applied sciences

MDPI

*Article*

# Deep Fusion Feature Extraction for Caries Detection on Dental Panoramic Radiographs

**Toan Huy Bui [1,*], Kazuhiko Hamamoto [2,*] and May Phu Paing [3]**

[1] Course of Science and Technology, Graduate School of Science and Technology, Tokai University, Tokyo 108-8619, Japan
[2] School of Information and Telecommunication Engineering, Tokai University, Tokyo 108-8619, Japan
[3] Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand; 59601030@kmitl.ac.th
* Correspondence: 9ltad006@mail.u-tokai.ac.jp (T.H.B.); hama@keyaki.cc.u-tokai.ac.jp (K.H.)

**Abstract:** Caries is the most well-known disease and relates to the oral health of billions of people around the world. Despite the importance and necessity of a well-designed detection method, studies in caries detection are still limited and show a restriction in performance. In this paper, we proposed a computer-aided diagnosis (CAD) method to detect caries among normal patients using dental radiographs. The proposed method mainly consists of two processes: feature extraction and classification. In the feature extraction phase, the chosen 2D tooth image was employed to extract deep activated features using a deep pre-trained model and geometric features using mathematic formulas. Both feature sets were then combined, called fusion feature, to complement each other defects. Then, the optimal fusion feature set was fed into well-known classification models such as support vector machine (SVM), k-nearest neighbor (KNN), decision tree (DT), Naïve Bayes (NB), and random forest (RF) to determine the best classification model that fit the fusion features set and perform the most preeminent result. The results show 91.70%, 90.43%, and 92.67% for accuracy, sensitivity, and specificity, respectively. The proposed method has outperformed the previous state-of-the-art and shows promising results when none of the measured factors is less than 90%; therefore, the method is promising for dentists and capable of wide-scale implementation caries detection in hospitals.

**Keywords:** caries; tooth decay; dental radiographs; deep learning; features extraction; machine learning

check for updates

**Citation:** Bui, T.H.; Hamamoto, K.; Paing, M.P. Deep Fusion Feature Extraction for Caries Detection on Dental Panoramic Radiographs. *Appl. Sci.* **2021**, *11*, 2005. https://doi.org/10.3390/app11052005

Academic Editor: Arjan Kuijper

Received: 31 December 2020
Accepted: 19 February 2021
Published: 24 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Oral health plays a main role in people's overall health and quality throughout their lifetime, regardless of nationality, region, or religion. It is healthy conditions without mouth and facial pain, oral and throat cancer, oral infection and sores, periodontal (gum) diseases, tooth decay, tooth loss, and disorders that limit an individual's capacity in biting, chewing, speaking, and psychosocial wellbeing. The World Health Organization (WHO) estimated that around 3.5 billion people were affected by oral diseases in 2016 and continuedly increasing [1]. Caries, also known as tooth decay or oral cavities, is the most common disease that affects the quality of life worldwide. Around 60%–90% of school children and almost 100% of adults have dental cavities. Caries is the breakdown of teeth due to acid made by bacteria. The symptoms of untreated caries come in different forms or colors, such as yellow or black, resulting in oral pain, facial pain, tooth loss, and is a major cause of noncommunicable disease. Treatment of oral diseases is usually expensive and not part of universal health coverage. Dental treatment costs 5% of total health spending and is generally a 20% out-of-pocket expenditure in many developed countries. The condition seems to be worse in most developing countries where people cannot afford oral health treatment services. Most caries conditions are treatable and preventable in the early stage,

therefore reducing the dentist's effort and expenditure. Figure 1 describes an example of a healthy tooth and a tooth with cavities.

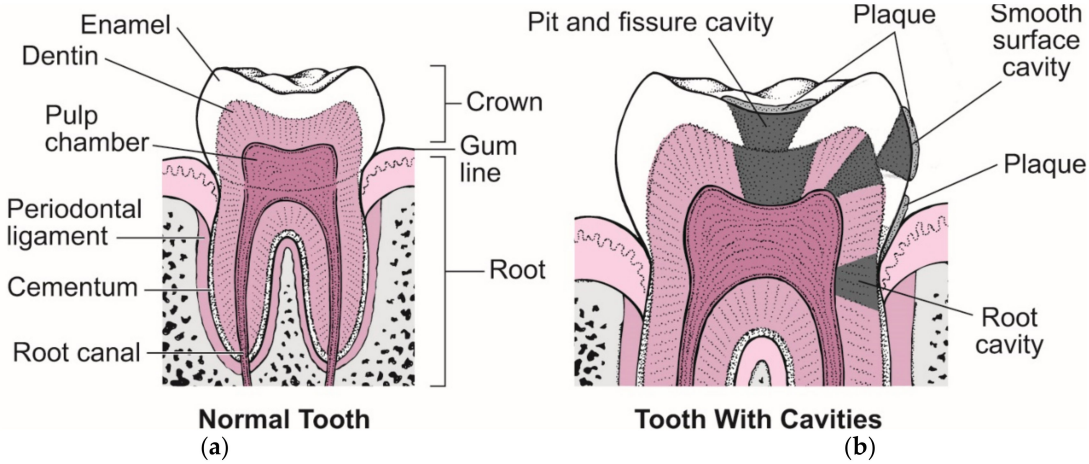



**Figure 1.** Sample of healthy and caries tooth image [2]. (**a**) Structure of healthy tooth and (**b**) a tooth with decay.

Detection of caries may consist of three phases: (1) segment (or isolate) the diagnosis tooth from other teeth; (2) preliminary diagnosis to determine whether a tooth has decay; (3) comprehensive diagnosis to make a treatment for the decaying tooth as well as classify the stage of decay into four groups (C1–C4) based on the condition and damage of the tooth. Although a nurse could perform phase one, phases two and three may need practical experience from a dentist. In this research, we aim to develop a method to make a preliminary diagnosis at phase two to reduce the dentist's effort on non-caries patients.

Recently, with the development of medical imaging technology, computer-aided diagnosis systems (CADs) play the main role in the early detection of several diseases such as cancer, diabetes, or even caries [3,4]. Caries can be detected using several different methods and techniques. Some researchers proposed detection using photoacoustic images, wavelengths, or ultrasound images [5–7]. Other research has detailed an approach using an RGB oral endoscope image [8,9]; however, most approaches cannot provide a detailed structure of the tooth, especially the tooth root, and therefore struggle to provide caries diagnosis. Compared to oral endoscope imaging, dental radiographs provide greater image quality and detect detailed structure deformity in the tooth [10]; therefore, the dental radiographic image is the most well-known approach, and is preferable for the detection of caries in the early stage.

