applied sciences

Article

DCP: Prediction of Dental Caries Using Machine Learning in Personalized Medicine

In-Ae Kang [1,†], Soualihou Ngnamsie Njimbouom [1,†], Kyung-Oh Lee [1] and Jeong-Dong Kim [1,2,*]

[1] Department of Computer and Electronics Convergence Engineering, Sun Moon University, Asan-si 31460, Korea; inae5004@sunmoon.ac.kr (I.-A.K.); salehrico@sunmoon.ac.kr (S.N.N.); leeko@sunmoon.ac.kr (K.-O.L.)
[2] Genome-Based BioIT Convergence Institute, Sun Moon University, Asan 31460, Korea
* Correspondence: kjd4u@sunmoon.ac.kr
† These authors contributed equally to this work.

Abstract: Dental caries is an infectious disease that deteriorates the tooth structure, with tooth cavities as the most common result. Classified as one of the most prevalent oral health issues, research on dental caries has been carried out for early detection due to pain and cost of treatment. Medical research in oral healthcare has shown limitations such as considerable funds and time required; therefore, artificial intelligence has been used in recent years to develop models that can predict the risk of dental caries. The data used in our study were collected from a children's oral health survey conducted in 2018 by the Korean Center for Disease Control and Prevention. Several Machine Learning algorithms were applied to this data, and their performances were evaluated using accuracy, F1-score, precision, and recall. Random forest has achieved the highest performance compared to other machine learnings methods, with an accuracy of 92%, F1-score of 90%, precision of 94%, and recall of 87%. The results of the proposed paper show that ML is highly recommended for dental professionals in assisting them in decision making for the early detection and treatment of dental caries.

Keywords: machine learning; dental caries; random forest; logistic regression; gradient boot decision tree; support vector machine; artificial neural network; convolutional neural network; long short time memory

Citation: Kang, I.-A.; Ngnamsie Njimbouom, S.; Lee, K.-O.; Kim, J.-D. DCP: Prediction of Dental Caries Using Machine Learning in Personalized Medicine. *Appl. Sci.* **2022**, *12*, 3043. https://doi.org/10.3390/app12063043

Academic Editor: Gabriele Cervino

Received: 22 January 2022
Accepted: 14 March 2022
Published: 16 March 2022

1. Introduction

Dental caries is a common chronic infectious oral disease that affects adolescents and adults worldwide. There are approximately 1.8 billion caries-related disease cases recorded annually, which imply that deterioration of dental health will ultimately lead to more records of tooth loss cases [1–3].

The prevalence of dental caries rapidly increases from infancy to adolescence. It exhibits an epidemiologic characteristic that is gently maintained after adolescence, and then periodontal disease increases from adolescence [4]. Hence, it is necessary for dental caries to be treated at the infancy stage because untreated childhood caries can lead to caries of permanent teeth and oral diseases. Therefore, as it affects the overall quality of life, it is recognized as a public health problem rather than an individual problem [5–7]. Early detection of dental caries can be resolved with preventive treatment and restorative treatment. However, long-term treatment such as extensive tooth cleaning, root canal treatment, and prosthetic treatment after tooth extraction is required if it affects dentin or pulp due to long-term negligence [8]. It is most desirable for everyone to pay attention to one's oral health and to visit the dentist for regular diagnosis and treatment. However, in many cases, prevention or examination is neglected before noticing symptoms due to economic conditions or awareness of the importance of oral health [9].

Globally used as the main population-based measure of caries-related diseases worldwide, the decayed–missing–filled teeth (DMFT) index is used for the summation of the number of permanent teeth experienced with caries (decayed, missing, or filled). Oral health is known to be affected by an individual's improper oral health management habits, eating habits, use of dental care, and economic factors [8]. Hence, identifying DMFT and its related factors and predicting permanent teeth with dental caries can be an essential basis for establishing an individual oral prevention strategy. In Korea, in accordance with Article 9 of the Oral Health Act, children's oral health surveys are being conducted. This survey is used as basic data for effective dental caries management for children's oral health by examining the public's oral health status, related forms, and use of dental care products. The survey conducted in 2018 surveyed 27,568 people, and it is recognized as big data in the field of oral epidemiology. The survey results show that DMFT tends to increase significantly as age increases [10].

Machine learning (ML) has become an essential instrument to comprehend and analyze data as massive as the survey mentioned above and is being used in various ways in the medical area. ML is a technique used to predict future outcomes by learning existing patterns between elements of targeted data [11]. The history of ML started in the 1950s and progressed until the 1980s and 1990s and then stood still. However, in the mid-2000s, the vast amount of big data accumulated on the Internet and hardware development to handle it amounted to a colossal leap forward [12]. Recently, remarkable achievements have been made in image recognition, speech recognition, and translation using ML. In the field of dentistry in Korea, deep learning (DL) research using panoramic, dental computed tomography, X-ray, and intraoral photos are being conducted [13,14]. However, compared to the importance and usefulness of the technology, it is not being used smoothly in dentistry, and attempts to use it in more diverse fields are needed.

Unlike the conventional method, ML can predict carious teeth in the population with surveys or basic information before performing a detailed diagnosis by an expert. Human resources, time, and cost required for an oral health examination will considerably be reduced. ML can also help classify the high-risk individuals to receive an accurate diagnosis and necessary treatment from a specialist. In addition, if the predictive model identifies contributing factors that significantly impact DMFT, caries prevention can be accomplished by controlling those factors.

This paper proposes the prediction of dental caries model using machine learning in personalized medicine. The proposed model, called DCP, uses methods such as random forest (RF), support vector machine (SVM), gradient-boosted decision trees (GBDT), and logistic regression (LR). A grid search algorithm and cross-validation were applied to increase the prediction probability of dental caries. Moreover, the DCP model uses DL models (such as artificial neural network (ANN), convolutional neural network (CNN), long and short-term memory (LSTM)). Layers settings and max-pooling were applied to reduce overfitting and to provide good prediction results. Hence, the DCP model will replace the time-consuming process of detecting dental caries, as dentists and patients can use the results to evaluate the future crescendo of potentially severe dental infections.

The remainder of this paper is organized as follows. Section 2 presents the related works. Section 3 describes the DCP model and the characteristics of the ML models for the experiments. Experimentation that incorporates the dataset and hyperparameter tuning are depicted in Section 4. The experimental results and discussion are presented in Section 5. Finally, Section 6 provides the conclusion and future work.

2. Related Work

Artificial intelligence (AI) technology can be applied in a variety of ways in the medical field, from early detection of disease, patient condition monitoring using software, clinical decision making using big data, and wearable devices and sensors that use IoT technology to monitor a patient's health status in real time. In addition, ML can improve diagnosis and treatment effectiveness by allowing medical experts to analyze real-time data to identify changes in patients.

