SJMI(2022) 01; 21-41 ISSN : 2635-4608

DOI http://dx.doi.org/10.31916/sjmi2022-01-02

Grading Classification of Tongue Cancer in Oral Images Using CNN Transfer Learning

Chung Hwan Kang, Tae Soo Lee

Received: 7 March 2022 / Accepted: 30 July 2022 / Published online: 30 December 2022

Abstract– Oral cancer is one of the top 10 diseases in cancer incidence worldwide and its prognosis is not good as early diagnosis is not made and so the survival rate from it has not been improved greatly. Oral cancer occurs most commonly in tongue, floor of mouth, and lower lip and 5-year survival rate is very low at 50% and if diagnosed early, the average time of survival becomes longer. Therefore, to improve the survival rate of oral cancer patients, early diagnosis and discovery and patient habit improvement including smoking are very important. Oral cancer can be diagnosed with imaging equipment like CT and MRI or through invasive biopsy, but it is not easy to make an approach and generally difficult to discover early. Therefore, to identify if early diagnosis would be possible using the intraoral images obtained through ordinary camera, the reliability of tongue cancer grade classification was evaluated by applying deep learning analysis. The collected images of oral cavity were classified into four groups: normal, inflammatory lesion, pre-cancerous lesion, and malignant tumor depending on clinical importance and three (3) types of image pretreatment were applied. In case of deep neural network, Inception-ResNet-v2 was used and 70%(1,307 images) of the pretreated images were used as training data and after image enhancement processing, transfer-learned with SGDM technique. As test data, the remaining 30% data (561 images) were used.

In confusion matrix, it appeared that sensitivity was 90.37%, specificity 95.03%, accuracy 93.21%, positive predictability 92.06%, and negative predictability 93.93%, and in ROC curve, AUC was normal because normal was 0.9781, inflammatory lesion 0.9102, precancerous lesion 0.8944, and malignant tumor 0.9688, and the classification performance of inflammatory lesion or pre-cancerous lesion was inferior to that of malignant tumor. The accuracy depending on pretreatment did not show a statistically significant difference in HE-1 and HE-2 compared to standard image. The performance of the diagnostic classifier applying deep learning into intraoral images showed accuracy at the average rate of 82.9% and so showed more excellent performance than the precedent studies using different RGB images. Deep learning study using such RGB images tends to be rapidly commercialized as it has disclosed big data whose images were easy to obtain and that were preclassified and so noninvasive data can be easily obtained unlike biopsy or blood test. Intraoral image does not have preclassified and disclosed big data yet, but have a distortion in an artifact or lesion in obtaining an image and so if we can standardize the image acquisition and organize big data through correct preclassification in cooperation with clinicians although this is limited, it will be helpful for early diagnosis of tongue cancer.

Key word: Tongue cancer, Deep neural network, Classification, Transfer Learning

Chung Hwan Kang
Researcher in Konkuk University Medical Center

TaeSooLee(✉) ***corresponding author***
Dept. of Biomedical Engineering, Graduate School, Chungbuk National University, Cheongju, Korea
e-mail : tslee@chungbuk.ac.kr

I. Introduction

Oral cancer is one of the ten most common diseases worldwide. Oral cancer has no biomarkers and is difficult to detect clinically, so the treatment prognosis is poor and treatment costs are high.In order to solve these difficulties, studies for diagnosis and treatment have been continued for the past few years, but the survival rate has not improved significantly. Oral cancer is a malignant neoplasm that occurs in the lips or oral cavity. In the field of dentistry, it is traditionally defined as Oral Sqamous Cell Carcinoma (OSCC) because 90% of cancers originate from squamous cells[1]. The remaining 10% are characterized for different levels of differentiation and lymph node metastases[2]. Oral cancer is two to three times more common in men than women in most racial groups. Cancer of the oral cavity and pharynx is known as the 6th most common cancer in the world in worldwide reports[3]. According to the latest report from the International Agency for Research on Cancer (IARC) on oral cancer (ICD-10 code C00-08: lip, oral cavity), it has been found to be located in the lips, tongue, gums, floor of the mouth, parotid glands, etc. The most common sites presenting oral cancer are the tongue (ventral-marginal, 40%), floor of the mouth (30%), and lower lip[4]. Oral cancer has a worldwide incidence of over 300,000 diagnoses and an annual mortality rate of approximately 145,000. Age-standardised incidence rates are high in developed countries, but are characterized by high mortality rates in developing countries[5]. In clinical examination, most oral cancers are diagnosed in a malignant stage. The reason for this is that it is often diagnosed late because there is little pain in the early stages. Therefore, the possibility of survival is reported to be low[6]. The 5-year survival rate for oral cancer is very low at 50%[7]. Looking at the survival rate according to TNM (tumor, node and metastasis) classification, the 5-year survival rates of 66.2% for T1, 57.9% for T2, 43.0% for T3, and 22.2% for T4 were reported for each T classification[8]. In addition, it is known that the average survival time decreases from T1 to T4, and that the survival rate decreases as the size of the primary tumor increases[9]. Looking at the survival rates related to drinking and smoking, non-drinkers and non-smokers showed a survival rate of 73.1%, patients who only smoked 61.4%, and finally, patients who drank and smoked at the same time showed a survival rate of 41.4%. Drinking and smoking are the most dangerous factors that can cause oral cancer, and it has already been reported that drinking and smoking are even more dangerous[10]. Therefore, in order to improve the survival rate of oral cancer patients, early diagnosis and detection of oral cancer and improvement of patient habits such as smoking and drinking are very important. In general, lesions in the oral cavity can be confirmed visually and by palpation, and most oral cancers can be diagnosed histologically as squamous cell carcinoma that occurs on the mucosal surface. Imaging tests not only confirm the lesion, but also help determine the treatment policy and determine the treatment effect by diagnosing whether the lesion has spread to the submucosa or deep tissues, and whether or not lymph node invasion has occurred. CT and MRI are helpful for the imaging diagnosis of oral cancer. In CT images, there are many areas that are difficult to evaluate due to artifacts caused by teeth or mandibular foramen, and most lesions are located within the soft tissue plane with similar shades. Therefore, there is little contrast for differentiation. Because many cases of oral cancer are invasive lesions, it is difficult to determine the size change by left-right comparison because there is no elevation on the mucosal surface. In particular, in the case of the tongue composed of muscles, it is difficult to detect internal changes with

CT, and only the relatively fatty lingual septum is distinguished by low contrast. Comparatively, MRI is advantageous in identifying the space of the lesion because it can produce multiplanar images, and it is advantageous because the tissue contrast is superior to that of CT and a small amount of fat between the muscle layers can be seen relatively sensitively. In addition, since contrast enhancement can be observed, contrast-enhanced lesions can be easily identified by comparing images before and after contrast enhancement[11]. With the 4th industrial revolution, interest in artificial intelligence is increasing, and research is continuing to utilize it in various fields by convergence with existing information and communication technologies. Innovative changes using artificial intelligence have led to the development of the internet of things, big data, cloud computer, and 3D printing technology, and continue to expand the field of application. The development of artificial intelligence improves work efficiency and reduces human errors, so the form of work that humans have been working on is being replaced by digital labor instead of manual work[12]. The medical field is no exception, and over the past 20 years, Computer Aided Diagnosis (CAD) has developed significantly due to increased access to digital medical data, improved computing power, and advances in artificial intelligence[13]. This development has been adopted not as a replacement for medical professionals, but rather as a tool to increase the accuracy of diagnosis. Not only specific medical fields, but gradually various medical fields are showing a movement to increase the reliability of diagnosis by using artificial intelligence. Oral cancer can be diagnosed using imaging equipment such as CT and MRI or through invasive biopsy, but it is not easy to access and is generally difficult to detect in an early stage. In this study, it was judged that intraoral images can be collected relatively easily when intraoral images are taken using a general camera, as well as accessibility and usability. In order to minimize the effect on structures such as teeth and gums from oral images, an experiment was conducted using images of the side of the tongue, which is the main site of tongue cancer, which is the most frequently occurring among oral cancers. Currently, research on the possibility of using deep learning models using intraoral images is insufficient. Therefore, in this paper, we tried to evaluate the reliability of tongue cancer grading by applying CNN transfer learning to intraoral images.