Clinically, dental radiographs, which are used to identify teeth problems and evaluate oral health, were taken by X-ray with a low level of radiation to capture images of the interior of teeth and gums. Radiographs are usually shown in grayscale images or sometimes in color images; however, color radiographs require significant investment, which provides a barrier to entry for most hospitals, especially hospitals in low-income countries; to account for this, we focused on grayscale radiographs. Unfortunately, there is no reliable dataset that provides high-quality images, descriptions, and reliable ground truth. In this field, most data are only shared within some strict conditions, such as all the researchers must publish to a specific journal or must be a member of some group or event. Some researchers publish their private data used in their research. The data usually have problems with the quality of the image, size of data, lack of description and ground-truth, and/or lack of data availability in the long-term. In this research, dataset and ground truth were provided by Dr. Kumon Makoto, director of Shinjuku East Office, under a research contract with Tokai University. Dr. Kumon Makoto received The Academy of Clinical Dentistry Certified Physician and registered as a professional dentist with No.148529 on 19 May 2003. With 18 years of experience as a dentist and responsibility for over 200 patients per month, he could reliably provide a truthful dataset. More importantly, all the patients

who participated in the dataset collection were real patients of Dr. Makoto and under his treatment. Each caries tooth in the dataset was already confirmed in the patient's medical history during the treatment. For reasons mentioned above, we believe that our dataset is trustworthy and can be used for research and publication purposes.

## 2. Related Works

In the dental examination using radiographs, caries can be recognized as a break in the tooth, parts missing from a tooth, or tooth loss. There is no obvious symptom or criteria on the shape, size, or intensity for tooth decay except the dentist's diagnosis experience, which causes a huge challenge for computer-aided diagnosis systems based on image processing. Wei Li et al. [11] proposed a method to detect tooth decay using a support vector machine (SVM) and a backpropagation neural network (BPNN). The method uses two features set separately for features extraction: Autocorrelation Coefficient and Gray-level Co-occurrence Matrix. Then, a model of SVM and BPNN was applied separately for classification purposes. The result shows that SVM has around 79% accuracy on the testing set, whereas BPNN is around 75% accuracy. The result is inefficient and needs more works for improvement. Besides this, in the article, the dataset's description is not mentioned; thus, it may lead to questions about the research's reliability.

Yang Yu et al. [12] tried to enhance the backpropagation neural network layer and features extraction of autocorrelation coefficient matrix. The method was tested on 80 private tooth images (55 images for training and 35 images for testing and shows 94% accuracy); however, there is a great computational burden when the number of layers in the backpropagation neural network is increased. In addition, effective measures, such as sensitivity (SEN), specificity (SPEC), precision (PRE), and F-measure, are not mentioned. Further, the pretty small testing data (35 images) without cross-validation also shows weakness, which cannot address the whole problem of tooth decay.

Shashikant Patil [13] proposed an intelligent system with dragonfly optimization. Multi-linear principal component analysis (MPCA) was applied to extract the feature set. The features set were then fed into a neural network classifier trained using an optimization method, which was the adaptive dragonfly algorithm (ADA). The proposed MPCA model non-linear Programming with ADA (MNP-ADA) was tested with 120 private tooth images divided into three test cases. Each test case consisted of 40 images, 28 images were used for training, and 12 images were used for testing. The other classifiers, such as fruit fly (FF) [14] and grey-wolf optimization (GWO) [15], and feature sets, such as linear discriminant analysis (LDA) [16], principal component analysis (PCA) [17], and independent component analysis (ICA) [18], were also used in the testing for comparison. The final average results show that the MNP-ADA model reaches 90% accuracy, 94.67% sensitivity, and 63.33% specificity. The result shows a low performance of specificity, which describes non-caries patients misclassified as caries patients; therefore, the distinction between caries and non-caries patients is not efficient, so the performance needs to be improved. Because the result shows a high accuracy value despite a low specificity value, this may lead to hesitations about the balance of data between caries and non-caries images. This study also shows other measure values, such as precision and f1-score, which are discussed in more detail in the results section.

Nowadays, deep learning makes a great breakthrough in the machine learning field [19]. The convolutional neural network (CNN) is the most well-known deep learning model, which could be used for many purposes, such as the detection of new unknown objects (transfer learning), fine-tuning the weight, or feature extraction [20–23]; however, so far as we are aware, there is no previous study that applied deep learning for caries classification problems, especially in dental radiographs; meaning that there may be a need for research in this area. In addition, a single CNN model may result in an unsatisfied performance and neglect a large space of unexplored potential of image data. Thus, there is a need for improvement from the deep activated features by combining other sources for more features. Consequently, in this study, we propose a deep activated model that can best

describe dental radiographs and improve the performance of the feature set by combining other mathematic features such as mean, standard deviation (STD), and texture features. Each deep activated model features set is extracted carefully by testing the result of each considerably deep layer. The mathematic features are also tweaked to get the minimum features set while maintaining optimal performance. The combination features set, called the "fusion feature" in this study, is later fed into different classification models to find the best models that fit the features set and perform the best distinction in data. This study focused on two key objectives:

(i) Stability, which based on the large data, can describe the problem and cross-validation to measure the different situation of problem;
(ii) Performance is a better result in accuracy and improves specificity since the balance between sensitivity and specificity is sometimes more important than accuracy. Other measures are also shown for comparison with the previous study.

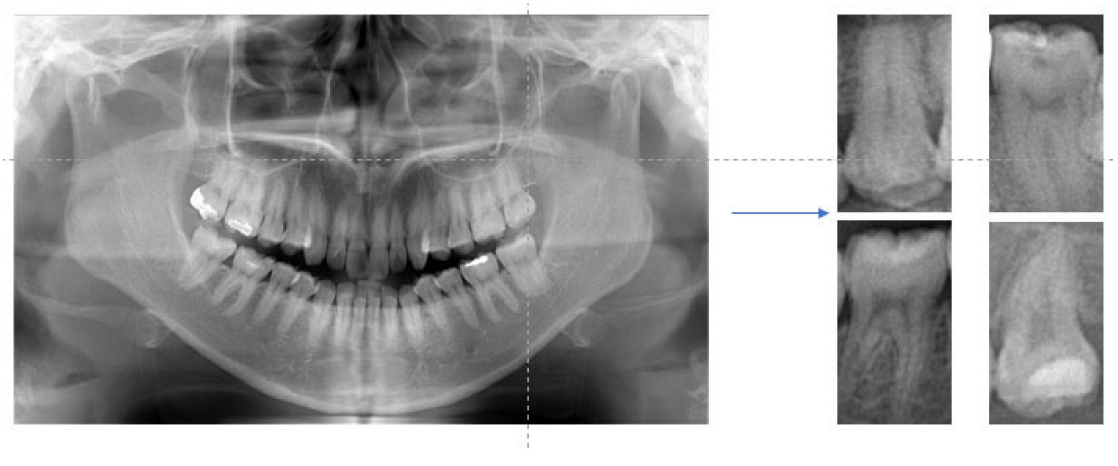
The rest of the paper is organized as follows. Section 3 describes the dataset and proposed method and describes how to implement our method step-by-step. Section 4 shows the results of each step described in Section 3. The results of previous studies are also mentioned for comparison. Section 5 provides discussion, a summary, and conclusions.

## 3. Materials and Methods

This section describes the proposed method as well as gives information about our dataset. Since there is no specific well-known public dataset in this field, a carefully prepared dataset is important for evaluating the proposed method; thus, most researchers prefer to build their own datasets for experiments [11–13].