The field of medical image analysis and diagnosis has the highest annual growth rate in the AI-based healthcare industry, and a large amount of medical image data available facilitate the research venture in this field. Based on this paradigm, many AI methods have been introduced in the medical area to alleviate the time-consuming and cost-related aspect of performing some diagnoses. VUNO, a representative case, helps perform bone age assessment using DL by integrating and analyzing medical image information such as X-ray and CT [15]. Lunit improves the early diagnosis rate and reduces the false diagnosis rate by developing an AI-based solution for chest X-rays and mammography images, using DL technology [16]. Consent agreement related to medical data has restricted some research's vulgarization and is currently heavily performed only by medical doctors.

With the entry into an aging society, oral health is one of the essential factors determining the quality of life. Periodontitis is one of the most common diseases afflicted by humankind and is the sixth most prevalent disease globally. Loss of alveolar bone is one of the main symptoms of periodontitis and causes tooth loss, edentulous jaw, and masticatory dysfunction [17]. A classification system for periodontitis was needed to provide a standard for research on the etiology and treatment of periodontitis [18]. Since then, the classification system for periodontitis has been continuously revised according to the newly identified scientific and clinical evidence [19]. In general, clinical attachment loss (CAL) evaluates periodontal health using a periodontal probe. This method has limitations in terms of reliability; thus, a computer-aided diagnosis (CAD) system can help clinicians make decisions by extracting important features from medical images taken in various environments [20]. However, the existing CAD approach has limitations in that feature extraction is not easy and takes much time due to the diversity of disease patterns [21].

Nevertheless, current methods based on DL, a subset of ML, have the advantage of automatically extracting critical features from images through learning [22]. Convolutional neural network (CNN) is the most used method to segment, classify, and detect organs or related diseases in medical images [23–25]. In the field of oral and maxillofacial surgery, the CAD approach integrated with CNN has been used to detect landmarks in cephalograms [26], tooth detection and classification [27,28], caries diagnosis [29], and maxillary sinusitis detection [30]. A dentist can provide a more accurate and rapid diagnosis result by adding a second opinion to the results obtained through this method. Recently, studies on RBL detection using CNN in dental panoramic images have also been conducted [31,32].

Research on dental caries classification by image interpretation mainly controls the contrast of the X-ray image. The authors of [33] present a novel approach that detects the presence of dental caries-hybridized negative transformation and statistical analysis for the dental image containing dental caries along with cysts. Radon transform and discrete cosine transform were applied to cavities, and some studies diagnose by classification using various classification techniques such as decision tree, random forest, and naïve Bayes. This technique improved the diagnostic accuracy of caries by adjusting the image contrast, but the limitation of the study was revealed due to the generation of image noise. The histogram equalization technique was employed to overcome image noise limitations [34,35]. The authors of [36] proposed an image enhancement technique using watershed and introduced a kernel-modified SVM to facilitate the initial caries diagnosis at clinics. The authors of [37] proposed a review study for the early detection and diagnosis of dental caries on periapical radiographs using DL CNN algorithms. The authors of [38] presented work on classifying the presence or absence of root caries as a dichotomous variable based on public data from the National Center for Health Statistics in the United States. The top variables influencing dental root caries were extracted using a statistical technique and ranked based on their F1-Score. Various ML techniques were applied, and SVM displayed the highest performance. A case study based on open data in Korea, classifying the presence or absence of dental caries using the decayed–missing–filled teeth (DMFT), was proposed by [39]. This study constructed a multiple linear regression model targeting approximately 20,000 children using the influence of individual oral health management habits, eating habits, dental care products used, and socioeconomic factors as variables in the Oral Health Survey. Using

ML, they found feature importance in the order of gender, region, oral health perception status, number of dental treatments, and daily snack intake for one year.

3. DCP Model

This study focused on predicting dental caries by proposing a prediction model: the DCP, which consists of data collection, data prepossessing, and prediction model. The proposed DCP model, consisting of ML and DL models, uses features from the preprocessed dataset obtained from the survey to predict the presence or absence of dental caries from children. Using various ML methods, our model takes the relationships between parameters in a dataset and predicts dental caries as meta-knowledge. Our proposed model is schematically illustrated in Figure 1.

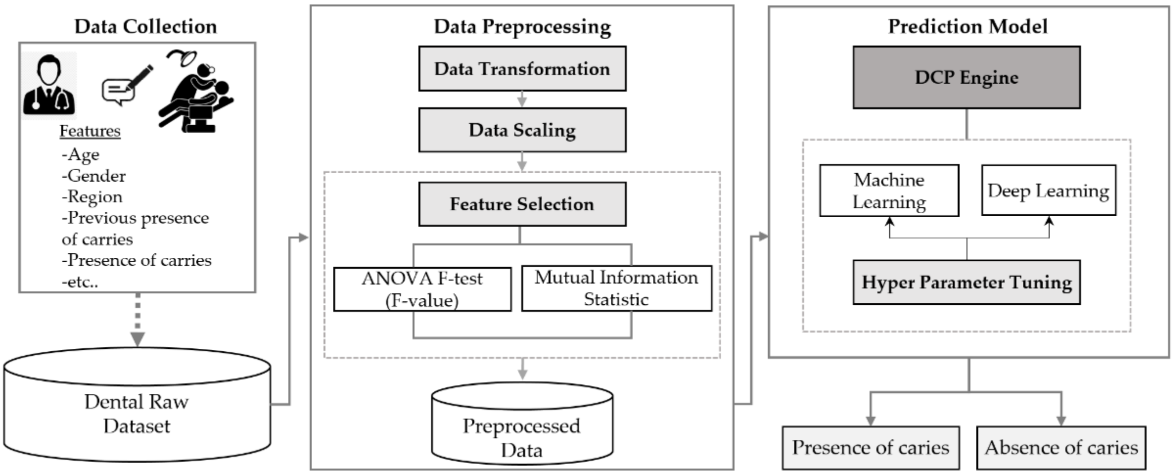


Figure 1. Proposed model for dental caries prediction.

3.1. Data Collection

A labeled dataset made of 22,287 samples consisting of features such as age, region, gender, etc., were made available by the Korean Center for Disease Control and Prevention under the Ministry of Health and Welfare after a children's oral health survey in 2018. This dataset is the outcome of consultation operations performed by medical doctors. The total interviewed patients were grouped into two classes, 20,593 of them were not suffering from dental caries, and 1695 were suffering from it. The patients; personal information was kept secret and was not recorded.

