chi-square	15	0.8300	0.8400	0.8300	0.8400

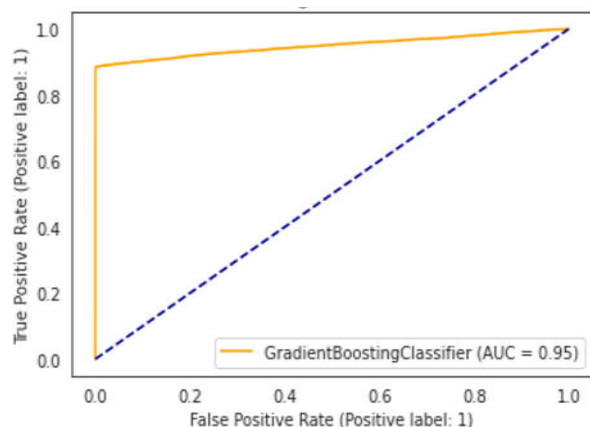
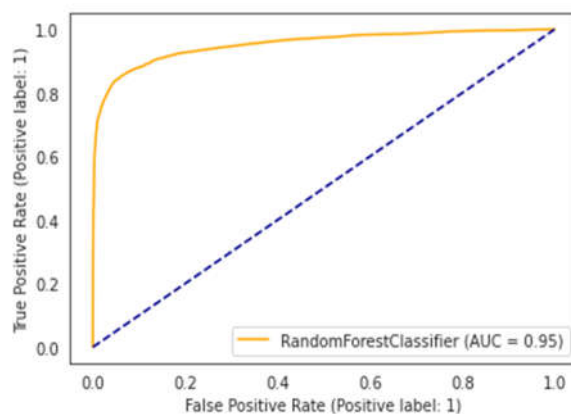
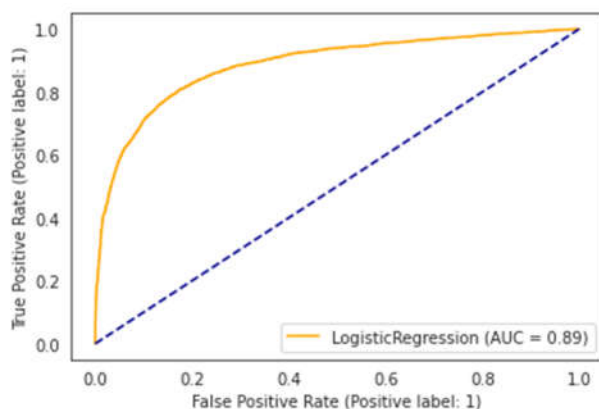
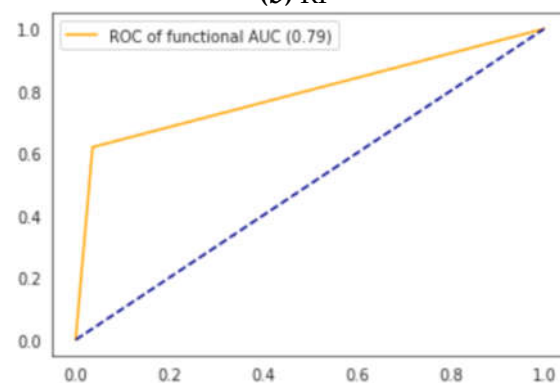
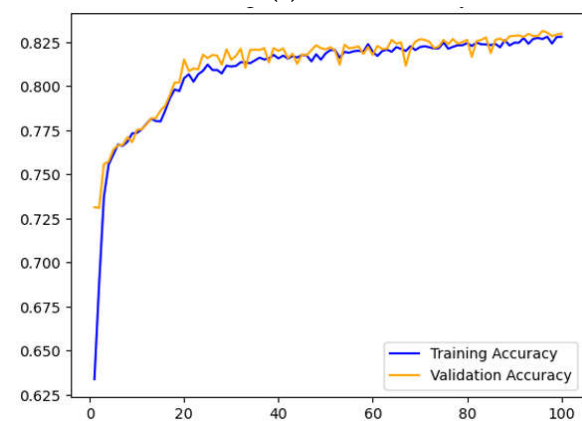
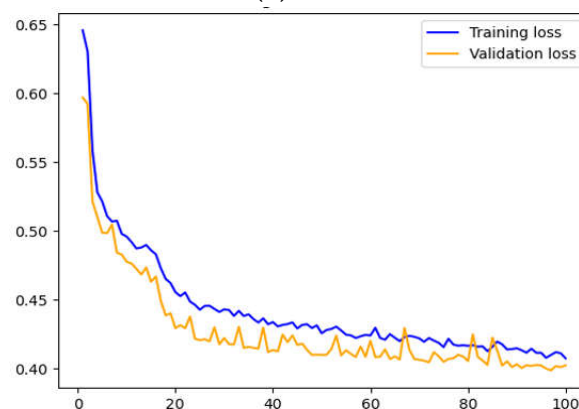
**(a)** GBDT**(b)** RF**(c)** LR**(d)** SVM**(e)** LSTM validation accuracy**(f)** LSTM validation loss**Figure 3** Evaluation of different models used (a), (b), (c), (d) ROC curve of the classifiers trained using the subsets. (e), (f) Training vs. validation loss LSTM.