### *3.1. Radiographs Dataset*

To the best of our knowledge, the tooth is diverse in size, shape, and structure. Characteristics of tooth decay contribute even more to this diversity; therefore, the larger a dataset is, the better it can describe the tooth decay issues. Our dataset was collected and labeled by a dentist from the Tokai hospital. The dataset was assessed for quality and ethics by Tokai university's committee for the right of use and publication; however, the dataset's images are panoramic oral radiographs of all teeth, whereas dental diagnosis and treatment should be made for every individual tooth. Consequently, we needed to manually segment the tooth into each sub-image, which consists of the target tooth, which needs the diagnosis, and its label. The segmentation is simple and can be done by any dentist or nurse; therefore, we anticipate no considerable effect in this study (Figure 2). To simulate the real cases, where the area determined for each tooth varies between whoever performs the segmentation, we do not take the area and range of tooth fixed in any size but consider it very flexible depending on the tooth's size and position and surrounding space.



(**a**) (**b**)

**Figure 2.** Samples of oral and tooth image. (**a**) oral panoramic radiograph and (**b**) segmented tooth radiographs.

After the segmentation, the dataset comprised 533 image samples: 229 caries teeth and 304 non-caries teeth. Since the difference in the number of caries and non-caries images remain a small proportion (caries/non-caries is approximate 0.43/0.57), the dataset can be considered as balanced data. Each image is a two-dimensional grayscale image that consists of the target tooth and its surrounding areas, such as black empty space or a part of neighboring teeth. The images present the original condition of teeth without any modification in color, size, or angle. All images are flexible in size, which matches the standard segmentation process and is then fed into the same size layer for the feature extraction step.

### *3.2. Method*

Caries detection mainly consists of two stages: feature extraction and classification. In the first stage, we experimented to find the deep activated features from pre-trained models that best describe radiography, such as Alexnet [20], Googlenet [24], VGG16 [25], VGG19 [25], Resnet18 [26], Resnet50 [26], Resnet101 [26], and Xception [27] networks. The experiments were done in the deepest layers of each model. Later, the mathematic features, such as mean, STD, and texture features such as Haralick's features [28], were extracted to improve the feature's information. Both features set are later combined into fusion features. The second stage is where we test the feature set in the classification models, such as support vector machine (SVM), Naïve Bayes (NB), k-nearest neighbor (KNN), decision tree (DT), and random forest (RF). The whole process, along with other sub-stages, are shown in Figure 3.

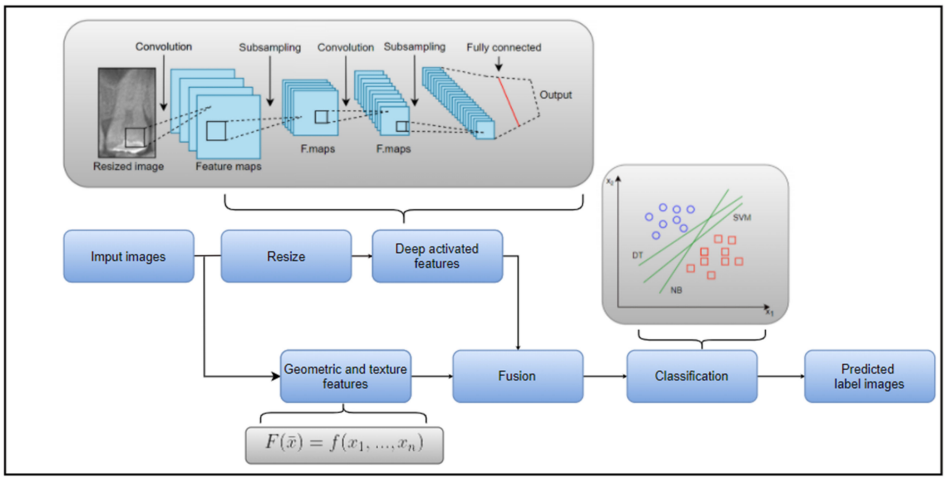



**Figure 3.** Diagram for caries prediction.

#### 3.2.1. Features Descriptors using Pre-trained CNN Deep Networks

A pre-trained CNN is used in this study as a feature descriptor to extract the deep activated features. The eight most well-known networks, Alexnet, Googlenet, VGG16, VGG19, Resnet18, Resnet50, Resnet101, and Xception network, were applied to find the best descriptor pre-trained networks. Table 1 describes each pre-trained model specifications in detail, such as depth, parameter, size, and input size. The most common recommended layers for extraction are usually the latest layer before the "prediction" layer for the deepest learning rate; therefore, for our experiment so far, we tested several layers before the "prediction" layer (except the "drop" layer because the drop layer likely shows the same information as the previous respective layer). The image needs to be resized at a specific size before feeding it into a particular network. Technically, the network processes RGB images when the radiographs are grayscale; therefore, we multiplied the grayscale channel to replace the missing channel in the image. The layers and network tests are shown in the Results section.

**Table 1.** Convolutional neural network (CNN) model specification.

| Network Model | Depth | Size (MB) | Parameter ($\times 10^6$) | Input Size |
|---|---|---|---|---|
| Alexnet | 8 | 227 | 61.0 | 227×227×3 |
| Googlenet | 22 | 27 | 7.0 | 224×224×3 |
| VGG16 | 23 | 528 | 138.4 | 224×224×3 |
| VGG19 | 26 | 549 | 143.7 | 224×224×3 |
| Resnet18 | 18 | 45 | 11.5 | 224×224×3 |
| Resnet50 | 50 | 98 | 25.6 | 224×224×3 |
| Resnet101 | 101 | 171 | 44.7 | 224×224×3 |
| Xception | 126 | 88 | 22.9 | 299×299×3 |

#### 3.2.2. Features Descriptors using Geometric Features

Geometric features are a fundamental factor to describe any kind of problem. Since the features are extracted using a mathematic formula, the features are therefore understandable and explainable. Despite of the contribution of deep activated features descriptors, geometric features can contain sufficient and relevant information that is noticeable to human. Furthermore, deep activated features usually explore the data in a way impenetrable for human, whereas the geometric features are usually learned from the expert's experience in the field; therefore, the geometric features are necessary and irreplaceable for solving a complex problem.

In clinical practice, dentists manually determine the difference between caries and non-caries depending on the damage of structure in the tooth. The damage to the tooth's structure can be explained by the difference in size, shape, contrast, margin, intensity, and so on. Based on their characteristic, the suspicious features, which describes the state of the tooth, is extracted such as mean, Haralick's features [28], and gray level co-occurrence matrices (GLCM) features [29,30]. Table 2 describes the name and formula of used features in detail. In the formula, $I(x, y)$ presents the pixel value $I$ at the coordinate point $x, y$ of the candidate image $N$. $p(i, j)$ presents the $(i, j)^{th}$ entry of GLCM matrix. $N_g$ presents number of distinct gray levels in the image. $\mu$ and $\sigma$ present the mean and standard deviation values.