3.2. Data Preprocessing

We cleaned our dataset by removing useless attributes and samples with several empty features. After that, we performed data scaling to normalize the range of the independent variables and standardize the functionality range of the input dataset, therefore having a standard deviation of 1 and a mean value of 0. Then, we performed feature selection to identify the subset of features that are more important to the target variable. analysis of variance (ANOVA) F test, feature selection using random forest, and mutual information statistic methods were used, and features independent of the target variable were removed [40]. The ANOVA statistical feature selection techniques selected features based on analysis of the variance using the F value methods; F value here was used to determine the level variance between the sample means relative to the variation within the data sample. This feature selection process can be seen in one of the source code files on our repository. The dataset consisted of two classes: presence of caries (1695 samples) and absence of caries (20,593 samples). Handling imbalanced classes distribution in the dataset is a challenge to overcome to improve the prediction accuracy of the minority class. Given that classifiers learning most of the patterns from the majority class tend to be less sensitive to the minority class, unreliability assumption may occur in their prediction performances [41]. During our experiments, to solve the unbalanced data problem, this paper used synthetic minority oversampling technique (SMOTE), which generates new data instances by taking samples of the feature space of each target class combined with the feature of its nearest neighbors [42]. The newly generated instances are different from the existing samples present in the minority class. In our experiments, classes distribution was of the ratio of 92.40% (20,593) vs. 7.6% (1695) for the absence of caries vs. presence of caries, respectively. The data samples were oversampled to a ratio of 58.82% (20,593) for the absence of caries vs. 41.18% (14,415) for the presence of caries after the use of the SMOTE technique to prevent the unbalancing faced.

3.3. Prediction Models

The prediction model of our approach, called DCP, was designed to predict the presence or absence of caries for patients in a specified age range, considering the specific oral health features recorded. The basic assumption is having evenly distributed and independent data. The studied features were pooled together and used as input data for classification models to predict dental caries.

Machine learning is a computer algorithm built to automatically improve its learning through experience from example data. Supervised learning (task-driven), unsupervised learning (data-driven), and reinforcement learning (learning from error) are the most known types of machine learning. We have experimented using traditional ML methods such as RF, GBDT, LR, SVM and DL methods such as ANN, CNN, LSTM. These ML algorithms learn patterns in the input data and generate rules to mine on our dataset to predict dental caries. We then compared the predicted outputs of these algorithms with their actual values to evaluate the effectiveness of different methods in predicting caries. The test dataset was used to calculate precision, recall, F1 score, and accuracy. Predictive accuracy is most relevant for clinical use in dental healthcare. The characteristic of each ML method and DL method used in this study will be described in more detail.

3.3.1. Random Forest (RF)

RF is an advanced technique of decision tree analysis, and it is an ML method that analyzes data by expanding it into multiple trees instead of one tree. Because variable selection is free, overfitting to the dataset can be prevented, and a model with high predictiveness is obtained. Multiple pieces of data of the same size as the original data are composed by random sampling that allows for duplication through bagging, and each decision tree is constructed based on this. The final observation is predicted by combining the results extracted from each decision tree with the mode or weighted average. About 37% of the total data is used as out of bag (OOB) data in the RF. Through OOB error, it is possible to calculate the performance evaluation of the built model and the importance of input variables affecting the prediction performance. RF has an advantage over other data mining techniques such as ANN and can be used to improve the understanding of research models and predictive performance [43].

3.3.2. Gradient Boosting Decision Tree (GBDT)

XGBoost is a tree-based ensemble ML algorithm. It uses a technique called gradient boosting. Boosting is a technique for predicting accuracy by grouping weak learners into a set. Gradient descent is applied, weighing the learning errors of weak prediction models and sequentially reflecting them in the next learning model to create a strong prediction model that minimizes losses. XGBoost is an algorithm that dramatically reduces the learning time such that parallel learning is implemented using gradient boost, a representative boosting algorithm, and both classification and regression can be applied. XGBoost has strong durability, has a regulation function to prevent overfitting, and provides early stopping [44].

3.3.3. Support Vector Machine (SVM)

SVM is a powerful supervised learning model used for linear and non-linear classification, regression, and outlier detection. The goal is to establish a hyperplane in a high-dimensional or infinite-dimensional space that maximizes the spacing between two kinds of data samples. The mathematical form of the hyperplane is defined in Equation (1).

$$g(x) = w{\cdot}x + b = 0 \tag{1}$$

where the separation of the hyperplane is represented by the normal vector w. b stands for the offset between the origin plane and the hyperplane. The objective function is transformed into a dual optimization problem by introducing the Lagrangian coefficient in Equation (2) [45]:

$$\begin{gathered} \underset{w,b}{min}\ \underset{\alpha}{max} L(w,b,\alpha) = \tfrac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(w{\cdot}x_i + b) - 1] \\ (0 \le \alpha_i \le c.i = 1, 2, \ldots n) \\ s.t.\ loss\ function\ :\ \sum_{i=1}^{n} [y_i(w{\cdot}x_i + b) - 1] \end{gathered} \tag{2}$$

where $L(w,b,a)$ denotes the Lagrangian function, α_i indicates the Lagragian coefficient, and c is a penalty factor which is the upper bound of α_i. When x_i belongs to the first class, the value of y_i = 1, while when x_i belongs to the second class, y_i = -1. The error value also referred to as loss value of the objective function is affected by the penalty factor c, whereby the more significant the value of c, the greater the loss. The SVM may easily suffer from overfitting if the error is significant. Otherwise, the SVM may have an underfitting problem when c is too small [46–48].

3.3.4. Logistic Regression (LR)

LR analysis is a type of categorical data modeling in which the dependent variable consists of a nominal scale and a sequence scale and is used when the value of the dependent variable is 0 and 1, which are binary coefficients, of which the odds ratio is verified. The probability that the dependent variable has a value of 1 is called odds [49].

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon \tag{3}$$

- Z (dependent variable): the dependent variable, the value to be predicted (here, the presence or absence of dental caries);
- β (coefficients): a coefficient calculated through logistic regression, a value showing the relationship between the dependent variable and the independent variable.
- X (explanatory variables): independent or explanatory variables, influence factor variables considered to predict the dependent variable (here, the set of features used to predict the dental caries).