II. Theoretical background

1. History of artificial intelligence

Artificial intelligence is based on the idea that “human or animal intelligence can be expressed in detail and accurately enough to be copied by a computer” and is defined in various ways like the definition of intelligence. Artificial intelligence was established as a field of research in a workshop held at Dartmouth College in the summer of 1956[14]. Computers began to solve problems in algebra, prove theorems in geometry, and understand the structure of language. Artificial intelligence research in the early 1950s and 1960s achieved remarkable results in the areas of theorem proofs and games, but after that, disappointment and decline due to excessive expectations and the development of new models and theories were repeated[15]. In the 1970s and 1980s, research on expert systems was active. Since the rediscovery of the backpropagation algorithm in the mid-1980s, research on artificial neural network (ANN) models have become active. Artificial intelligence research in the 1990s utilized methods

from various fields such as statistics, information theory, and optimization, and was equipped with a solid theoretical foundation such as learning theory[16]. In the early 21st century, with the construction of big data and the development of high-performance computers, artificial intelligence technology was successfully applied to the economic field. As a result, the AI-related market exceeded $8 billion in 2016[17]. As convolutional neural networks (CNNs) and recurrent neural networks (RNNs), one of the artificial intelligence technologies, have been developed, research on video, text, and voice recognition has been actively conducted[18].

2. Machine learning

Machine learning (ML) is a field of inquiry dedicated to understanding and building how to 'learn', that is, how data can be leveraged to improve the performance of some set of tasks[19]. It is considered part of artificial intelligence. Machine learning algorithms build models based on sample data, called training data, to make predictions or decisions without being explicitly programmed[20.]. Machine learning algorithms are used in a variety of applications, such as healthcare, email filtering, speech recognition, agriculture, and computer vision, where it is difficult or impossible to develop existing algorithms to perform required tasks[21]. A subset of machine learning is closely related to computational statistics, which focuses on using computers to make predictions, but not all machine learning is statistical learning. Mathematical optimization research provides methods, theories, and applications to the field of machine learning. Some implementations of machine learning use data and neural networks in a way that mimics the operation of a biological brain[22] .

3. Classification of machine learning systems

(1) Supervised learning

Supervised learning (SL) is a machine learning paradigm for problems where available data consists of labeled examples. That is, each data point contains a feature (covariate) and an associated label. The goal of a supervised learning algorithm is to learn a function that maps feature vectors (inputs) to labels (outputs) based on example input-output pairs[23]. Infer functions from labeled training data consisting of a set of training examples[24]. In supervised learning, each example is a pair consisting of an input object (usually a vector) and a desired output value (also called a supervised signal). Supervised learning algorithms analyze training data and generate inferred functions that can be used to map new examples. In an optimal scenario, the algorithm can correctly determine the class label for unseen instances. This requires that the learning algorithm generalize to unseen situations in the training data in a "reasonable" way. This statistical quality of an algorithm is measured through the so-called generalization error.

(2) Unsupervised learning

Unsupervised learning is a type of algorithm that learns patterns from untagged data. The hope is to enable machines to succinctly represent their world through imitation, an important way people learn, and then generate imaginative content from it. Unlike supervised learning, where experts tag data, unsupervised methods tagged as “balls” or “fish” convert patterns into probability densities[25] or combinations of neural function preferences encoded in weights and activations in the machine. Indicates the self-organization that captures.

(3) Reinfocement learning

Reinforcement learning (RL) is an area of machine learning concerned with how an intelligent agent

should take actions in its environment in order to maximize the concept of cumulative reward. Reinforcement learning is one of the three basic machine learning paradigms, along with supervised learning and unsupervised learning. Reinforcement learning differs from supervised learning in that it does not require you to present labeled input/output pairs and does not require explicit workarounds to be corrected. Instead, it focuses on finding a balance between exploration and exploitation[26]. The environment is usually described in the form of a Markov decision process (MDP), as many reinforcement learning algorithms for this context use dynamic programming techniques[27]. The main difference between classical dynamic programming methods and reinforcement learning algorithms is that the latter do not assume knowledge of the exact mathematical model of the MDP and target large-scale MDPs for which exact methods are not feasible.

4. Artificial Neural Networks(ANNs)

An artificial neural network (ANN), commonly referred to simply as a neural network (NN)[28], is a computing system inspired by the biological neural networks that make up the brains of animals[29]. ANNs are based on collections of connected units, or nodes, called artificial neurons that loosely model neurons in the biological brain. Each connection, like a synapse in a biological brain, can transmit signals to other neurons. Artificial neurons can receive signals, then process them and send signals to connected neurons. The "signal" of the connection is real, and the output of each neuron is computed as some non-linear function of the sum of the inputs. Connections are called edges. Neurons and edges usually have weights that adjust as learning progresses. Weight increases or decreases signal strength when connected. A neuron may have a threshold so that a signal is sent only when the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. A signal travels from the first layer (the input layer) to the last layer (the output layer), and it is possible after passing through the layers several times.

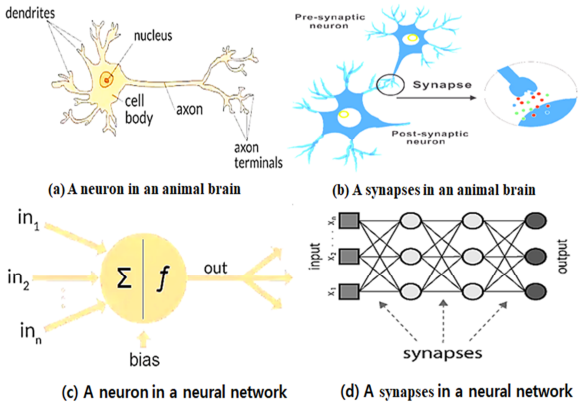


Figure 1. Artificial neural networks

Just as neurons become nerves when gathered, artificial neurons become artificial neural networks when aggregated. The general algorithmic structure of a neural network consists of a number of hidden layers between an input layer (data to be trained) and an output layer (results from a trained model). The hidden layer plays the role of receiving the weights of the previous input layer or other hidden layers and passing them to the next layer. When data information is handed over from the input layer, weights are calculated by going through hidden layers for each feature of the data, and these values are finally moved to the output layer to determine the final value (Zell, 1997). Just as a signal is transmitted when a neuron exceeds a certain threshold value of a stimulus, it can be transmitted to the next neural network only when a value above a threshold value is derived through the role of various functions.