#### 3.2.3. Fusion Features

The extracted feature in deep networks and geometric features are combined in this step. The whole extract geometric features are connected to each deep activated feature. The fusion feature is then fed into a classification model in the next step. Also, to measure the efficiency of the geometric feature as well as fusion feature over deep activated features, we measured the performance by feeding each deep activated feature and fusion features into the classifier at the same condition as fusion features (Figure 4). The comparison of the result between fusion and deep activated features is discussed in the Results section.

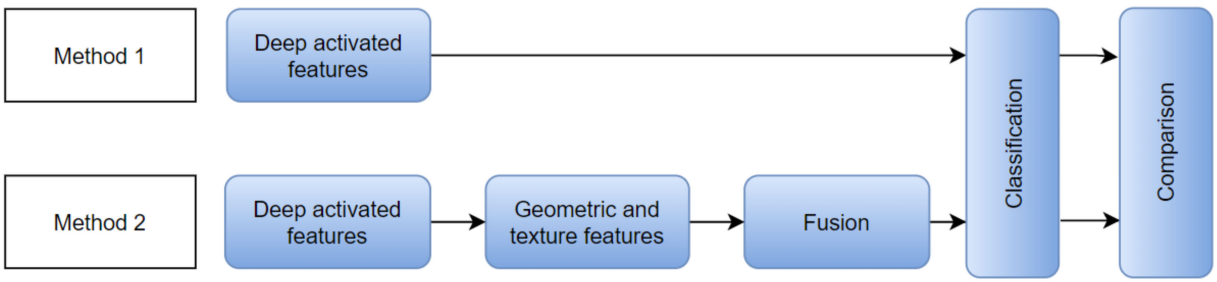



**Figure 4.** Diagram for experiment deep activated and fusion features.

**Table 2.** Geometric features and formula.

| Features | Name | Formula |
|---|---|---|
| F1 | Mean | $\mu = \frac{1}{n} \sum_{(x,y)\in N} I(x,y)$ |
| F2 | Entropy | $E = -\sum_i \sum_j p(i,j) \log p(i,j)$ |
| F3 | Autocorrelation | $AutoCorr = \sum_i \sum_j (i \cdot j) p(i,j)$ |
| F4 | Contrast | $Cont = \sum_i \sum_j \lvert i-j \rvert^2 p(i,j)$ |
| F5 | Correlation | $Corr = \sum_i \sum_j \frac{(i-u_x)(j-u_y)p(i,j)}{\sigma_x \sigma_y}$ |
| F6 | Cluster prominence | $Cont = \sum_i \sum_j (i+j-\mu_x-\mu_y)^4 p(i,j)$ |
| F7 | Cluster shade | $Shade = \sum_i \sum_j (i+j-\mu_x-\mu_y)^3 p(i,j)$ |
| F8 | Dissimilarity | $Cont = \sum_i \sum_j \lvert i-j \rvert \cdot p(i,j)$ |
| F9 | Maximum probability | $MaxProb = \max_{i,j} p(i,j)$ |
| F10 | Sum of square variance | $SumVar = \sum_i \sum_j \lvert i-\mu^2 \rvert p(i,j)$ |
| F11 | Sum of average | $SumAvg = \sum_{i=2}^{2N_g} i \cdot p_{x+y} i$ |
| F12 | Sum of entropy | $SumEnt = \sum_{i=2}^{2N_g} p_{x+y}(i) \log p_{x+y}(i)$ |
| F13 | Sum of variance | $\text{SumVar} = \sum_{i=2}^{2N_g} (i - SumEnt)^2 \cdot p_{x+y}(i)$ |
| F14 | Difference entropy | $DiffEnt = -\sum_{i=0}^{N_g-1} p_{x-j}(i) \log p_{x-y}(i)$ |

#### 3.2.4. Classification

Each deep activated feature is combined with geometric features and then fed into the classification model. To test the fusion efficiency between deep activated and geometric features, the deep activated features are also tested separately and compared with the fusion features. Most tests were conducted using a well-known "optimal margin classifier," also known as support vector machine (SVM) [31].

The SVM model aims to find the optimal hyperplane that can best describe the distinction between data, caries, and non-caries in this case. To moderate the number of training points, we apply the Gaussian radial basis function in the classifier. For a given training data $D = \{(x_i, y_i),\ i = 1 \ldots N\}$ and $y_i \in \{-1, 1\}$ the SVM classifier and the mapping function of the Gaussian kernel can be described as follows in Equations (1) and (2):

$$\min_{\omega, b,\ \xi} \frac{1}{2} ||W||^2 + C \sum_i \xi_i^2 \ subject\ to\ y_i \left( W^T X_i + b \right) \geq 1 - \ \xi_i,\ \xi_i \geq 0,\ \forall i \tag{1}$$

where C > 0 is the selected parameter and $\xi$ is a set of slack variables.

$$\mathrm{K}\ (\mathrm{X},\ \mathrm{Y})\ =\ e^{\frac{||X-Y||^2}{A}} \tag{2}$$

where K is the kernel function and A is the constant.

Furthermore, to guarantee the best classification which fit the features set, we also test the best features set on k-nearest neighbor (KNN) [32], decision tree (DT) [33], Naïve Bayes (NB) [34], and Random Forest (RF) [35].

## 4. Experimental Results

This section describes how we conducted the experiment as well as gives information on the experiment environment. The result of each step is explained in detail. The best result was compared to the previous state-of-the-art.

### 4.1. Measures

Performance assessment of the proposed method in this study relied on three well-known measures: accuracy (ACC), sensitivity (SEN), and specificity (SPEC). Plus, we also present precision or positive predictive value (PPV), negative predictive value (NPV), f1-score, the area under the curve (AUC), and processing time to give a comprehensive view about the advantage of the proposed method and for other research reference purposes. The calculation of the measure can be explained as follows in Equations (3)–(8):

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$SEN = \frac{TP}{TP + FN} \tag{4}$$

$$SPEC = \frac{TN}{TN + FP} \tag{5}$$

$$PPV = \frac{TP}{TP + FP} \tag{6}$$

$$NPV = \frac{TN}{TN + FN} \tag{7}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{8}$$

where:

True positive (TP) presents the number of caries images classified correctly as caries;
True negative (TN) presents the number of non-caries images classified correctly as non-caries;
False-positive (FT) means the number of non-caries images classified wrongly as caries;
False-negative (FN) means the number of caries images classified wrongly as non-caries.

### 4.2. Experiment and Result

In the first stage of the experimental result, we define the optimal layers in each deep pre-trained network that best represents the problem. Table 3 explains the features set extracted from each deep pre-trained network respective to their layer. The extracted features set are tested with a support vector machine model to reach the final classification result. There is no reference for choosing the layer in each network; therefore, we tried several layers before its prediction layer. Some of them show that the best layers are the pooling layer, whereas others may choose the layer before. The highest performance can be reached from the "fc8" layer in the VGG16 model, which presents an accuracy of 90.57%, sensitivity of 91.30%, and specificity of 90%. Furthermore, Resnet50, Resnet101, and Xception also show a very promising result of around 88% accuracy. Noticeably, none of the deep activated features have less than 80% accuracy, proving that deep activated features are effective.