Table 4 summarizes the study's findings and describes the performance of each model. Caries prediction was performed using binary classification in this experiment,

and the GBDT model demonstrated the best predictive performance with an accuracy of 0.9519, F1-score of 0.9379, prediction of 0.9984, and recall of 0.8844. Furthermore, the suitable subsets for each model differs, and the number of features used for each model has been reduced. The best performance was obtained in the SVM model when training 43 features using the Relief F method, which showed an accuracy of 0.9039, which is more than 9% higher than the accuracy of 0.8128 when using the 43 sets of Full Features. When training models using Feature Selection or Feature Selection combine to Feature Importance techniques, all models performed better than when using the Set of Full Features. This experiment demonstrates how the optimized dataset affects ML performance.

In order to demonstrate the usefulness of the caries prediction decision support system proposed in this paper, it is explained using the Receiver Operating Characteristic (ROC) curve graph. The X-axis of the ROC curve graph expressed in the paper means specificity. In other words, it is a value predicted that there is no caries. Y-axis means sensitivity. This means a value that accurately predicts that a patient has caries. Figure 2 shows the performance of each model using the Full Features dataset. The performance evaluation of the GBDT, RF, and SVM models gives high performance with more than 90% accuracy. LSTM and LR models also provide more than 80% accuracy. Figure 3 depicts the performance of each model applied to the feature selection method dataset. Outstanding evaluations of 0.95 or higher for the GBDT and RF models, 0.90 for the SVM, 0.89 for the LR, and 0.83 for the LSTM models. All models performed quite well overall, according to the evaluation.

5. Discussion

Dental caries is rapidly increasing and recognized as a significant public health problem beyond personal health care. In addition, prevention and early detection of dental caries are essential to reduce the social cost of dental caries that will occur in the future. Currently, caries diagnosis is performed using radiographs or probes. These methods showed remarkable differences in the accuracy and reliability of dental caries diagnosis according to the clinical experience of dental experts. Considering the problems and shortcomings of diagnosis established on the clinician's subjective judgment and experience, research and development of a ML-based DSS are needed. ML based dental caries prediction DSS will help prevent dental caries, manage oral hygiene, improve dietary habits concerning dental caries, and reduce the time and cost required for diagnosis.

The main goals of this study are two first, to select the most informative features to enable the effective prediction of dental caries. The second goal was to propose a model that can be effectively and accurately classified based on caries inducing features. Optimized datasets reduce the performance and cost of ML models. In this study, Chi-Square, Relief F, mRMR, Correlation, and GINI method were used to obtain an optimal data set. Additionally, in this study five classification models (GBDT, RF, LR, SVM, LSTM) were trained using a full feature and subsets. From the experimental results, the proposed data set and classification model are a data set using a combination of mRMR and GINI techniques and a GBDT classification model. This data set consisted of a feature set reduced to 18 out of 43 features, resulting in a 6% improvement in accuracy over training with the full feature set.

Dental caries is rapidly increasing and recognized as a significant public health problem further than just personal health care. In addition, prevention and early detection of dental caries are essential to reduce the socioeconomic cost of dental caries that will occur in the future. Currently, caries diagnosis is performed using radiographs or probes. According to clinical experience of dental specialists, these approaches demonstrated significant variations in the accuracy and reliability of dental caries diagnosis. Considering the problems and shortcomings of diagnosis established on the clinician's subjective judgment and experience, research and development of a ML-based DSS are needed. This ML-based dental caries prediction DSS will help prevent dental caries, manage oral hygiene, improve dietary habits concerning dental caries, and reduce the time and cost required for diagnosis.