By applying the logistic regression equation presented in Equation (3) to the logistic function in Equation (4), the probability of the occurrence of a future event can be calculated.

$$g(z) = \frac{1}{(1 + e^{-z})} \tag{4}$$

3.3.5. Artificial Neural Networks (ANN)

ANN is a mathematical model created for prediction or classification by mimicking the brain nervous system in ML and cognitive science. The basic structure of ANNs is that natural neurons receive signals through synapses. In the case of ANNs, inputs correspond to synapses and are weighted according to the strength of individual signals, and activation functions calculate the outputs of ANN. In other words, ANN refers to the overall model that has the problem-solving ability by changing the bonding strength of synapses through learning. The computation of an ANN begins by introducing an array of numbers x_i into the input layer of the processing node. These signals then travel along with the connections to each node in the adjacent layer and can be suppressed or amplified via connection-specific weights (w_t). Adjacent layer nodes are summaries for incoming signals. Then, the incoming signal is a threshold function [50].

$$f(x) = \frac{1}{1+e^{-x}} \tag{5}$$

In Equation (5), the output $f(x)$ ranges between 0 and 1, and the output of the processing unit is calculated as:

$$Q_j = \frac{1}{1+e^{-\sum x_i w_i}} \tag{6}$$

This output signal Q_j defined in Equation (6), continues through the weighted linkage to the next node layer. The process repeats until the signal reaches the output layer. The output signal can then be interpreted as the ANN's response to a given input stimulus.

3.3.6. Convolution Neural Network (CNN)

CNN is a layer-based neural network that extracts features using convolution. Due to the nature of convolution, only nearby data are grouped and processed. The feature detection layer performs convolution, rectified linear unit (ReLU), and pooling processes, and convolution plays a role in activating features by using a convolution filter on the input data [51]. ReLU keeps the data as positive values such that the network learns quickly, therefore avoiding the vanishing gradient problem. In addition, pooling simplifies the output by performing nonlinear downsampling and reducing the network's number of parameters to learn. The classification layer is a node connected after the CNN passes through the feature detection layer and is generally composed of a fully connected layer (FC), and outputs a K-dimensional vector, where K represents the number of classes to be predicted by the network, here K = 2 (presence and absence of caries). A CNN is a particular neural network for processing data with known grid-like topologies. These data include time-series data, which can be thought of as 1D, and image data are considered a 2D grid of pixels. Networks are called CNN because they use a mathematical operation called convolution.

3.3.7. Long Short-Term Memory (LSTM)

RNN is an ANN that learns the time-dependent relationship between input data by combining the past information stored in the hidden layer with the current input value. However, RNN may have gradient loss problems or explosions for long-length sequences [52]. Out of the proposed variant of RNN, LSTM has become a state-of-the-art model for some problems in recent years. The LSTM unit structure comprises cells composed of a forget gate, an input gate, and an output gate. This LSTM design has overcome the RNN vanishing gradient problem. The forget gate is a form of taking the sigmoid function to the weighted sum of the information of the previous cell (h_{t-1}), and the current input data (x_t), is biased to determine how much the previous cell state is to be forgotten and converts a value between 0 and 1, indicating the degree to which information from the past is remembered.

$$f_t = \sigma(W_f \cdot [h_{t-1},\ x_t] + b_f) \tag{7}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}\ , x_t] + b_i) \tag{8}$$

$$\widetilde{C}_t = tanh(W_C \cdot [h_{t-1},\ x_t] + b_C) \tag{9}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{10}$$

$$o_t = \sigma(W_0 [h_{t-1}\ ,\ x_t] + b_o) \tag{11}$$

$$h_t = o_t * tanh(C_t) \quad (12)$$

Equation (7) means that the sigmoid layer decides the retention and removal of information. If it has a value of 1, it is "kept", and if it has a value of 0, it is "removed".

Equations (8) and (9) are the steps to decide the storage of new information. In Equation (8), the value to be updated is determined through sigmoid. Equation (9) creates a new candidate value $\widetilde{C}_t$ to be added to the cell state.

Equation (10) is the update step, and the past cell state, that is, (C_{t-1}), is updated to a new state (C_t).

Equations (11) and (12) are the processes of deciding the output value to the output and how much to extract the finally obtained cell State value.

4. Experimentation

4.1. Dataset

The Children's Dental Health dataset is Korea's public dental health dataset [4]. This dataset provides oral health information of 22,288 children collected from a survey. The dataset contains 45 features such as gender, previous caries history, calculus information, frequency of oral health care product usage, and more. All participants were divided into two classes. Patients with dental caries were labeled as "1", and participants unaffected by caries labeled as "0". Our dataset had an unbalanced distribution of data: over 20,593 samples without dental caries "labeled 0" and 1695 samples with dental caries "labeled 1", as depicted in Figure 2.

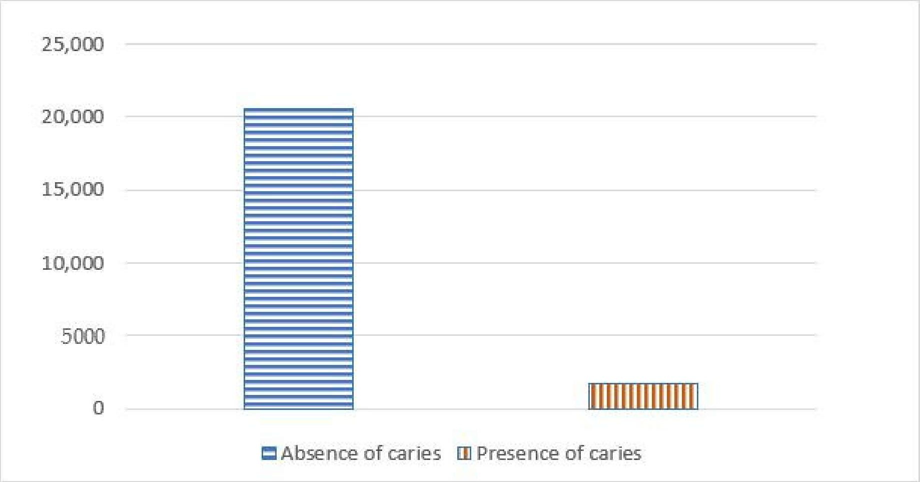


Figure 2. The distribution of data sample.

4.2. Hyperparameter of Different Machine Learning Models

A grid search algorithm was used to tune the hyperparameters of our traditional ML methods. By providing a custom value range, the grid search method thoroughly searches the set of hyperparameters that optimizes the result of our models. The parameters that maximize the correct probability value were obtained by applying grid search with the maximum likelihood estimation concept (MLE) for our dental caries prediction task. The basic equation of MLE can be represented, as in Equation (13):

$$p^* = arg\ max_p P(y|\vec{x}) \quad (13)$$

Here, $P(y|\vec{x})$ represents the entropy (that is the uncertainty), and p^* represents the set of hyperparameters that minimizes the uncertainty and maximize the performance. Table 1 describes the different optimizable parameters used while predicting dental caries using traditional models as well as the deep neural network algorithm. From Table 1, we can see the selected hyperparameters value of each algorithm after the grid search algorithm, which meticulously determined the best values from a range of values as shown in experimentation source code.

Table 1. Optimizable parameters for different models.