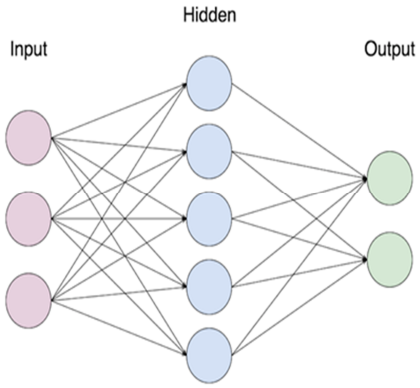


Figure 2. Examples of artificial neural networks

5. Deep learning

Deep learning (also known as deep structured learning) is part of a broader family of machine learning methods based on artificial neural networks with representation learning. Learning can be supervised, semi-supervised or unsupervised[30]. The adjective "deep" in deep learning refers to the use of multiple layers in the network. Deep learning is a modern variation that is concerned with an unbounded number of layers of bounded size, which permits practical application and optimized implementation while retaining theoretical universality under mild conditions. Deep learning is basically designed based on the structure of an artificial neural network, but as shown in Figure 3, it has developed into a simpler and better form by supplementing the overfitting phenomenon and slow learning time, which are disadvantages of existing artificial neural networks. There are many structures that can compose deep learning, and most of them are derived from several representative structures, and the structure of the algorithm can change depending on the type of data to be analyzed. Representatively, there are CNNs that show good performance in video and audio fields, and RNNs that show good performance in character recognition such as handwriting recognition.

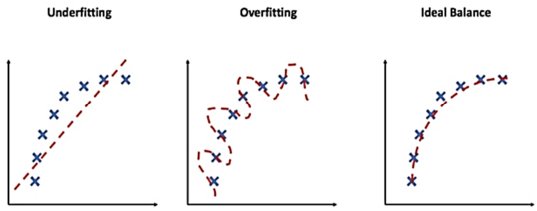


Figure 3. Understanding overfitting and overfitting

6. Overfitting

In machine learning, overfitting means learning the training data too well. In general, training data is usually a subset of real data. Therefore, as shown in the graph in Figure 4, there may be a point where the error decreases for the training data but increases for the actual data. From this point of view, overfitting is a phenomenon in which errors for actual data increase due to excessive learning on the training data. For example, a phenomenon similar to overfitting is a phenomenon in which a person who has learned cat characteristics by seeing a yellow cat sees a black or white cat and fails to recognize it as a cat.

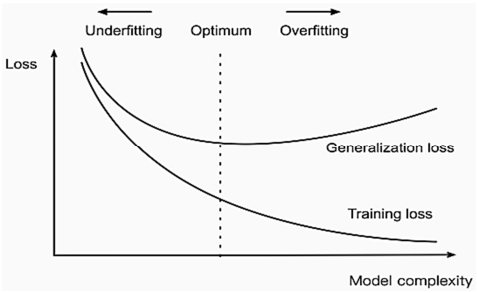


Figure 4. Impact of model complexity on underfitting and overfitting

7. Convolutional Neural Network: CNN

(1) CNN structure and understanding

In deep learning, convolutional neural networks (CNNs or ConvNets) are a class of artificial neural networks (ANNs) most commonly applied to analyze visual images[31] CNNs, also known as Shift Invariant or Space Invariant Artificial Neural Networks (SIANN), are based on a shared-weight

architecture of convolutional kernels or filters that slide along the input features and give a transform-equivalent response known as a feature map[32].

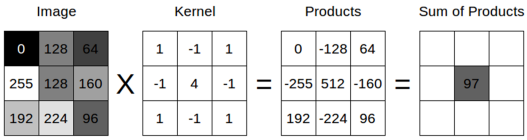


Figure 5. Convolution principle

Counterintuitively, most convolutional neural networks are not transformation invariant due to the downsampling operation applied to the input[33]. They have applications in image and video recognition, recommender systems[34] image classification, image segmentation, medical image analysis, natural language processing[35] brain-computer interfaces, and financial time series[36]. A CNN is a normalized version of a multilayer perceptron. A multilayer perceptron usually refers to a fully connected network. That is, each neuron in one layer connects to every neuron in the next layer. The "perfect connectivity" of these networks tends to overfit the data. So, on the scale of connectivity and complexity, CNNs are at the lower end. Convolutional networks are inspired by biological processes[37] in that the connection patterns between neurons are similar to the organization of animal visual cortex. Individual cortical neurons respond to stimuli only in a limited area of the visual field known as the receptive field. The receptive fields of different neurons partially overlap to cover the entire field of view. CNNs use relatively little preprocessing compared to other image classification algorithms. This means that networks learn how to optimize filters (or kernels) through automated learning, whereas in traditional algorithms these filters are designed by hand. This independence from prior knowledge of feature extraction and human intervention is a major advantage. CNN shows a different aspect from conventional neural networks in the way it processes images, and it is characterized by enabling more complex calculations using specific filters rather than extracting features of images through simple calculations. The image we see is composed of a pixel structure, and if it is a color image instead of a horizontal and vertical position information, and a gray image, it becomes 3D data by adding color information (Red, Green, Blue, RGB).

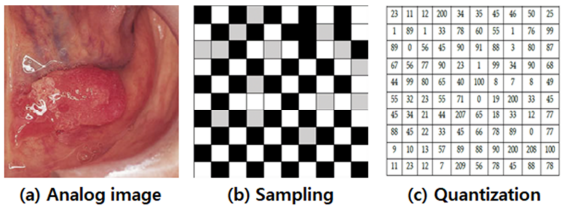


Figure 6. Creation of digital images

In order to extract the features of this three-dimensional information well, the CNN structure consists of a convolutional layer, a max pooling layer, and finally a fully connected layer. Figure 7 the backpropagation method is used to update the weight by forwarding the error of the output value in the reverse direction.

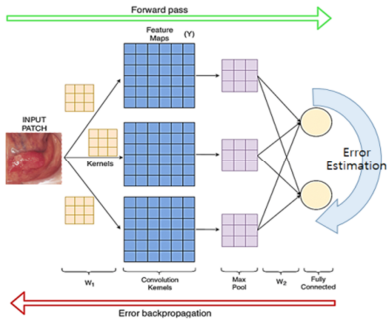


Figure 7. Backpropagation

(2) Converlutional layer

To process 3D image data, a kernel of a specific size is used. The kernel is a two-dimensional concept multiplied by the X-axis and the Y-axis. For example,

if the kernel size is defined as 3 × 3, this kernel has a weight (or parameter) for each cell and moves from the upper left of the image to the side one by one, It goes through an arithmetic process. Filter size is the definition of how many times an operation process through such a kernel size is performed. A filter is a three-dimensional concept on the Z axis, which is different from kernel size (Figure 8).

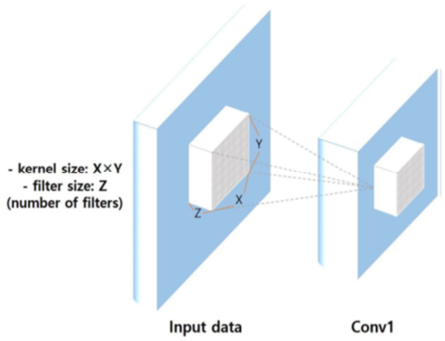


Figure 8. Difference Between Kernel Size and Filter Size

As shown in <Figure 9>, the values obtained through this operation represent the characteristics of the image and are called features. That is the convolutional layer proceeds with the kernel and filter operation of the data information transmitted earlier and passes the value to the next layer. These values represent the characteristics of the image and facilitate image classification.