Various feature selection techniques were used in the data preprocessing stage of the proposed dental caries prediction DSS to obtain various datasets. By using the optimal dataset to be applied to ML, yielded better performance than the previously published questionnaire-based dental caries prediction DSS.

Karhade, Deepti S. et al. [22] used data from 6,404 people to conduct a study on the caries decision of primary teeth in children. The researcher yielded an AUROC of 0.74, a sensitivity of 0.67, and a positive predictive value of 0.64. Park Y.H et al. [24] obtained AUROC values of LR 0.784, XGBoost 0.785, RF 0.780, and LightGBM 0.780 in a study of children under the age of 5. The results of the proposed study are superior to the results of Karhade, Deepti S. et al. [22] and Park Y.H et al. [24]. For AUROC, our model result, showed RF 0.96, GBDT 0.95, LR 0.89, SVM 0.90, and LSTM 0.83. The DSS proposed in this study used more than 22,287 survey data. In addition, the GBDT model showed an accuracy of 95%, F1-Score of 93%, prediction of 99%, and recall of 88%. However, since the data used in this study were for the Korean child population, there is a problem in that they represent the Korean population better than populations in other countries.

6. Conclusions

Analysis of Dental caries is one of the most frequent fields for modern day re-search. It is because of the severity of dental caries. Facts and figures published according to the 2019 Global Burden of Disease Study; dental caries is reported as the most common oral disease affecting approximately 3.5 billion people. Different approaches have been used to analyze dental caries, but the major concern has been the accuracy of prediction or detection. In this study, we propose a DSS that predicts dental caries centered on the data from the 2018 children's oral health survey. Prediction of dental caries was performed by binary classification, and after training ML by applying various feature selection techniques to the dataset, the optimal model was selected. As a result, the best performance was shown when the GBDT model was trained on the dataset to which the mRMR and GINI techniques were applied. Using 18 features out of a total of 43 features, the results of accuracy 0.9519, precision 0.9984, F1-score 0.9380, recall 0.8844, and AUC-ROC 0.95 were obtained. The proposed model has demonstrated excellent performance and can be applied as a diagnostic aid to identify patients with dental caries. In addition, effective suggestions can be presented for caries prevention, dietary planning, and treatment plan design. This can significantly reduce the social costs caused by dental caries and the patient's diagnosis time and cost.

Future research could also be conducted with a dataset that incorporates information gleaned from samples of the populations of Korea and other nations. Additionally, we intend to employ permutation significance, which can be applied to another ML models, to compare and analyze the experimental outcomes of the model.