Machine Learning	Parameters	Range	Description
RF	n_estimator max_depth min_samples_leaf max_features	160 None 1 auto	Number of decision trees to use in the forest Maximum depth of each tree Number of samples required at the leaf node Number of features to consider for each tree
GBDT	n_estimator max_depth min_samples_leaf max_feature	100 None 2 Auto	Identical to the above
	subsample	0.85	Controls the percentage of observations; sed to create a score for each tree
SVM	kernel	rbf	Kernel type to be used by the algorithm
	gamma	0.0001	Kernel factor in controlling the distance of influence of a training point
	C	1000	Adds a penalty for each misclassified data
	probability	True	The parameter that enables probability estimate
LR	penalty	l_1	Specifies the norm of penalty (none, $l_1 l_2$ or elastic net)
	C	1	Specifies weight importance for training data or complexity penalty
	solver	liblinear	Specifies the algorithm used in the optimization problem
ANN CNN LSTM	learning rate	0.001	Determines the step size at each iteration while minimizing the loss.
	epoch	200 (ANN, CNN) 100 (LSTM)	Specifies the number of passes on the entire training dataset that the ML performs.
	batch size	1000 (ANN) 1380 (CNN) 480 (LSTM)	Indicates the number of training samples used during one epoch.
	patience	30 (ANN) 10 (CNN) 50 (LSTM	Specifies early stop if no progress on the validation set

5. Results and Discussion

5.1. Evaluation Metrics

To evaluate the performance of our prediction models and avoid overfitting, our dataset was split into training (80% of data samples) and test (20% of data samples) sets. The training set was further divided into training (80% of the Training set) and validation (20% training set) sets. This work compared the commonly used evaluation metric (accuracy, precision, recall, and F1-score) of the machine learning experimented on RF, SVM, GBDT, LR, ANN, LSTM, and CNN. The mathematical equations of our evaluation metrics are defined below:

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{16}$$

$$Accuracy = \frac{TP + FP}{(TN + TP + FN + FP)} \tag{17}$$

Precision: Equation (14) is defined as the ability of a classifier to identify the relevant data points. *TP* (true positive) is when a positive observation is also predicted as positive. *FP* (false positive) is a negative observation that is indicated as being positive.

Recall: Equation (15) is the measurement that defines how correctly a model has identified the positive classes in our dataset classes. *FN* (false negative) represents a positive observation that was predicted as being negative.

*F*1-*Score*: Equation (16) is the measurement of the model's *accuracy* on the dataset. The weighted average between the *precision* and *recall* are defined in Equations (14) and (15).

Accuracy: Equation (17) is the measurement that determines a model's efficiency at identifying patterns and relationships between features in the datasets, which is the ratio between the correctly predicted value by the total number of assessments, where *TN* (true negative) is the negative observation correctly predicted as negative.

5.2. Evaluation Results of Our Models

After proper feature selection and training of our ML methods, performance (accuracy, F1-score, precision, and recall) is essential in selecting the appropriate ML for our prediction model. This work performs a binary classification on the presence or absence of dental caries. Table 2 shows the performance results of all the ML methods on the Test set. We observed that the RF outperformed the other applied ML methods and any neural network-based methods. Our proposed approach using RF displayed an accuracy, F1-score, precision, and recall of 0.92, 0.90, 0.94 and 0.87, respectively. Therefore, the presented results of RF can be seen as highly competitive compared to other prediction methods applied.

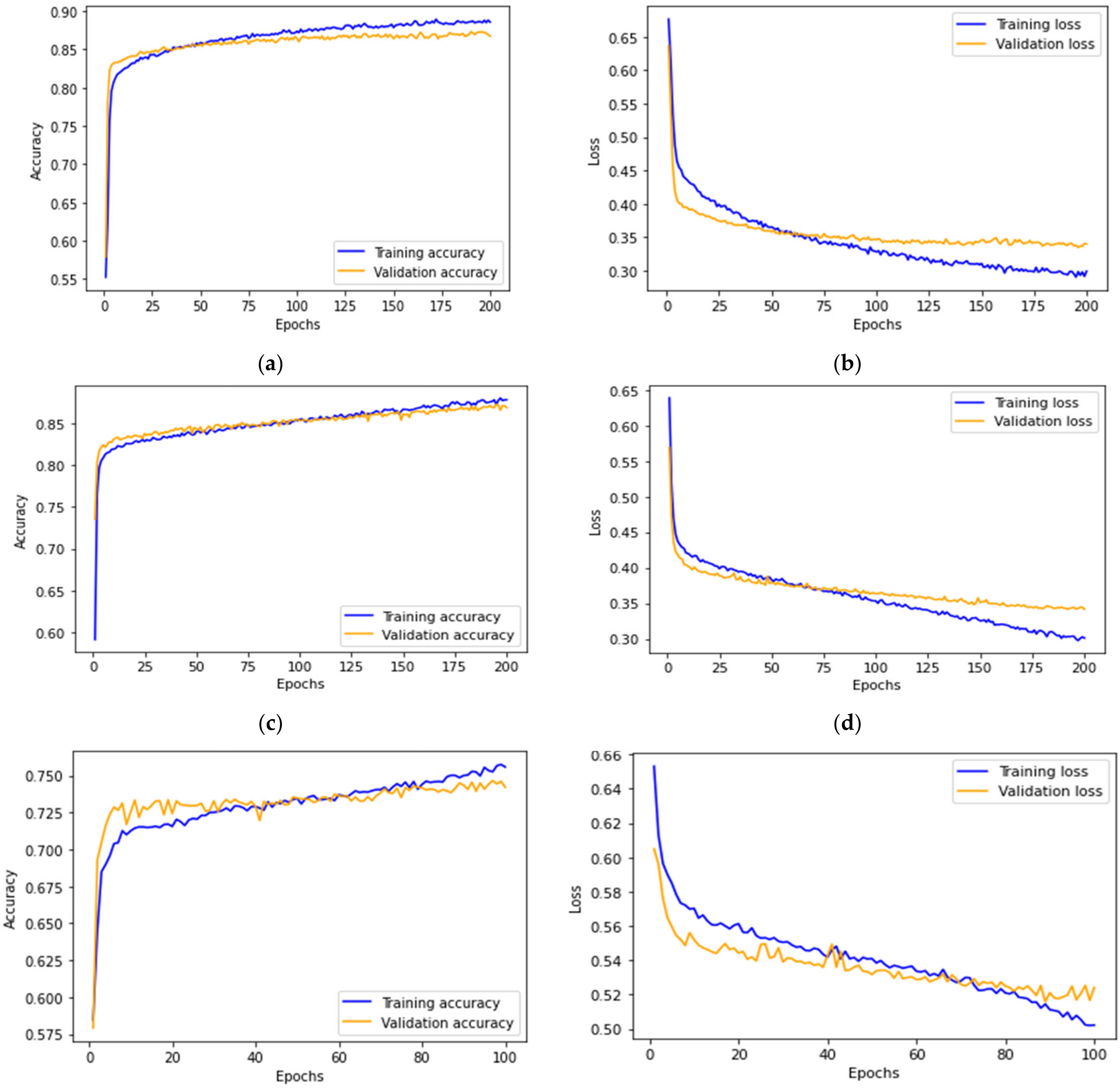
Table 2. Performance of different ML methods used.