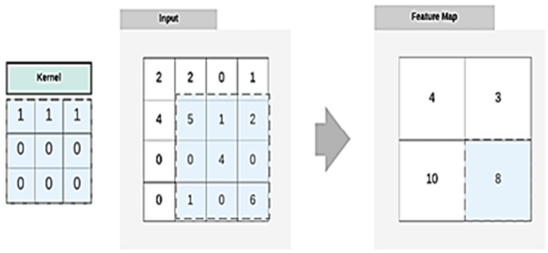


Figure 9. Feature map

(3) Pooling layer

The pooling layer serves to reduce the amount of data information by sampling some of the most meaningful data among the previously delivered data (Liu et al., 2017). Data sampling methods include minimum pooling, average pooling, and maximum pooling. Maximum pooling, which can best reflect the characteristics of the corresponding image, is generally used. Assuming that there is one 4×4 size data in Figure 10, if it is assumed that only the largest data value is extracted by moving a 2×2 size filter by 2 spaces along the x-axis and y-axis, the most characteristic data can be extracted and the amount of data can be reduced by 1/4.

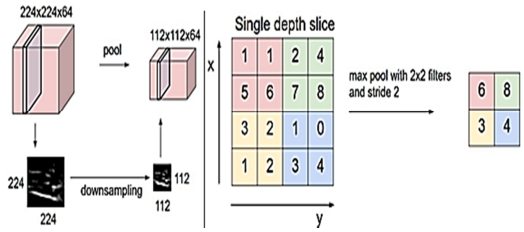


Figure 10. Example of maximum Pooling

(4) Fully connected layer

The fully connected layer is a dense layer that finally represents the result value through the convolutional layer and the pooling layer. In this layer, the probability of having an n-dimensional vector value is represented, and the number of n is equal to the number of classes to be checked (Figure 11). If we classify 10 numbers from 1 to 10, the final fully connected layer will generate 10 result data. The data generated in this way is output using an activation function in order to be converted into a probability.

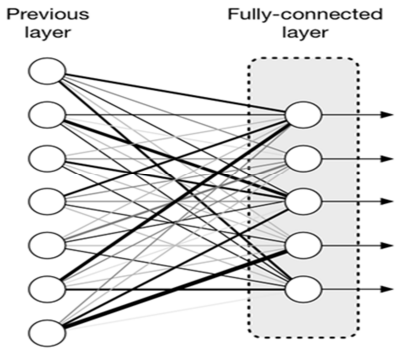


Figure 11. Fully connected layer

8. Terminology related to deep learning technology

(1) Cost function

The cost function is a function that explains how well the designed model is applied to given data and the error is small (W et al., 2016). A smaller value means that a more suitable model is designed. In the regression model, the method of least squares minimizes the sum of residual squares to represent the appropriate dependent variable value by the explanatory variable, or the mean squared error that represents the average of the residual squares Using the method, find the straight line that can best be explained in the observed data. Similarly, in machine learning, various methods are used to reduce the value of the cost function. A typical example is cross entropy. The cross-entropy for a given set of distributions (p, q) is defined as Equation (1).

$$CE = -\sum_{x} p(x) \log q(x) \quad \text{---------(1)}$$

Since the probability q value for a distribution p-value is a real number from 0 to 1, taking the logarithm gives a negative number. To compensate for this negative value, add a minus value to the logarithm to produce a positive real number. As the probability q value approaches 1, the log value converges to 0, so entropy (degree of uncertainty) decreases. A sample is drawn from a probability distribution, and the cost function is reduced by minimizing the cross-entropy between this distribution and the target distribution to find more suitable weights in the next iteration.

(2) Optimizer

An optimizer is an algorithm that updates weights for errors.The most representative is gradient descent and stochastic gradient descent (SGD). As shown in Figure 12, the gradient descent method decreases the weight when the derivative of the cost function, that is, the slope is a positive value, and increases the weight when the slope is a negative value, as shown in Figure 12.

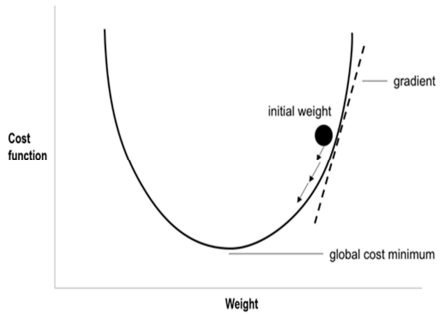


Figure 12. Optimization by Gradient Descent Algorithm

As the slope value converges to 0, the value of the cost function also decreases. The downside of gradient descent is that the amount of computation becomes enormous because the gradient is calculated at every step. To compensate for this shortcoming, SGD is mainly used. Since SGD calculates the gradient several times in the same time by calculating the gradient only from some data instead of calculating all the data at each step, the weight can be updated at a high speed. However, even though there is a global minimum for error, there may be a problem that it stays at the local minimum and is not updated anymore (see Figure 13).

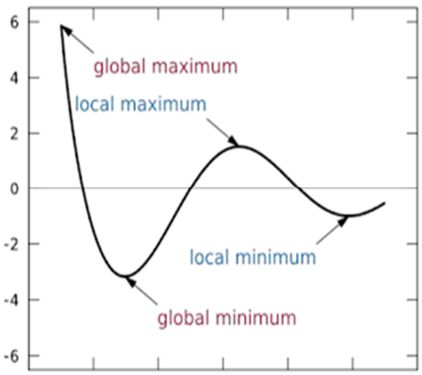


Figure 13. Global Minimum and Local Minimum

To solve this problem, the momentum technique is

used. Momentum has the meaning of 'acceleration' in the dictionary, and as a method of changing the speed in the direction of updating the weight, it is accelerated and updated in the direction of the current slope. That is, when the error range is large, the update interval is further widened to accelerate and update to the optimal value, thereby reducing the probability of being at the local lowest point.

(3) Activation function

The values derived through the convolutional layer use an activation function to further highlight and refine the characteristics of the data. Relatively unnecessary information is reduced, and weights extracted from key input data that reduce the cost function are increased. Typically, three activation functions are used to send output values to the next layer.

① **Sigmoid function**

It represents the data value as a value between 0 and 1 and makes the negative input value close to 0. At the end of each layer, the shape of the value is transformed or normalized under the influence of this function. Refer to <Figure 14>

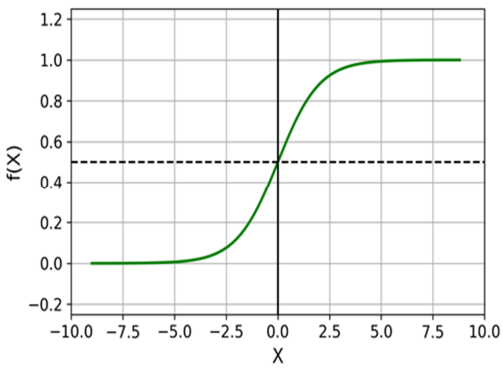


Figure 14. Sigmoid function

② **Rectified Linear Unit, ReLU function**

As a representatively used activation function, it has the characteristic of turning all negative values to 0 (Agarap, 2019). If it has a positive value even slightly, it is calculated to have a positive slope to maintain this characteristic. In the CNN model using the backpropagation method, the value of the gradient disappears as the neural network of the model deepens due to the sigmoid function. This is called the vanishing gradient problem, and the ReLU function effectively prevents this (see <Figure 15>).