**Table 3.** Performance of deep activated features layer corresponding to networks.

| Network | Alexnet | Googlenet | VGG16 | VGG19 | Resnet18 | Resnet50 | Resnet101 | Xception |
|---|---|---|---|---|---|---|---|---|
| Layer | fc8 | pool5-7x7_s1 | fc8 | fc8 | pool5 | avg_pool | pool5 | avg_pool |
| ACC | 0.8679 | 0.8302 | **0.9057** | 0.8113 | 0.8491 | 0.8868 | 0.8868 | 0.8868 |
| SEN | 0.7826 | 0.8261 | **0.9130** | 0.7391 | 0.8261 | 0.8696 | 0.8261 | **0.9130** |
| SPEC | **0.9333** | 0.8333 | 0.9000 | 0.8667 | 0.8667 | 0.9000 | **0.9333** | 0.8667 |
| PPV | 0.9000 | 0.7919 | 0.8750 | 0.8095 | 0.8261 | 0.8696 | **0.9048** | 0.8400 |
| NPV | 0.8485 | 0.8621 | **0.9310** | 0.8125 | 0.8667 | 0.9000 | 0.8750 | 0.9286 |
| F1-score | 0.7200 | 0.6786 | **0.8077** | 0.6296 | 0.7037 | 0.7692 | 0.7600 | 0.7778 |
| AUC | 0.9087 | 0.8333 | **0.9587** | 0.8674 | 0.9014 | 0.9565 | 0.9072 | 0.9464 |

The highest performance for each measured factor regarding to network was highlighted in bold.

To enhance the performance so far, we combined each deep activated feature set with geometric features and fed them into the SVM model (Table 4). The result of the fusion feature shows that the fusion Xception feature evolved. After the combination, Figure 5 shows that the fusion features of the Xception network have become the most prominent features and improved the performance to 90.45%, 100%, and 86.67% for accuracy, sensitivity, and specificity, respectively. The highest difference is the improvement of sensitivity from 91% to 100%; therefore, the fusion Xception features set has demonstrated geometric contribution in the proper combination with deep activated features. Although the performance is compatible with Xception fusion features, Resnet18 and Googlenet also show an improvement of 83.02% to 86.79% accuracy and 84.91% to 88.68% accuracy, respectively. Noticeably, none of the fusion feature sets have lower accuracy than their respective deep activated features. In conclusion, fusion features show an obvious advantage on deep activated features set alone.

**Table 4.** Performance of fusion features corresponding to networks.

| Network | Alexnet | Googlenet | VGG16 | VGG19 | Resnet18 | Resnet50 | Resnet101 | Xception |
|---|---|---|---|---|---|---|---|---|
| ACC | 0.8679 | 0.8679 | 0.9057 | 0.8113 | 0.8868 | 0.8868 | 0.8868 | **0.9245** |
| SEN | 0.7826 | 0.8696 | 0.9130 | 0.7826 | 0.8696 | 0.8696 | 0.8261 | **1.0000** |
| SPEC | 0.9333 | 0.8667 | 0.9000 | 0.8333 | 0.9000 | 0.9000 | **0.9333** | 0.8667 |
| PPV | 0.9000 | 0.8333 | 0.8750 | 0.7826 | 0.8696 | 0.8696 | **0.9048** | 0.8519 |
| NPV | 0.8485 | 0.8966 | 0.9310 | 0.8333 | 0.9000 | 0.9000 | 0.8750 | **1.0000** |
| F1-score | 0.7200 | 0.7407 | 0.8077 | 0.6429 | 0.7692 | 0.7692 | 0.7600 | **0.8519** |
| AUC | 0.9087 | 0.8949 | 0.9594 | 0.8659 | 0.9123 | 0.9580 | 0.9087 | **0.9688** |

The highest performance for each measured factor regarding to network was highlighted in bold.

We randomly divided the training and testing set for cross-validation to design and evaluate the appropriate caries detection method. The k-fold cross-validation is a well-known reliable technique to test the robustness of the method. The application of k-fold cross validation proves the proposed method's reliability to cover the whole problem and adapt to the unknown samples; this technique was also used to prevent the overfitting of the method on our testing data.

We then used the most prominent results obtained by the different classification models to determine which classification model best fit the features set. In this study, decision tree (DT), k-nearest neighbor (KNN), Naïve Bayes (NB), random forest (RF), and support vector machine were used (Table 5). In this step, we also applied k-fold cross-validation to prevent the overfitting of the method and to calculate the final average assessment. The support vector machine is obviously the most dominant model that shows an accuracy of 91.70%, sensitivity of 90.43%, and specificity of 92.67%. As mentioned earlier in Section 3.1, the used dataset is considered balanced and shows a small difference in the number of caries and non-caries samples; therefore, precision (also known as positive predictive value) and recall pair (also known as sensitivity) also show promising values of 91.51% and 90.43%, respectively. Finally, we generated the receiver operating characteristic(ROC) curves for each classifier; the ROC curves describe each classifier in each fold of the experi-

ment. The mean ROC curve of each classifier is interpolated in each graph from Figure 6a to Figure 6e and compared in Figure 6f.