Model	Accuracy	F1-Score	Precision	Recall
RF	0.92	0.91	0.94	0.88
ANN	0.88	0.87	0.87	0.87
CNN	0.87	0.87	0.87	0.87
GBDT	0.85	0.81	0.83	0.78
SVM	0.83	0.79	0.82	0.76
LR	0.82	0.78	0.80	0.76
LSTM	0.75	0.74	0.74	0.74

It is essential to mention the performances observed when DL was applied during our proposed paper. Three DL algorithms were analyzed: ANN, CNN, and LSTM, out of which the ANN outperformed the other DL methods. The results of the different DNNs are shown in Figure 3. We performed our experiments with the tuned parameters learning rate of 0.001, epochs of 200 for ANN and CNN, and 100 for LSTM and drop out of 0.20.

Figure 4 presents a graphical representation of accuracy, F1-score, precision, and recall ranking prediction results of all the ML methods used. As a result of analyzing the performance of the ML methods on our dataset, RF showed the highest performance among the ML used, followed by ANN, CNN, GBDT, SVM, LR, and then LSTM. When compared to each other, RF accuracy was better than ANN by 4%, CNN by 5%, GBDT by 7%, SVM by 9%, LR by 10%, and then LSTM by 17%. Similarly, F1-score, precision, and recall of RF also outperformed all other ML used. Regardless of the high-end result achieved by the ANN among the DNN algorithms used on the test set, RF still stands out to be the best algorithm for our proposed paper.

During our experiments, the dataset was divided into training and test sets. The training set was further divided into a training and validation set. *K*-fold cross-validation was used during our grid search hyperparameter tuning step on the validation set to tune the hyperparameters, while models were trained on the training set. In our case, we used *K* = 10, in which during each fold, training and validation were partitioned into a *K* subset for validation and *K*-1 subset set for training. The test set was left aside to evaluate the performances of our different ML methods. These results revealed that our models performed well on unseen data as they did on the training Set.



(**a**)

(**b**)

(**c**)

(**d**)

(**e**)

(**f**)

Figure 3. Accuracy vs. loss performance evaluation of different neural network models used. (**a**) Training vs. validation accuracy ANN. (**b**) Training vs. validation loss ANN. (**c**) Training vs. validation accuracy CNN. (**d**) Training vs. validation loss CNN. (**e**) Training vs. validation accuracy LSTM. (**f**) Training vs. validation loss LSTM.

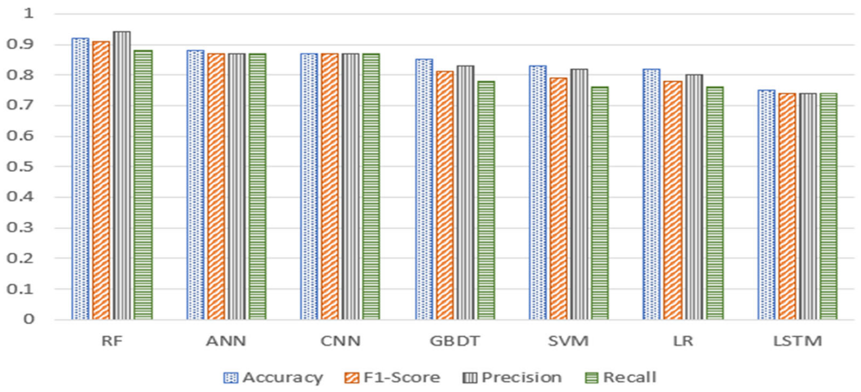


Figure 4. Graphical representation of machine learning model metric.

5.3. Discussion

The usage of ML in oral healthcare has shown its success in increasing the oral health condition of individuals by providing dental professionals with a tool, giving them an ability to make early decisions to prevent dental caries to individuals, hence improving their overall lifestyle condition.

This paper has explored and investigated several ML methods (SVM, LR, GBDT, RF, and DL) on our dental caries dataset and identified RF as the model displaying the highest performance. The proposed model highly predicted dental caries with an accuracy of 92% while using RF as a classification algorithm. When feature calibration was performed to increase the model performance using feature selection techniques such as ANOVA and RF, all ML methods witnessed a drop in their performances except RF. We performed experiments on various DL (CNN and LSTM) regardless of their usage specificity (CNN appropriate for the images dataset and LSTM for the time series dataset due to the history between the training sample) on the specific type of dataset. Nonetheless, these experiments were conducted for experimentation purposes, an observation of how the model would perform on such data.

A similar study on dental caries prediction by [38] had an accuracy of 97%; this study found some existing correlation between parameters such as age, income, date of last dental visit, hours of television watching, and the presence of caries in the studied population. According to their study, a person's age was one of the essential predictive values because of the exposure of the root surface of aged people. One of the barriers to dental care access was identified as to be the low income of some families, given they were unable to obtain proper and regular health care checkups. Parameters such as oral health variables, lifestyle, and demographics were essential features to their oral health status. In the current study, such parameters were not considered due to our dataset's lack of such information. Similarly, age was not considered an essential feature during our study and was removed. All patients in our study were of the same age (12 years old). Therefore, other specific parameters such as infectious tooth parameters collected by experts were used for our prediction. Some of the most critical features playing an essential role in predicting dental caries were captured by our ANOVA F test feature selection, as shown in our source file of feature selection on our repository. We demonstrated that using the complete set of features in this study was more effective than reducing the dimensionality of our dataset. In this investigation, ML methods helped us identify the probable cause (attribute features) of dental caries presence in patients and thus could help in the implementation of a recommendation and decision support system to facilitate diagnostic, early detection, and recommendation to future patients, given the rarity of oral health decision support system based on ML. This paper proposes a solution that would significantly reduce the time, cost, and human power required to perform a similar job in current dental healthcare.