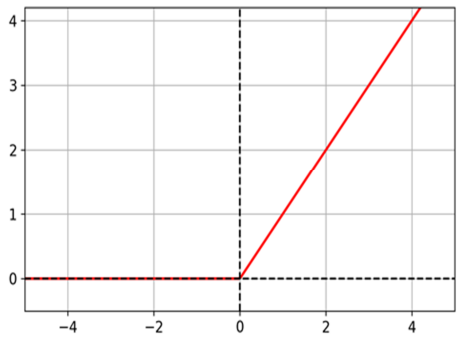


Figure 15. ReLU function

③ Softmax function

For n variables, it is a function that always makes the sum of the variables equal to 1 while maintaining the ratio of the magnitude of this number. It is used when converting to a probability value for each variable and is mainly used when dividing into three or more categories in the last layer (see <Figure 16>). It is a function that calculates the probability of each possible class in a multi-class classification model, and the sum of the probabilities is exactly 1.0. For example, Softmax can determine the probability that a particular image is a dog with a probability of 0.9, the probability that it is a cat with a probability of 0.08, and the probability that a particular image is a horse with a probability of 0.02.

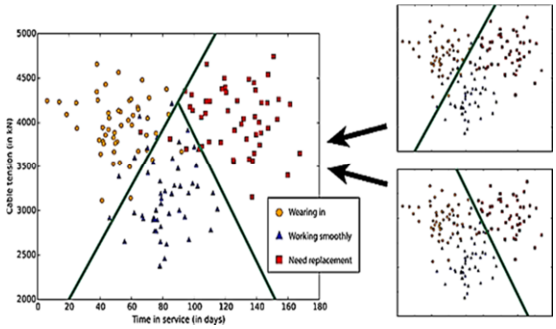


Figure 16. Polynomial Classification and Softmax

8. Transfer learning

Transfer learning means using some of the abilities of a neural network learning in a specific field to learn a neural network used in a similar or completely new field. Taking image classification as an example, the front end of neural networks such as Resnet or VGG is composed of CNN layers. CNN can be divided into a part that extracts the features of an image and a part that classifies a class. The feature extraction area is composed of multiple layers of convolution layers and pooling layers. The convolution layer is an essential element that reflects the activation function after applying a filter to the input data.

The pooling layer located after the convolution layer is an optional layer. At the end of the CNN, a fully connected layer for image classification is added.

Between the part that extracts the features of the image and the part that classifies the image, there is a flattened layer that creates an array of image-type data. The ability of neural networks to extract features from these images can be used in other fields as well. That is, using the feature extraction capability of the high-performance Resnet or VGG neural network learned through tens of thousands to tens of millions of images as it is, and changing only the affine layer as the last output layer to relearn only this changed layer . It is transfer learning. Transfer learning is effective even when the number of training data is small, has a fast learning speed, and provides much higher accuracy than learning without transfer learning.

9. Transfer learning application model

Table 1 presents pre-trained networks and properties in ImageNet.

Table 1. Pretrained networks on ImageNet

Network	year	depth	paremeter	Input image size
AlexNet	2012	8	61×10^6	227×227
VGG	2014	19	138×10^6	227×227
GoogLeNet	2015	22	7×10^6	227×227
Inception-v3	2015	48	23.9×10^6	299×299
Inception-ResNet-v2	2016	164	55.9×10^6	299×299

(1) AlexNet

CNN, which is attracting attention as the two major mountain ranges of deep learning models along with RNN, is basically based on the structure proposed by Yann LeCun in 1989 (see <Figure 17>). In the 2012 ILSVRC (ImageNet Large-Scale Visual Recognition Challenge) competition, which can be called the 'Olympics' in the field of computer vision, Professor Jeffrey Hinton's AlexNet recorded a top-5 error of 15.4%, taking second place (26.2%). He beat him by a margin and took first place. Here, the top-5 error refers to the error rate when there is no correct answer among the top 5 categories predicted by the model. At the time, the ILSVRC dataset was a 1000-category prediction problem, and thanks to AlexNet, deep learning, especially CNN, gained attention.

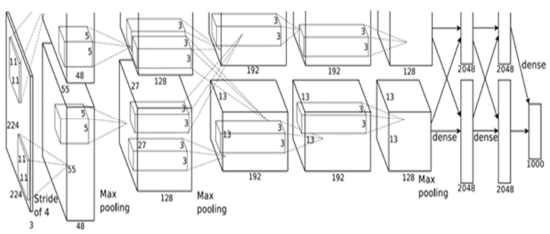


Figure 17. AlnexNet

(2) GoogLeNet

After AlexNet, VGGNet in 2014, and GoogLeNet in 2015 continued attempts to improve performance by stacking layers deeper. GoogLeNet has a more complicated structure than VGGNet, so it was not widely used, but it attracted attention in terms of architecture. In general, one convolutional filter is used in one convolutional layer. However, GoogleNet researchers have presented a creative idea that individual layers can be extended thickly by introducing various types of filters or pooling even in one layer. And the structure proposed by them is the inception module, and it is the 1×1 convolutional layer that has received particular attention (Szegedy et al., 2014). For example, if the number of dimensions of the current layer input data image is 100 × 100 × 60 and 20 1 × 1 convolutional layers are used, the number of dimensions of the data is reduced to 100 × 100 × 20 and 60 channels (dimensions) It can also be seen that one pixel reaching 20 is linearly transformed and dimensionally reduced into a 20-dimensional feature space.

(3) Inception v2 & v3

Inception v2 and v3 are presented in the same paper as the developed form of the previous GoogLeNet. In the paper, two problems were defined and solutions were presented (Szegedy et al., 2016). The first problem is the representational bottleneck, which is a phenomenon in which the amount of information is greatly reduced when the dimension is excessively reduced in the neural network net. The second problem is factorization, which improves the kernel used in the existing convolutional operation. This can further reduce computational complexity. The structure of Inception v2 and Inception v3 is almost the same. Inception v3 adds four skills to the same structure. First, replace the 7 × 7 convolutional operations of the stem layer with three 3 × 3 operations, secondly use "'RMSProp" as an optimizer, and thirdly use an auxiliary classifier Batch normalization was applied using only one. Finally, to prevent overfitting, normalization was performed on the answer data using label smoothing.

(4) Inception-ResNet-v2

Google announced a new convolutional model, Inception-ResNet -v2, through its research blog. As a result of ILSVRC test of this model, which absorbed the advantages of ResNet into Inception v3 model, as shown in Table 2, the previous model's record was updated, but the memory and computation amount increased by almost twice compared to Inception v3 .

Table 2. Inception-ResNet-v2 model accuracy (Google AI Blog, 2016)

Model	Top-1 Accuracy	Top-5 Accuracy
Inception-ResNet-v2	80.4%	95.3%
Inception v3	78.0%	93.9%
ResNet 152	76.8%	93.2%
ResNet V2 200	79.9%	95.2%

In Figure 18, (a) is a pictorial representation of the entire network, and (b) is a simplified representation of overlapping parts.