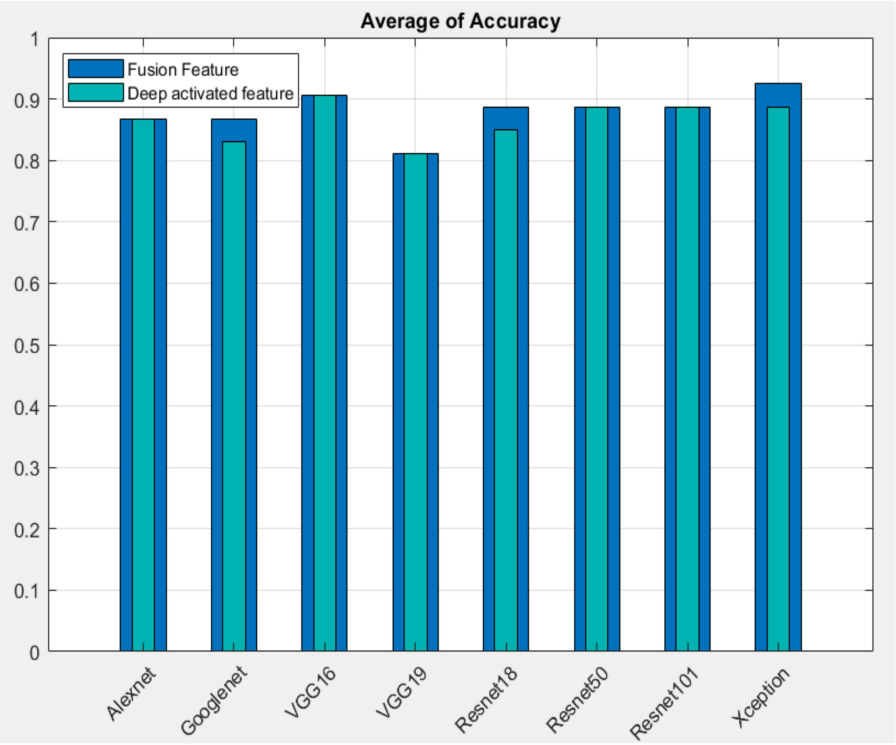



**Figure 5.** Overlay bar graphs distribution of average accuracy between fusion feature and deep activated feature.

**Table 5.** Performance of fusion features based on classifiers.

| Classifier | Measure | Five-Fold Cross-Validation | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Fold-1** | **Fold-2** | **Fold-3** | **Fold-4** | **Fold-5** | **Mean** |
| Decision Tree | Accuracy | 0.6415 | 0.6038 | 0.7170 | 0.6038 | 0.6981 | 0.6528 |
| | Sensitivity | 0.6522 | 0.7826 | 0.7391 | 0.6957 | 0.6087 | 0.6957 |
| | Specificity | 0.6333 | 0.4667 | 0.7000 | 0.5333 | 0.7667 | 0.6200 |
| | PPV | 0.5769 | 0.5294 | 0.6538 | 0.5333 | 0.6667 | 0.5920 |
| | NPV | 0.7037 | 0.7368 | 0.7778 | 0.6957 | 0.7188 | 0.7265 |
| | F1-score | 0.4412 | 0.4615 | 0.5313 | 0.4324 | 0.4667 | 0.4666 |
| | AUC | 0.6696 | 0.6507 | 0.7717 | 0.6159 | 0.7043 | 0.6825 |
| K-Nearest Neighbor | Accuracy | 0.8491 | 0.8302 | 0.7736 | 0.7547 | 0.7170 | 0.7849 |
| | Sensitivity | 0.6522 | 0.6957 | 0.6087 | 0.6087 | 0.6522 | 0.6435 |
| | Specificity | 1.0000 | 0.9333 | 0.9000 | 0.8667 | 0.7667 | 0.8933 |
| | PPV | 1.0000 | 0.8889 | 0.8235 | 0.7778 | 0.6818 | 0.8344 |
| | NPV | 0.7895 | 0.8000 | 0.7500 | 0.7429 | 0.7419 | 0.7649 |
| | F1-score | 0.6522 | 0.6400 | 0.5385 | 0.5185 | 0.5000 | 0.5698 |
| | AUC | 0.8261 | 0.8145 | 0.7543 | 0.7377 | 0.7094 | 0.7684 |
| Naïve Bayes | Accuracy | 0.7358 | 0.7333 | 0.7170 | 0.7547 | 0.7547 | 0.7391 |
| | Sensitivity | 0.6087 | 0.7308 | 0.6087 | 0.6522 | 0.6522 | 0.6505 |
| | Specificity | 0.8333 | 0.7353 | 0.8000 | 0.8333 | 0.8333 | 0.8071 |
| | PPV | 0.7368 | 0.6786 | 0.7000 | 0.7500 | 0.7500 | 0.7231 |
| | NPV | 0.7353 | 0.7813 | 0.7273 | 0.7576 | 0.7576 | 0.7518 |
| | F1-score | 0.5000 | 0.5429 | 0.4828 | 0.5357 | 0.5357 | 0.5194 |
| | AUC | 0.8101 | 0.8066 | 0.8043 | 0.7674 | 0.8094 | 0.7996 |

**Table 5.** *Cont.*

| Classifier | Measure | Five-Fold Cross-Validation | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Fold-1** | **Fold-2** | **Fold-3** | **Fold-4** | **Fold-5** | **Mean** |
| Random Forest | Accuracy | 0.9057 | 0.8679 | **0.9245** | 0.7736 | 0.7925 | 0.8528 |
| | Sensitivity | 0.8696 | 0.9565 | 0.9565 | 0.7391 | 0.6522 | 0.8348 |
| | Specificity | 0.9333 | 0.8000 | 0.9000 | 0.8000 | 0.9000 | 0.8667 |
| | PPV | 0.9091 | 0.7857 | 0.8800 | 0.7391 | 0.8333 | 0.8295 |
| | NPV | 0.9032 | 0.9600 | 0.9643 | 0.8000 | 0.7714 | 0.8798 |
| | F1-score | 0.8000 | 0.7586 | 0.8462 | 0.5862 | 0.5769 | 0.7136 |
| | AUC | 0.9551 | 0.9261 | 0.9623 | 0.8087 | 0.8652 | 0.9035 |
| Support Vector Machine | Accuracy | **0.9623** | **0.9245** | 0.8868 | **0.8868** | **0.9245** | **0.9170** |
| | Sensitivity | 0.9565 | 0.8696 | 0.7391 | 0.9565 | 1.0000 | 0.9043 |
| | Specificity | 0.9667 | 0.9667 | 1.0000 | 0.8333 | 0.8667 | 0.9267 |
| | PPV | 0.9565 | 0.9524 | 1.0000 | 0.8148 | 0.8519 | 0.9151 |
| | NPV | 0.9667 | 0.9063 | 0.8333 | 0.9615 | 1.0000 | 0.9336 |
| | F1-score | 0.9167 | 0.8333 | 0.7391 | 0.7857 | 0.8519 | 0.8253 |
| | AUC | 0.9971 | 0.9899 | 0.9681 | 0.9652 | 0.9688 | 0.9778 |

The highest performance for accuracy regarding to network was highlighted in bold.