6. Conclusions

In this paper, we proposed the DCP model that informs preventive measures or diagnostic solutions for dental caries using data collected from the children's oral health survey conducted in 2018 by the Korean Center for Disease Control and Prevention under the Ministry of Health and Welfare. The data were preprocessed, unnecessary features were removed (using random forest, ANOVA, and mutual information statistic methods) and then balanced using the widely used oversampling technique SMOTE. We conducted a comparative study using three DNN methods (ANN, CNN, and LSTM) and four binary classification ML models (RF, GBDT, SVM, and LR) on the cleaned and balanced dataset. After proper hyperparameters tuning, RF displayed the highest performance of 0.92, 0.90, 0.94, and 0.87 for accuracy, F1-score, precision, and recall, respectively, in predicting the presence or absence of dental caries.

The AutoEncoder can learn essential features of the various features available in the dataset that contain as much information as possible while using as little data as possible. For future works, we intend to apply the AutoEncoder during our data preparation phase to improve the performance of our model by dimensionality reduction, given the high number of features in the dataset. Simultaneously, we intend to build a new kernel method for our model and test it on our dataset. We are currently working on implementing a decision support system to help predict dental caries and provide recommendations to patients using this DCP model.

Author Contributions: Conceptualization, I.-A.K. and S.N.N.; methodology, I.-A.K., S.N.N., J.-D.K. and K.-O.L.; software, S.N.N.; validation, I.-A.K., S.N.N., J.-D.K. and K.-O.L.; formal analysis, I.-A.K.; investigation, I.-A.K. and S.N.N.; resources, I.-A.K. and S.N.N.; data curation, J.-D.K.; writing—original draft preparation, I.-A.K. and S.N.N.; writing—review and editing, S.N.N. and J.-D.K.; visualization, S.N.N.; supervision, J.-D.K. and K.-O.L.; project administration, J.-D.K. and K.-O.L.; 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: The dental caries training data were collected by statistical operation and can be found at https://www.korea.kr/common/download.do?tblKey=EDN&fileId=188769457 (accessed on 3 January 2021). The material used in our work is available at https://github.com/Soualihou237/Dental-Caries-Prediction-Using-Machine-Learning (accessed on 26 December 2021).

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

References

1. Dye, B.A.; Li, X.; Thorton-Evans, G. *Oral Health Disparities as Determined by Selected Healthy People 2020 Oral Health Objectives for the United States, 2009–2010*; National Center for Health Statistics: Washington, DC, USA, 2012; Volume 104, pp. 1–8.
2. Dye, B.A.; Tan, S.; Smith, V.; Lewis, B.G.; Barker, L.K.; Thornton-Evans, G.; Eke, P.I.; Beltrán-Aguilar, E.D.; Horowitz, A.M.; Li, C.-H. Trends in oral health status: United States, 1988–1994 and 1999–2004. *Vital Health Stat.* **2007**, *248*, 1–92.
3. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **2017**, *390*, 1211–1259. [CrossRef]
4. Korea National Children's Oral Health Survey. 2018. Available online: https://www.korea.kr/common/download.do?tblKey=EDN&fileId=188769457 (accessed on 3 January 2021).
5. Casamassimo, P.S.; Thikkurissy, S.; Edelstein, B.L.; Maiorini, E. Beyond the dmft: The human and economic cost of early childhood caries. *J. Am. Dent. Assoc.* **2009**, *140*, 650–657. [CrossRef] [PubMed]
6. Petersen, P.E.; Bourgeois, D.; Ogawa, H.; Estupinan-Day, S.; Ndiaye, C. The global burden of oral diseases and risks to oral health. *Bull. World Health Organ.* **2005**, *83*, 661–669.
7. Cummins, D. Dental caries: A disease which remains a public health concern in the 21st century—The exploration of a breakthrough technology for caries prevention. *J. Clin. Dent.* **2013**, *24*, 1–14.