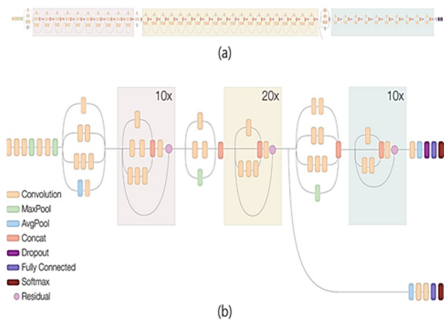


Figure 18. Inception Resnet V2 network (Google AI Blog, 2016)

III. Research methods

1. Computing Environment and Dataset

The image data used in this study were images acquired through Internet searches and oral images from the Department of Oral and Maxillofacial Surgery at S University and C University Dental Hospitals. Among the oral images, lateral images of the tongue were selected and divided into four categories (normal, inflammatory lesions, precancerous lesions, and malignant tumors). According to clinical importance, class 0 (normal) is 1,141 (61%), class 1 (inflammatory lesion) is 218 (12%), class 2 (precancerous lesion) is 140 (7%), class 3 (malignant tumor) was classified as 369 (20%). The computing environment for tongue cancer classification is Windows 10, Matlab 2019b's deep learning toolbox was installed to construct a deep neural network, and a GTX1650 GPU was used for parallel processing.

2. Image pre-processing

After preprocessing the collected images to fit the deep learning model, the performance of each model was compared and analyzed by dividing them into a training set and a test set. Intraoral images were taken at various resolutions and consisted of 1.3M to 3.5M pixels, which were normalized to images with a resolution of 512 × 512. In the image acquisition stage, preprocessing was performed to make the data less sensitive to environmental changes such as lighting. Two methods were applied for preprocessing. First, after color conversion of RGB color into CIE 1976 L*a*b* color, Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to the brightness channel (L). After applying, it was reconverted to RGB color. Second, the RGB color image was converted into a new RGB color image by applying CLAHE to each of the R, G, and B channels. After each method was applied, a median filter was commonly applied to each RGB channel. In CLAHE, the contrast limit was set to 0.05, the window size was set to 8 × 8, and the median filter kernel to remove artifacts that could occur when CLAHE was applied was set to 3 × 3.

3. Deep Neural Networks

The deep neural network used Inception-ResNet-v2, which accepts 299×299 input data with the highest resolution among pre-trained DNNs and has the deepest convolutional layer with 164 layers. Inception v3, which has recently been widely used in the medical field, has been used to diagnose diabetic retinopathy or skin cancer and reported excellent results (Esteva et al., 2017; Gulshan et al., 2016), 299 Due to the limitation of ×299 resolution and 48 layers of convolutional layers, it was judged that the use of Inception-ResNet-v2 with excellent classification performance using 55.9 million parameters was suitable for the purpose of this study.

4. Augmenting training data

For augmentation of the training data, the left and right vertical and horizontal translation pixel area of the image was -30 to 30, the rotation angle area was set to -90 to 90 degrees, the size control area was set to 0.5 to 1.5, and flipping was applied. Flipping was randomly determined during each training session.

5. Transfer learning

Using the pre-trained Inception-ResNet-v2, the learning algorithm performed transfer learning using the stochastic gradient descent with momentum

(SGDM) technique. The learning environment parameters were 50 for the maximum epoch, 8 for the mini-batch, 0.0001 for L2 adjustment, and 0.0016 for the initial learning rate. For learning data, 70% of 1,308 images out of 1,868 images were used.

6. Validation dataset

Verification data used 560 oral images, 30% of the total data completely separated from learning data.

IV. Result

1. Image pre-processing

Since the size of the image varies, the image was standardized by converting it to 512 × 512 through preprocessing After that, two methods of histogram equalization (HE) were applied to the standard image, and the HE-1 image was converted from RGB color to CIE 1976 L*a*b* color in the standard image. CLAHE processing was performed on the brightness channel, and after reconversion to RGB color, a 3×3 median filter was applied to each RGB channel. For the HE-2 image, CLAHE was applied to each of the R, G, and B channels from the standard image to RGB color, converted into a new RGB color image, and a 3×3 median filter was applied to each RGB channel . It can be seen that HE was applied in various ways to reduce the influence of the shooting environment or ambient brightness, etc., and the lesion area was processed as an image with better perceptual contrast than the standard image (see Figure 19) .

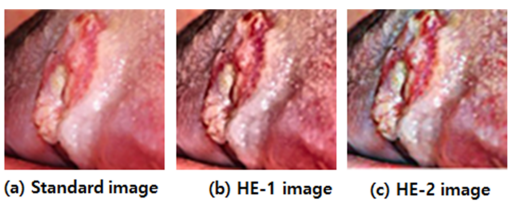


Figure 19. Pre-processed tongue cancer image

2. Augmenting training data

Figure 20 shows how the first 8 chapters of the training data are converted through data augmentation and input to the deep neural network to be used for learning.

Figure 20. Augmented training images

3. Transfer learning

In Figure 21, the line connected with the black dots is the change in the accuracy and loss value of the entire verification data according to the number of epochs. The thin solid line represents the change in the classification accuracy and loss value of each mini-batch, and the thick solid line represents the change in the value obtained by smoothing the thin solid line value. In the last part, the accuracy of the mini-batch rises to 100% and the loss drops to 0.17%, while the accuracy and loss of the verification data are 81.96% and 97.85%, respectively. Therefore, although learning by training data is successful, it can be seen that there is an increase in the loss value due to overfitting (overlearning) in the verification data.

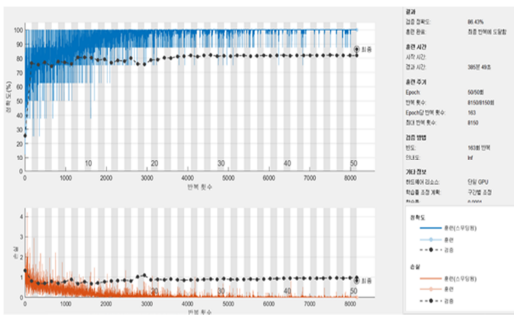
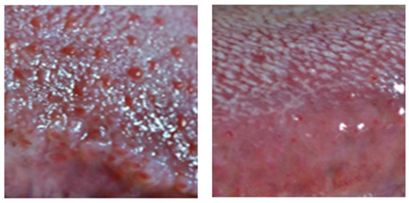


Figure 21. Example of change in accuracy and loss during training and validation.

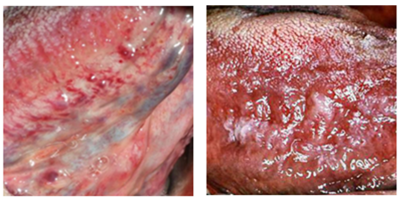
4. Classification example

Figure 22 is an image that accurately classifies class 0 oral images with 100% and 99.7% probability. Figure 23 shows an example of incorrectly classifying class 0 oral images into class 2 (75.3%) and class 3 (59.1%). The final trained DNN indicates that it provides objective and quantitative information by showing the probability that a given oral image for verification will be included in each class.