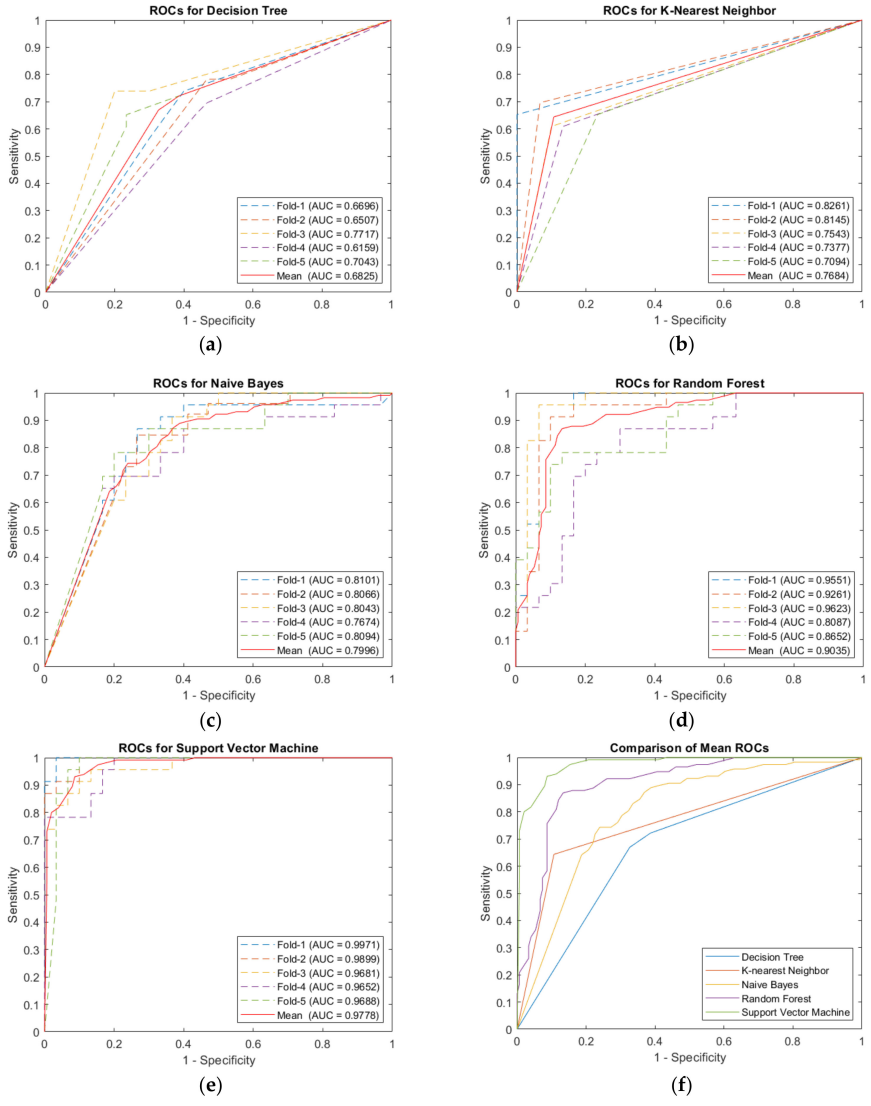



**Figure 6.** Comparison of the ROC curves for five classifiers. (**a**) Decision tree, (**b**) K-nearest neighbor, (**c**) Naïve Bayes, (**d**) Random Forest, (**e**) Support vector machine, and (**f**) Comparison of mean of receiver operating characteristic (ROC) curves for each classifier.

For a comprehensive assessment, the execution time of the proposed caries detection method is also computed. The experiments were conducted using the Matlab2020a environment in Windows 10. The main process was performed using a CPU core i7-9750 HF, supported by a GeForce GTX 2060 graphic card.

Each function process is carefully considered since they are factors used to determine the complexity of the method. Table 6 shows that the total process takes 13.79 s in total, and the most complex function, which is deep activated feature extraction, takes less than 10 s. Also, the geometric feature calculation was extracted smoothly in only 2.52 s. Based on these results, we consider the proposed method to perform considerably well and capable of wide implementation, even with low computer specifications. Based on the prediction and evaluation time, an image of a tooth after segmentation will take only 0.28 s (less than 1 s) to know if it is caries or non-carries. This is optimal for dentists, even in a large hospital with a huge number of patients.

**Table 6.** Total execution time for each function of the proposed system.

| Function Name | Time(s) |
|---|---|
| Load data | 0.37 |
| Deep activated features extraction | 9.99 |
| Geometric features extraction | 2.52 |
| Fusion features combination | 0.01 |
| Training classification model | 0.62 |
| Predict and evaluation | 0.28 |
| Total | 13.79 |

Lastly, the proposed method was compared with the previous state-of-the-art techniques (Table 7). Because the different methods were conducted in different datasets, each dataset's size and complexity will make a difference in performance. For a fair comparison, we detail the method and describe the difference and the advantages/disadvantages of each method. In addition, because some methods are not fully described but have been tested on other datasets in other papers, we provide a reference to the appropriate study and provide a description. The comparison table shows that [11,12] have a disappointing performance, whereas [13] performs much better; however, considering the accuracy of 90.00%, the sensitivity of 94.67, and specificity of 63.33%, we can see an imbalance in data as well as a low-performance result. Our proposed method achieved 92.67% specificity compared to the other methods, which is a 29.34% improvement, and the remaining sensitivity values are better than 90%. The 4.24% decrease in sensitivity value is a worthy exchange for the improvement of specificity.

**Table 7.** Performance comparison of the proposed method and with the previous methods.

| References | Method | Samples | ACC% | SEN% | SPEC% | PPV% | NPV% |
|---|---|---|---|---|---|---|---|
| [11,13] | • Autocorrelation and GLCM features<br>• SVM classification | 120 | 53.33 | 59.33 | 06.67 | 73.67 | 6.67 |
| [12,13] | • Autocorrelation coefficients matrix<br>• Neural network classification | 120 | 73.33 | 77.67 | 53.33 | 90.33 | 53.33 |
| [13] | • Multi-linear principal component analysis<br>• Non-linear programming with adaptive dragonfly algorithm<br>• Neural network classification | 120 | 90.00 | 94.67 | 63.33 | 91.00 | 63.33 |
| Proposed method | • Deep activated features<br>• Geometric features<br>• SVM classification | 533 | 91.70 | 90.43 | 92.67 | 91.51 | 93.36 |

## 5. Conclusions

In this article, we present a caries detection method using radiography images. Firstly, the radiography images were manually defined by dentists as either caries or non-caries. Later, in the feature extraction process, tooth images were used to extract the deep activated feature. The proper layer used to extract deep activated features from each deep pre-trained model was defined during experiments. Then, the geometric feature was also extracted and combined with deep activated features to build fusion features. The optimal features set was explored by a performance comparison between deep activated features fusion features. The set of geometric features was reduced to its minimum while retaining the optimal information. Next, we fed the fusion into classification models such as support vector machine (SVM), decision tree (DT), k-nearest neighbor (KNN), Naïve Bayes (NB), and random forest (RF) to classify between caries and non-caries images. Our proposed method has achieved 91.70%, 90.43%, and 92.67% for accuracy, sensitivity, and specificity, respectively. We improved the accuracy by 1.7%, from 90% to 91.70%, and the specificity by 29.34%, from 63.33% to 92.67%; the sensitivity was also good at 90.43% compared to previous state-of-the-art methods. The proposed method gives two key contributions: the first contribution is to find the best features set, which is the combination between deep activated features and geometric features, and then fit a proper classification model to describe the problem. The second contribution is to enhance the performance by improving the specificity measure factor. The performance of the deep activated feature is not proportional to the complexity or size of the model. The VGG16 deep activated feature is better than Xception, whereas the fusion result is the opposite. Our choice of the deep activated feature plays an important role; however, choosing analytically calculated features also contributed to the result equally. The finding of which deep activated feature features are compatible with analytically calculated features is more important than finding the best deep activated feature among all pre-trained models. While most research tries to build networks as deep as possible to improve the learning performance, our result proves that the performance is sometimes irrelevant to the network's depth. More importantly, the calculated feature's combination may play a key role in improving the performance and, therefore, unexchangeable for the pre-trained model's depth. The processing time, which takes 13.79 s for the whole experiment and 0.28 s for prediction, demonstrates that the method can be widely implemented in a low-tech computer for a trivial time-consuming. Nonetheless, despite the advantage compared with the previous state-of-the-art, this study's limitation is that we conducted the detection of caries based on the manually segmented teeth. In future work, we will improve our research as a fully automated system by performing automatic segmentation. We also have a great interest in extending our method to classify different caries stages by using three-dimensional approaches. By that, our system will be an adjunct tool for both experienced and junior dentists.