Author Contributions: Conceptualization, I.-A.K.; methodology, I.-A.K., J.-D.K.; software, I.-A.K.; validation, I.-A.K., J.-D.K. ; formal analysis, I.-A.K.; investigation, I.-A.K.; resources, I.-A.K.; data curation, J.-D.K.; writing—original draft preparation, I.-A.K. writing—review and editing : J.-D.K. visualization, supervision, J.-D.K.; project administration, J.-D.K.; funding acquisition, J.-D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1F1A1058394).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Institute for Health Metrics and Evaluation (IHME) Explore Results from the 2019 Global Burden of Disease (GBD) Study Available online: <https://vizhub.healthdata.org/gbd-results/> (accessed on 14 December 2022).
2. Health Insurance Review and Assessment Service HIRA 2021-06-08 Available online: <https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020041000100&brdScnBltno=4&brdBltno=10368&pageIndex=1> (accessed on 14 December 2022).
3. Rimi, I.F.; Arif, Md.A.I.; Akter, S.; Rahman, Md.R.; Islam, A.H.M.S.; Habib, Md.T. Machine Learning Techniques for Dental Disease Prediction. *Iran Journal of Computer Science* **2022**, *5*, 187–195, doi:10.1007/s42044-022-00101-0.
4. Zhang, X.; Liang, Y.; Li, W.; Liu, C.; Gu, D.; Sun, W.; Miao, L. Development and Evaluation of Deep Learning for Screening Dental Caries from Oral Photographs. *Oral Diseases* **2022**, *28*, 173–181, doi:10.1111/odi.13735.
5. Lee, S.; Oh, S.; Jo, J.; Kang, S.; Shin, Y.; Park, J. Deep Learning for Early Dental Caries Detection in Bitewing Radiographs. *Scientific Reports* **2021**, *11*, 16807, doi:10.1038/s41598-021-96368-7.
6. Estai, M.; Tennant, M.; Gebauer, D.; Brostek, A.; Vignarajan, J.; Mehdizadeh, M.; Saha, S. Evaluation of a Deep Learning System for Automatic Detection of Proximal Surface Dental Caries on Bitewing Radiographs. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology* **2022**, *134*, 262–270, doi:10.1016/j.oooo.2022.03.008.
7. Megalan Leo, L.; Kalpalatha Reddy, T.; Dental Caries Classification System Using Deep Learning Based Convolutional Neural Network. *Journal of Computational and Theoretical Nanoscience*, **2020**, *17*, 4660-4665(6), DOI: <https://doi.org/10.1166/jctn.2020.9295>
8. Lee, J.-H.; Kim, D.-H.; Jeong, S.-N.; Choi, S.-H. Detection and Diagnosis of Dental Caries Using a Deep Learning-Based Convolutional Neural Network Algorithm. *Journal of Dentistry* **2018**, *77*, 106–111, doi:10.1016/j.jdent.2018.07.015
9. Mohammad-Rahimi, H.; Motamedian, S.R.; Rohban, M.H.; Krois, J.; Uribe, S.E.; Mahmoudinia, E.; Rokhshad, R.; Nadimi, M.; Schwendicke, F. Deep Learning for Caries Detection: A Systematic Review. *Journal of Dentistry* **2022**, *122*, 104115, doi:10.1016/j.jdent.2022.104115.
10. Ainas A. ALbahbah; Hazem M. El-Bakry; Sameh Abd-Elgahany Deep Learning for Caries Detection and Classification. *International Journal of Computer Engineering and Information Technology* **2016**, *8*, 163–170.
11. Lian, L.; Zhu, T.; Zhu, F.; Zhu, H. Deep Learning for Caries Detection and Classification. *Diagnostics* **2021**, *11*, 1672, doi:10.3390/diagnostics11091672.
12. Cantu, A.G.; Gehrung, S.; Krois, J.; Chaurasia, A.; Rossi, J.G.; Gaudin, R.; Elhennawy, K.; Schwendicke, F. Detecting Caries Lesions of Different Radiographic Extension on Bitewings Using Deep Learning. *Journal of Dentistry* **2020**, *100*, 103425, doi:10.1016/j.jdent.2020.103425.