(a) Classified 99.7% of class 0 images as class 0 (b Classified 100% of class 0 images as class 0

Figure 22. Examples of Accurate Tongue Cancer Classification



(a) Classified 75.3% of class 0 images as class 2 (b Classified 59.1% of class 0 images as class 3

Figure 23. Example of incorrect tongue cancer classification

5. Confusion matrix

When the snow cancer grading performance was calculated based on the error matrix in Figure 24, the overall accuracy was 83.57%. The accuracy between each class was high, with an average of 91.79%, but precision, recall, and F1-score, which are indicators used to evaluate the performance of the multi-classification model of imbalanced labels, averaged 0.67, 0.69, and 0.68. This indicates that the accuracy-to-classification performance is low. This indirectly indicates that inflammatory lesion images (class 1) and precancerous lesions (class 2) are difficult to distinguish from normal (class 0) or cancer (class 3) and are confused (see Table 3).

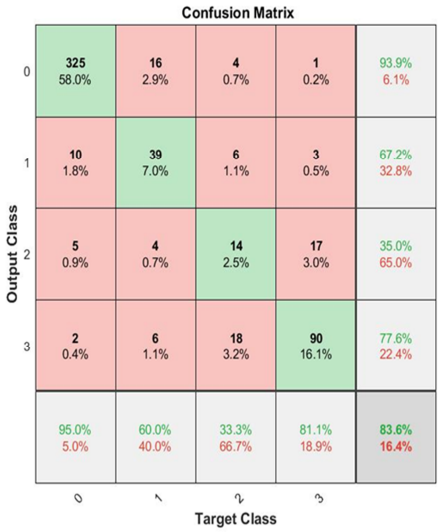


Figure 24. Confusion matrix of tongue cancer classifier

Table 3. Performance index of tongue cancer classifier

Class	Accuracy	Precision	Recall	F1 Score
0	93,21%	0,95	0,94	0,94
1	91,96%	0,60	0,67	0,63
2	90,36%	0,33	0,35	0,34
3	91,61%	0,81	0,78	0,79
Average	91,79%	0,67	0,69	0,68

6. ROC(Receiver operating characteristic) curve

Since the probability of being included in each class is output from the classifier, a Receiver Operating Characteristic curve (ROC curve) can be displayed for each class as shown in Figure 25. The area under the curve (ARC), which is a performance indicator of the classifier, can be calculated. AUC1 (class 0) = 0.9781, AUC2 (class 1) = 0.9102, AUC3 (class 2) = 0.8944, and AUC4 (class 3) = 0.9688, respectively. It can be seen that the AUC of class 1 and class 2 is relatively low, and the classification performance is

inferior compared to other classes.

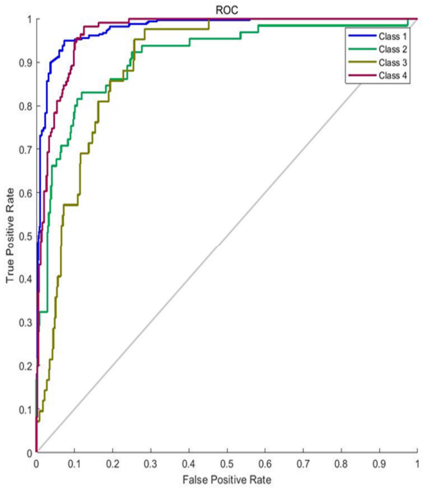


Figure 25. ROC curve of tongue cancer classifier

7. Cohen's Kappa

Cohen's Kappa value, one of the quantitative performance indicators of the classifier, was calculated. The Cohen's Kappa value calculated from the numerical values shown in the error matrix in Figure 24 is 0.857. Khosravi et al. (2018) reported the highest Cohen's Kappa value as 0.81 in digital pathology image classification using various DNNs, and Kwasigroch et al. (2018) reported a Cohen's Kappa value of 0.776 in diabetic retinopathy stage classification using deep CNN. It showed better performance than the Cohen's Kappa value of previous studies using RGB images.

8. Accuracy comparison between preprocessing methods

The accuracy of the standard image without preprocessing and the HE-1 image and HE-2 image preprocessed in different ways were 82.77%, 82.55%, and 83.38% on average for 10 times. As a result of the paired-sample t-test, the preprocessed images (HE-1, HE-2) showed a p value of 0.05 or higher compared to the standard images, showing no statistically significant difference. Although each accuracy did not show a statistically significant difference, as shown in Figure 26, HE-2 generally showed slightly superior performance.

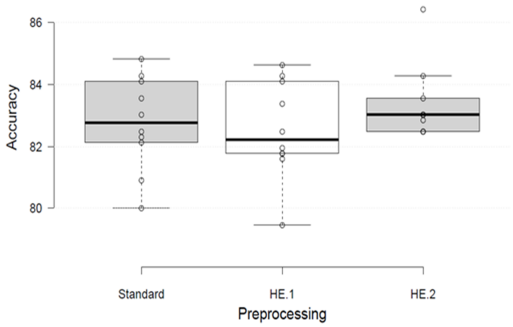


Figure 26. Accuracy comparison for each preprocessing

V. Discussion

This study is the first attempt to quantify and objectify the diagnosis of tongue cancer using oral images using Inception-ResNet-v2, a pre-trained DNN. As a result of comparing various pre-processing methods, the average accuracy of 82.77% (standard image), 82.55% (HE-1), and 83.38% (HE-2) was shown. In study on Age-related Macular Degeneration (AMD) classification using the NIH AREDS data set and DNN, the ophthalmologist's classification achieved an accuracy of 85.33% and the DNN's classification achieved an accuracy of 84.77%. As a result, it was reported that the classifications of doctors and DNNs were similar[38]. Codella et al. (2017) conducted a study to classify melanoma with deep learning in Dermoscopy images, and as a result, dermatologists showed 70.5% accuracy and DNN 76.0% accuracy, reporting that DNN classifies more accurately than doctors[39]. This study is the first attempt to quantify and objectify the diagnosis of tongue cancer using oral images using Inception-ResNet-v2, a pre-trained DNN. As a result of comparing various pre-processing methods, the average accuracy of 82.77% (standard image), 82.55% (HE-1), and 83.38% (HE-2) was shown. In study on Age-related Macular Degeneration (AMD)