8. Aggeryd, T. Goals for oral health in the year 2000: Cooperation between WHO, FDI and the national dental associations. *Int. Dent. J.* **1983**, *33*, 55–59.
9. Kim, A.H.; Han, S.Y.; Kim, B.I.; Kim, H.D.; Kwon, H.K. The characteristics of high caries risk group for 12-year-old children in Korea. *Korean Acad. Prev. Dent. Oral Health* **2010**, *34*, 302–309.
10. Hwang, D.-H.; Lee, J.-H. A study on DMFT index between elementary school students: Korea national health and nutrion examination survey. *J. Korean Soc. Oral Health Sci.* **2021**, *9*, 1–6. [CrossRef]
11. Seul, M.S. Current status and future developments of machine learning artificial intelligence in law: Focusing the cusp of machine learning in U.S. and discourses over legal profession and law school education. *Justice* **2016**, *156*, 269–302.
12. Velarde, G. Artificial Intelligence and its Impact on the Fourth Industrial Revolution: A Review. *Int. J. Artif. Intell. Appl.* **2019**, *10*, 41–48. [CrossRef]
13. 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. *J. Dent.* **2018**, *77*, 106–111. [CrossRef] [PubMed]
14. Lee, D. *Measurement and Evaluation of Bone Density for Dental Implant Surgery Using Synthetic Ghost-free Panoramic Radiograph*; Seoul National University: Seoul, Korea, 2020.
15. VUNO. "DeepBrain". Available online: https://www.vuno.co/en/boneage (accessed on 26 December 2021).
16. Lunit. "Lunit INSIGHT CXR1, Lunit INSIGHT CXR2, Lunit INSIGHT MMG". Available online: https://insight.lunit.io/ (accessed on 26 December 2021).
17. Tonetti, M.S.; Jepsen, S.; Jin, L.; Otomo-Corgel, J. Impact of the global burden of periodontal diseases on health, nutrition and wellbeing of mankind: A call for global action. *J. Clin. Periodontol.* **2017**, *44*, 456–462. [CrossRef] [PubMed]
18. Armitage, G.C. Development of a Classification System for Periodontal Diseases and Conditions. *Ann. Periodontol.* **1999**, *4*, 1–6. [CrossRef]
19. Caton, J.G.; Armitage, G.; Berglundh, T.; Chapple, I.L.; Jepsen, S.; Kornman, K.S.; Mealey, B.L.; Papapanou, P.N.; Sanz, M.; Tonetti, M.S.; et al. A new classification scheme for periodontal and peri-implant diseases and conditions—Introduction and key changes from the 1999 classification. *J. Periodontol.* **2018**, *89*, S1–S8. [CrossRef] [PubMed]
20. Kwon, O.; Yong, T.-H.; Kang, S.-R.; Kim, J.-E.; Huh, K.-H.; Heo, M.-S.; Lee, S.-S.; Choi, S.-C.; Yi, E.-J. Automatic diagnosis for cysts and tumors of both jaws on panoramic radiographs using a deep con-volution neural network. *Dentomaxillofac. Radiol.* **2020**, *49*, 20200185. [CrossRef] [PubMed]
21. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef]
22. Kallenberg, M.; Petersen, K.; Nielsen, M.; Ng, A.Y.; Diao, P.; Igel, C.; Vachon, C.M.; Holland, K.; Winkel, R.R.; Karssemeijer, N.; et al. Unsupervised Deep Learning Applied to Breast Density Segmentation and Mammographic Risk Scoring. *IEEE Trans. Med. Imaging* **2016**, *35*, 1322–1331. [CrossRef]
23. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef]
24. Hannun, A.Y.; Rajpurkar, P.; Haghpanahi, M.; Tison, G.H.; Bourn, C.; Turakhia, M.P.; Ng, A.Y. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat. Med.* **2019**, *25*, 65–69. [CrossRef]
25. Lee, H.; Park, M.; Kim, J. Cephalometric landmark detection in dental x-ray images using convolutional neural networks. In Proceedings of the SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis, Orlando, FL, USA, 13–16 February 2017; p. 101341W.
26. Ronneberger, O.; Fischer, P.; Brox, T. Dental X-ray image segmentation using a U-shaped Deep Convolutional network. In Proceedings of the International Symposium on Biomedical Imaging, New York, NY, USA, 16–19 April 2015.
27. Miki, Y.; Muramatsu, C.; Hayashi, T.; Zhou, X.; Hara, T.; Katsumata, A.; Fujita, H. Classification of teeth in cone-beam CT using deep convolutional neural network. *Comput. Biol. Med.* **2017**, *80*, 24–29. [CrossRef]
28. Hiraiwa, T.; Ariji, Y.; Fukuda, M.; Kise, Y.; Nakata, K.; Katsumata, A.; Fujita, H.; Ariji, E. A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography. *Dentomaxillof. Radiol.* **2019**, *48*, 20180218. [CrossRef] [PubMed]
29. Murata, M.; Ariji, Y.; Ohashi, Y.; Kawai, T.; Fukuda, M.; Funakoshi, T.; Kise, Y.; Nozawa, M.; Katsumata, A.; Fujita, H.; et al. Deep-learning classification using convolutional neural network for evaluation of maxillary sinusitis on panoramic radiography. *Oral Radiol.* **2018**, *35*, 301–307. [CrossRef] [PubMed]
30. Krois, J.; Ekert, T.; Meinhold, L.; Golla, T.; Kharbot, B.; Wittemeier, A.; Dörfer, C.; Schwendicke, F. Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Sci. Rep.* **2019**, *9*, 8495. [CrossRef] [PubMed]
31. Kim, J.; Lee, H.-S.; Song, I.-S.; Jung, K.-H. DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs. *Sci. Rep.* **2019**, *9*, 17615. [CrossRef] [PubMed]
32. Chen, C.; Bai, W.; Davies, R.H.; Bhuva, A.N.; Manisty, C.H.; Augusto, J.; Moon, J.; Aung, N.; Lee, A.M.; Sanghvi, M.M.; et al. Improving the Generalizability of Convolutional Neural Network-Based Segmentation on CMR Images. *Front. Cardiovasc. Med.* **2020**, *7*, 105. [CrossRef] [PubMed]
33. Veena, D.K.; Jatti, A.; Joshi, R.; Deepu, K.S. Characterization of dental pathologies using digital panoramic X-ray images based on texture analysis. In Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Korea, 11–15 July 2017; pp. 592–595.

34. Singh, P.; Sehgal, P. Automated caries detection based on Radon transformation and DCT. In Proceedings of the 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, 3–5 July 2017; pp. 1–6.
35. Ghaedi, L.; Gottlieb, R.; Sarrett, D.C.; Ismail, A.; Belle, A.; Najarian, K.; Hargraves, R.H. An automated dental caries detection and scoring system for optical images of tooth occlusal surface. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 1925–1928. [CrossRef]
36. Kuang, W.; Ye, W. A Kernel-Modified SVM Based Computer-Aided Diagnosis System in Initial Caries. In Proceedings of the Second International Symposium on Intelligent Information Technology Application, IEEE, Shanghai, China, 20–22 December 2008; Volume 3, pp. 207–211.
37. Nabilla, M.; Brenda, C.; Ichwan, S.J.A.; Arief, C. Deep learning convolutional neural network algorithms for the early detection and diagnosis of dental caries on periapical radiographs: A systematic review. *Imaging Sci. Dent.* **2021**, *51*, 237. [CrossRef]
38. 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. [CrossRef] [PubMed]
39. Yang, Y.-H.; Kim, J.-S.; Jeong, S.-H. Prediction of dental caries in 12-year-old children using machine-learning algorithms. *J. Korean Acad. Oral Health* **2020**, *44*, 55–63. [CrossRef]
40. Ibrahim, O.; Osman, A.H. A Novel Feature Selection Based on One-Way ANOVA F-Test for E-Mail Spam Classification. *Res. J. Appl. Sci. Eng. Technol.* **2014**, *7*, 625–638. [CrossRef]
41. Kotsiantis, S.; Kanellopoulos, D.; Pintelas, P. Handling imbalanced datasets: A review. *GESTS Int. Trans. Comput. Sci. Eng.* **2006**, *30*, 25–36.
42. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
43. Khoshgoftaar, T.M.; Golawala, M.; Van Hulse, J. An Empirical Study of Learning from Imbalanced Data Using Random Forest. In Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), Patras, Greece, 29–31 October 2007; Volume 2, pp. 310–317.
44. Dong, W.; Huang, Y.; Lehane, B.; Ma, G. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. *Autom. Constr.* **2020**, *114*, 103155. [CrossRef]
45. Suykens, J.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [CrossRef]
46. Smola, A.J. Learning with Kernels. Ph.D. Thesis, Informatik der Technischen Universit at Berlin, Birlinghoven, Germany, 1998.
47. Smola, A.J.; Schölkopf, B. *A Tutorial on Support Vector Regression*; NeuroCOLT Tech. Rep. TR 1998-030; Royal Holloway College: London, UK, 1998.
48. Shevade, S.K.; Keerthi, S.S.; Bhattacharyya, C.; Murthy, K.R.K. Improvements to the SMO Algorithm for SVM Re-gression. *IEEE Trans. Neural Netw.* **2000**, *11*, 1188–1193. [CrossRef] [PubMed]
49. Kurt, I.; Ture, M.; Kurum, A.T. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Syst. Appl.* **2008**, *34*, 366–374. [CrossRef]
50. Zhang, Z. A gentle introduction to artificial neural networks. *Ann. Transl. Med.* **2016**, *4*, 370. [CrossRef]
51. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
52. Shi, Z.; Chehade, A. A dual-LSTM framework combining change point detection and remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2021**, *205*, 107257. [CrossRef]