13. Srivastava, M.M.; Kumar, P.; Pradhan, L.; Varadarajan, S. Detection of Tooth Caries in Bitewing Radiographs Using Deep Learning 2017.
14. Ramzi Ben Ali; Ridha Ejbali; Mourad Zaied Detection and Classification of Dental Caries in X-Ray Images Using Deep Neural Networks. *ICSEA 2016 : The Eleventh International Conference on Software Engineering Advances* 2016.
15. Sornam, M.; Prabhakaran, M. A New Linear Adaptive Swarm Intelligence Approach Using Back Propagation Neural Network for Dental Caries Classification. In Proceedings of the 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI); IEEE: Chennai, September 2017; pp. 2698–2703.
16. Vinayahalingam, S; Kempers, S; Limon, L; Deibel, D; Maal, T; Hanisch, M.; Bergé, S.; Xi, T Classification of Caries in Third Molars on Panoramic Radiographs Using Deep Learning. *Sci Rep* **2021**, doi:doi:10.1038/s41598-021-92121-2.
17. Singh, P.; Sehgal, P. Automated Caries Detection Based on Radon Transformation and DCT. In Proceedings of the 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT); IEEE: Delhi, July 2017; pp. 1–6.
18. Berdouses, E.D.; Koutsouri, G.D.; Tripoliti, E.E.; Matsopoulos, G.K.; Oulis, C.J.; Fotiadis, D.I. A Computer-Aided Automated Methodology for the Detection and Classification of Occlusal Caries from Photographic Color Images. *Computers in Biology and Medicine* **2015**, *62*, 119–135, doi:10.1016/j.combiomed.2015.04.016.
19. Liu, L.; Wu, W.; Zhang, S.; Zhang, K.; Li, J.; Liu, Y.; Yin, Z. Dental Caries Prediction Based on a Survey of the Oral Health Epidemiology among the Geriatric Residents of Liaoning, China. *BioMed Research International* **2020**, *2020*, 1–12, doi:10.1155/2020/5348730.
20. Wang, Y.; Hays, R.D.; Marcus, M.; Maida, C.A.; Shen, J.; Xiong, D.; Coulter, I.D.; Lee, S.Y.; Spolsky, V.W.; Crall, J.J.; et al. Developing Children's Oral Health Assessment Toolkits Using Machine Learning Algorithm. *JDR Clinical & Translational Research* **2020**, *5*, 233–243, doi:10.1177/2380084419885612.
21. Hung, M.; Voss, M.W.; Rosales, M.N.; Li, W.; Su, W.; Xu, J.; Bounsanga, J.; Ruiz-Negrón, B.; Lauren, E.; Licari, F.W. Application of Machine Learning for Diagnostic Prediction of Root Caries. *Gerodontology* **2019**, *36*, 395–404, doi:10.1111/ger.12432.
22. Karhade, D.S.; Roach, J.; Shrestha, P.; Simancas-Pallares, M.A.; Ginnis, J.; Burk, Z.J.; Ribeiro, A.A.; Cho, H.; Wu, D.; Divaris, K. An Automated Machine Learning Classifier for Early Childhood Caries. **2021**.
23. Ramos-Gomez, F.; Marcus, M.; Maida, C.A.; Wang, Y.; Kinsler, J.J.; Xiong, D.; Lee, S.Y.; Hays, R.D.; Shen, J.; Crall, J.J.; et al. Using a Machine Learning Algorithm to Predict the Likelihood of Presence of Dental Caries among Children Aged 2 to 7. *Dentistry Journal* **2021**, *9*, 141, doi:10.3390/dj9120141.
24. Park, Y.-H.; Kim, S.-H.; Choi, Y.-Y. Prediction Models of Early Childhood Caries Based on Machine Learning Algorithms. *IJERPH* **2021**, *18*, 8613, doi:10.3390/ijerph18168613.
25. Singh, D.; Singh, B. Investigating the Impact of Data Normalization on Classification Performance. *Applied Soft Computing* **2020**, *97*, 105524, doi:10.1016/j.asoc.2019.105524.

26. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *jair* **2002**, *16*, 321–357, doi:10.1613/jair.953.
27. Agrawal, P.; Abutarboush, H.F.; Ganesh, T.; Mohamed, A.W. Metaheuristic Algorithms on Feature Selection: A Survey of One Decade of Research (2009-2019). *IEEE Access* **2021**, *9*, 26766–26791, doi:10.1109/ACCESS.2021.3056407.
28. Spencer, R.; Thabtah, F.; Abdelhamid, N.; Thompson, M. Exploring Feature Selection and Classification Methods for Predicting Heart Disease. *DIGITAL HEALTH* **2020**, *6*, 2055207620914777, doi:10.1177/2055207620914777.
29. scikit-learn developers Sklearn Feature Selection Chi Available online: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.chi2.html#sklearn.feature_selection.chi2 (accessed on 1 December 2022).
30. Zhang, B.; Li, Y.; Chai, Z. A Novel Random Multi-Subspace Based ReliefF for Feature Selection. *Knowledge- Based Systems* **2022**, *252*, 109400, doi:10.1016/j.knosys.2022.109400.
31. Ghosh, P.; Azam, S.; Jonkman, M.; Karim, A.; Shamrat, F.M.J.M.; Ignatious, E.; Shultana, S.; Beeravolu, A.R.; De Boer, F. Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques. *IEEE Access* **2021**, *9*, 19304–19326, doi:10.1109/ACCESS.2021.3053759.
32. Zhou, H.; Zhang, J.; Zhou, Y.; Guo, X.; Ma, Y. A Feature Selection Algorithm of Decision Tree Based on Feature Weight. *Expert Systems with Applications* **2021**, *164*, 113842, doi:10.1016/j.eswa.2020.113842.
33. Hu, Q.; Si, X.-S.; Qin, A.-S.; Lv, Y.-R.; Zhang, Q.-H. Machinery Fault Diagnosis Scheme Using Redefined Dimensionless Indicators and MRMR Feature Selection. *IEEE Access* **2020**, *8*, 40313–40326, doi:10.1109/ACCESS.2020.2976832.
34. Bugata, P.; Drotar, P. On Some Aspects of Minimum Redundancy Maximum Relevance Feature Selection. *Sci. China Inf. Sci.* **2020**, *63*, 112103, doi:10.1007/s11432-019-2633-y.
35. Wang, Y.; Li, X.; Ruiz, R. Feature Selection With Maximal Relevance and Minimal Supervised Redundancy. *IEEE Trans Cybern* **2022**, *PP*, doi:10.1109/tcyb.2021.3139898.
36. Saidi, R.; Bouaguel, W.; Essoussi, N. Hybrid Feature Selection Method Based on the Genetic Algorithm and Pearson Correlation Coefficient. In *Machine Learning Paradigms: Theory and Application*; Hassanien, A.E., Ed.; Studies in Computational Intelligence; Springer International Publishing: Cham, 2019; Vol. 801, pp. 3–24 ISBN 978-3-030-02356-0.
37. Ghasemi, F.; Neysiani, B.S.; Nematbakhsh, N. Feature Selection in Pre-Diagnosis Heart Coronary Artery Disease Detection: A Heuristic Approach for Feature Selection Based on Information Gain Ratio and Gini Index. In Proceedings of the 2020 6th International Conference on Web Research (ICWR); 2020; pp. 27–32.
38. Sung, S.-H.; Kim, S.; Park, B.-K.; Kang, D.-Y.; Sul, S.; Jeong, J.; Kim, S.-P. A Study on Facial Expression Change Detection Using Machine Learning Methods with Feature Selection Technique. *Mathematics* **2021**, *9*, 2062, doi:10.3390/math9172062.
39. Zhang, W.; Yu, J.; Zhao, A.; Zhou, X. Predictive Model of Cooling Load for Ice Storage Air-Conditioning System by Using GBDT. *Energy Reports* **2021**, *7*, 1588–1597, doi:10.1016/j.egy.2021.03.017.
40. Liang, W.; Luo, S.; Zhao, G.; Wu, H. Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms. *Mathematics* **2020**, *8*, 765, doi:10.3390/math8050765.
41. Kumar, M.S.; Soundarya, V.; Kavitha, S.; Keerthika, E.S.; Aswini, E. Credit Card Fraud Detection Using Random Forest Algorithm. **2019**.
42. Ao, Y.; Li, H.; Zhu, L.; Ali, S.; Yang, Z. The Linear Random Forest Algorithm and Its Advantages in Machine Learning Assisted Logging Regression Modeling. *Journal of Petroleum Science and Engineering* **2019**, *174*, 776–789, doi:10.1016/j.petrol.2018.11.067.
43. Otchere, D.A.; Arbi Ganat, T.O.; Gholami, R.; Ridha, S. Application of Supervised Machine Learning Paradigms in the Prediction of Petroleum Reservoir Properties: Comparative Analysis of ANN and SVM Models. *Journal of Petroleum Science and Engineering* **2021**, *200*, 108182, doi:10.1016/j.petrol.2020.108182.
44. Iwendi, C.; Mahboob, K.; Khalid, Z.; Javed, A.R.; Rizwan, M.; Ghosh, U. Classification of COVID-19 Individuals Using Adaptive Neuro-Fuzzy Inference System. *Multimedia Systems* **2022**, *28*, 1223–1237, doi:10.1007/s00530-021-00774-w.
45. Xiong, C.; Callan, J. EsdRank: Connecting Query and Documents through External Semi-Structured Data. In Proceedings of the Proceedings of the 24th ACM International on Conference on Information and Knowledge Management; ACM: Melbourne Australia, October 17 2015; pp. 951–960.
46. Norman R. Draper; Harry Smith *Applied Regression Analysis, Third Edition Book Series:Wiley Series in Probability and Statistics*; Wiley Online Library, 1998; ISBN 978-1-118-62559-0.
47. Michael P. LaValley *Statistical Primer for Cardiovascular Research*, 2008.
48. Sepp Hochreiter; Jurgen Schmidhuber Long Short-Term Memory. *Neural Computation* (Volume: 9, Issue: 8, 15 November 1997)Communicated by RonaldWilliams **1997**, 1735–1780, doi:10.1162/neco.1997.9.8.1735.
49. Hamayel, M.J.; Owda, A.Y. A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and Bi-LSTM Machine Learning Algorithms. *AI* **2021**, *2*, 477–496, doi:10.3390/ai2040030.
50. SVC' Object Has No Attribute "feature_importances_" Available online: <https://stackoverflow.com/questions/59681421/svc-object-has-no-attribute-feature-importances> (accessed on 1 May 2023).