**Author Contributions:** T.H.B. and K.H. conceived and designed this study. T.H.B. performed the experiments, simulations, and original draft preparation of the paper. M.P.P. helped in experimenting and evaluating the results. K.H. reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the Tokai University (protocol code 19212 and on March 6$^{th}$, 2020).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Restriction applies to the availability of these data. The data were obtained from Shinjuku East Dental Office (the director is Makoto Kumon) and are available from authors with permission of Makoto Kumon or by sending a request to Makoto Kumon at: http://www.shinjukueast.com/doctor-staff/.

**Acknowledgments:** The authors express sincere gratitude to the Japan International Cooperation Agency (JICA) for financial support. The author also thanks to Makoto Kumon, director of Shinjuku East Dental Office, for supporting the dataset in this research.

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

## References

1. Oral health. World Health Organization 2020. Available online: https://www.who.int/health-topics/oral-health/ (accessed on 1 October 2020).
2. Cavities. From the MSD Manual Consumer Version (Known as the Merck Manual in the US and Canada and the MSD Manual in the rest of the world), edited by Robert Porter. Copyright (2021) by Merck Sharp & Dohme Corp., a Subsidiary of Merck & Co., Inc., Kenilworth, NJ. Available online: https://www.msdmanuals.com/en-jp/home/mouth-and-dental-disorders/tooth-disorders/cavities (accessed on 26 January 2021).
3. Mosquera-Lopez, C.; Agaian, S.; Velez-Hoyos, A.; Thompson, I. Computer-Aided Prostate Cancer Diagnosis from Digitized Histopathology: A Review on Texture-Based Systems. *IEEE Rev. Biomed. Eng.* **2015**, *8*, 98–113. [CrossRef] [PubMed]
4. Mansour, R.F. Evolutionary Computing Enriched Computer-Aided Diagnosis System for Diabetic Retinopathy: A Survey. *IEEE Rev. Biomed. Eng.* **2017**, *10*, 334–349. [CrossRef] [PubMed]
5. Sampathkumar, A.; Hughes, D.A.; Kirk, K.J.; Otten, W.; Longbottom, C. All-optical photoacoustic imaging and detection of early-stage dental caries. In Proceedings of the 2014 IEEE International Ultrasonics Symposium, Chicago, IL, USA, 3–6 September 2014; pp. 1269–1272.
6. Hughes, D.A.; Girkin, J.M.; Poland, S.; Longbottom, C.; Cochran, S. Focused ultrasound for early detection of tooth decay. In Proceedings of the 2009 IEEE International Ultrasonics Symposium, Rome, Italy, 20–23 September 2009; pp. 1–3.
7. Usenik, P.; Bürmen, M.; Fidler, A.; Pernuš, F.; Likar, B. Near-infrared hyperspectral imaging of water evaporation dynamics for early detection of incipient caries. *J. Dent.* **2014**, *42*, 1242–1247. [CrossRef] [PubMed]
8. Li, S.; Pang, Z.; Song, W.; Guo, Y.; You, W.; Hao, A.; Qin, H. Low-Shot Learning of Automatic Dental Plaque Segmentation Based on Local-to-Global Feature Fusion. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 3–7 April 2020; pp. 664–668.
9. Maslak, E.; Khudanov, B.; Krivtsova, D.; Tsoy, T. Application of Information Technologies and Quantitative Light-Induced Fluorescence for the Assessment of Early Caries Treatment Outcomes. In Proceedings of the 2019 12th International Conference on Developments in eSystems Engineering (DeSE), Kazan, Russia, 7–10 October 2019; pp. 912–917.
10. Angelino, K.; Edlund, D.A.; Shah, P. Near-Infrared Imaging for Detecting Caries and Structural Deformities in Teeth. *IEEE J Transl. Eng. Health Med.* **2017**, *5*, 2300107. [CrossRef] [PubMed]
11. Li, W.; Kuang, W.; Li, Y.; Li, Y.; Ye, W. Clinical X-Ray Image Based Tooth Decay Diagnosis using SVM. In Proceedings of the 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, China, 19–22 August 2007; pp. 1616–1619.
12. Yu, Y.; Li, Y.; Li, Y.-J.; Wang, J.-M.; Lin, D.-H.; Ye, W.-P. Tooth Decay Diagnosis using Back Propagation Neural Network. In Proceedings of the 2006 IEEE International Conference on Machine Learning and Cybernetics, Dalian, China, 13–16 August 2006; pp. 3956–3959.
13. Patil, S.; Kulkarni, V.; Bhise, A. Intelligent system with dragonfly optimisation for caries detection. *IET Image Process.* **2019**, *13*, 429–439. [CrossRef]
14. Pan, W.-T. A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example. *Knowl. Based Syst.* **2012**, *26*, 69–74. [CrossRef]
15. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]
16. Loog, M.; Duin, R.P.W. Linear dimensionality reduction via a heteroscedastic extension of LDA: The Chernoff criterion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 732–739. [PubMed]
17. Lazcano, R.; Madroñal, D.; Salvador, R.; Desnos, K.; Pelcat, M.; Guerra, R.; Fabelo, H.; Ortega, S.; Lopez, S.; Callico, G.M.; et al. Porting a PCA-based hyperspectral image dimensionality reduction algorithm for brain cancer detection on a manycore architecture. *J. Syst. Archit.* **2017**, *77*, 101–111. [CrossRef]
18. Montefusco-Siegmund, R.; Maldonado, P.E.; Devia, C. Effects of ocular artifact removal through ICA decomposition on EEG phase. In Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), San Diego, CA, USA, 6–8 November 2013; pp. 1374–1377.
19. Jiao, Z.; Gao, X.; Wang, Y.; Li, J.; Xu, H. Deep Convolutional Neural Networks for mental load classification based on EEG data. *Pattern Recognit.* **2018**, *76*, 582–595. [CrossRef]
20. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Proceedings of the Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1, Lake Tahoe, NV, USA, 3–6 December 2012*; pp. 1097–1105.
21. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
22. Tiulpin, A.; Thevenot, J.; Rahtu, E.; Lehenkari, P.; Saarakkala, S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. *Sci. Rep.* **2018**, *8*, 1727. [CrossRef] [PubMed]

23. Stuhlsatz, A.; Lippel, J.; Zielke, T. Feature Extraction with Deep Neural Networks by a Generalized Discriminant Analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 596–608. [CrossRef] [PubMed]
24. Szegedy, C.; Wei, L.; Yangqing, J.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9.
25. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
27. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
28. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural Features for Image Classification. *IEEE Trans. Syst. ManCybern.* **1973**, *SMC-3*, 610–621. [CrossRef]
29. Soh, L.; Tsatsoulis, C. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 780–795. [CrossRef]
30. Clausi, D.A. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Can. J. Remote Sens.* **2002**, *28*, 45–62. [CrossRef]
31. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
32. Guo, G.; Wang, H.; Bell, D.; Bi, Y. KNN Model-Based Approach in Classification. In *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*; Springer: Berlin/Heidelberg, Germany, 2004.
33. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; CRC Press: New York, NY, USA, 1984.
34. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning, Second Edition*; Springer: New York, NY, USA, 2008.
35. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]