classification using the NIH AREDS data set and DNN, the ophthalmologist's classification achieved an accuracy of 85.33% and the DNN's classification achieved an accuracy of 84.77%. As a result, it was reported that the classifications of doctors and DNNs were similar[40]. Codella et al. (2017) conducted a study to classify melanoma with deep learning in Dermoscopy images, and as a result, dermatologists showed 70.5% accuracy and DNN 76.0% accuracy, reporting that DNN classifies more accurately than doctors. Compared to previous studies using these RGB images, this study showed excellent classifier performance with an average accuracy of 82.90% (79.46~86.43%)[38]. Recently, many research results have been reported in the transfer learning method using various pre-trained DNNs to improve the accuracy of diagnosis with DNN using various RGB images in the medical field. Feng Li et al. (2019) showed that diabetic retinopathy can be diagnosed with a sensitivity of 87.0% and specificity of 98.5% using Inception v3 DNN[41]. Gargeya et al. (2017) developed a new diagnostic tool for detecting diabetic retinopathy with a sensitivity of 94%, specificity of 98%, and AUC of 0.97 by training a DNN with 1024 depths using 75,137 published fundus images of diabetic patients[42]. Esteva et al. (2017) used 127,463 training and verification images for transfer learning using Inception v3 CNN for skin cancer level classification, in classification of 3 diseases[43], CNN 72.1% and 2 dermatologists showed accuracy of 65.56% and 66.0%, respectively, as a result, it has been reported that the level of skin cancer can be classified. Although the data set in this study is limited compared to similar studies, the overall accuracy was 82.90% on average, showing the possibility of grading tongue cancer. However, the F1-score, which reflects the data imbalance between labels, showed an average of 0.68, indicating that the classifier performance was degraded. This is because there are relatively few images of inflammatory lesions and precancerous lesions, and the boundary separating the two diseases is ambiguous. It is also understood that the performance of the F1-score classifier is degraded because it is reported that sight examination in clinical practice has limited ability as a screening method for oral cancer and precancerous lesions in the oral cavity[44]. DNN studies using these RGB images are characterized by having pre-classified open big data, easy acquisition compared to other diagnostic imaging equipment, and non-invasive acquisition unlike biopsy or blood tests. Therefore, it is easier to access than DNN research of other expensive imaging medical devices and can be commercialized quickly after verification. As a result, in 2018, the US Food and Drug Administration (FDA) approved IDx-DR, a disease diagnosis device using artificial intelligence, for the first time. IDx-DR analyzes pictures of the retina taken by a camera with artificial intelligence to determine whether the patient has diabetic retinopathy. When a doctor takes a picture of an eyeball with a retinal camera and uploads it to a cloud server, IDx-DR analyzes the picture. IDx-DR has excellent diagnostic power by diagnosing 87.4% of mild or more diabetic retinopathy and 89.5% of less than mild diabetic retinopathy, and the biggest feature is that doctors do not intervene in the diagnosis process[45].

VI. Conclusion

In this study, we proposed a method of classifying tongue cancer, for which early diagnosis is important compared to other cancers, using transfer learning techniques using deep neural networks. As a result of

the study, it showed an average accuracy of 82.90% (79.46~86.43%), which was similar to that of previous studies. A limitation of the study is that various techniques that help improve accuracy have not been applied due to limitations in data collection and computing capacity. First is the quality and quantity of training data. The dataset used in this study showed a large difference in image quality depending on limited data collection on the Internet and lighting or surrounding environment. In contrast, the images used in diabetic retinopathy or melanoma research are publicly available big data, and a large amount of data can be used, and the image quality is constant as it is captured at a certain distance. Therefore, it is considered necessary to standardize intraoral imaging using specific assistive devices, improve the imbalance in the amount of data between classes, and cooperate with clinicians to build big data. Second, it is necessary to apply various preprocessing besides the histogram equalization used in this study to minimize the influence on the above-mentioned data quality problem due to the preprocessing problem of oral images. In a study for recognizing basal cell carcinoma in histopathological images, a segmentation function was applied to improve accuracy through automated cell image search and segmentation[46]. However, recent studies analyzing deep learning methods for melanoma report that no complicated preprocessing techniques are required except for basic standardization tasks such as pixel value normalization, resizing, and cropping. It was said that it is more important to acquire a data set including images that have been edited[47]. Third, there is a phenomenon in which lesions are distorted due to artifacts or severe luminance differences in images acquired by air bubbles or light reflection caused by saliva in the oral cavity due to problems with acquisition characteristics of oral images. In fact, as a result of visually examining the images with errors in classification, in most images, blurring due to focus errors due to light reflection, artifacts due to air bubbles or tongue coating, and shaded areas due to extreme luminance differences are prominent. showed shape. Therefore, in order to minimize problems caused by light reflection during image acquisition, a circular polarizing lens (CPL) filter, a type of polarizing filter, was installed on the camera to minimize the effect of light reflection, and various artifacts were minimized. Images should be acquired. Fourth, it is necessary to minimize imbalance between labeled data. As shown in the results of this study, when the number of specific label images is relatively small, the overfitting phenomenon becomes prominent during training, resulting in low accuracy. Therefore, when sufficient data is not secured, the effect of data imbalance can be reduced by applying weights between each label or training with the same number of images through replication within the label, which is also considered a research task. Despite the above-mentioned limitations, this study was able to obtain a level of accuracy similar to that of other studies using RGB images, which is used as basic data for tongue cancer classification studies of oral images through DNN, which is still lacking in research. Will be available It is expected that it will help in the early diagnosis of tongue cancer if the shortcomings of the data set are supplemented in the future.

Competing interests

The authors declare that there are no competing interests.

[Reference]

[1] Cesar Rivera, "***Essential of oral cancer***", Int. J. Clin. Exp. Pathol,. Volum 8, No. 9, (2015) PP. 11884-11894. PMID: 26617944

[2] Chao-Hui Zuo, Hailong Xie, et al.,

"***Characterization of lymph node metastasis and clinical significance in the surgical treatment of gastric cancer***", Molecular and Clinical Oncology, Published online on: June 3, 2014 Page 821-826. https://doi.org/10.3892/mco.2014.303

[3] Samon Warnakulasuriya, "***Global epidemiology of oral and oropharyngeal cancer****"*, Oral Oncology, Vol. 45, Issue 4-5, (2009), PP. 309-316 https://doi.org/10.1016/j.oraloncology.2008.06.002

[4] Pablo H, Montero, Snehal G. Patel, "***Cancer of the Oral Cavity***", Surg Oncol Clin N Am, Vol. 24, No. 3, (2015) PP. 491-508. https://doi: 10.1016 /j. s o c.2015.03.006

[5] Neda Mahdavifar, Mahshid Ghoncheh, et al., "***Epidemiology, Incidence and Mortality of Bladder Cancer and their Relationship with the Development Index in th world***", Asian Pac J Cancer Prev., Vol. 17, No. 1, (2016) PP. 381-386. https://doi.org/10.7314/apjcp.2016.17.1.381.

[6] Nanditha Sujir, Junaid Ahmed, et al., "***Challenge in Early Diagnosis of Oral Cancer: Cases Series***", Acta Stomatol Croat, Vol. 53, No. 2, (2019) PP. 174-180. https://doi.org/10.15644/asc53/2/10

[7] Dong-Ho Geum, Young-Chea Roh, et al., "***The impact factors on 5-year survival rate in patients operated with oral cancer****"*, J Korean Assoc Oral Maxillofac Surg, Vol. 39, No. 5, (2013) PP 207-216. http://doi.org/10.5125/jkaoms.2013.39.5.207

[8] Volkan Erdogu, Necati Citak, Celal Bugra Sezen, "***Comparison of 6th, 7th, and 8th edition of the TNM staging in non-small cell lung cancer patients: Validation of the 8th edition of TNM staging***", Turk Gogus Kalp Damar Cerrahisi Derg, Vol. 30, No. 3, (2022) PP. 395-403. https://doi.org/10.5606/tgkdc.dergisi.2022.20089

[9] Yong-Seok Choi, Min Gyeong Kim, Jong-Ho Lee, et al., "***Analysis pf prognostic factors through survival rate analysis of oral squamous cell carcinoma patients treated at the National Cancer Center: 20 years of experience****", j Korean Assoc Oral Maxillofac Surg, Vol. 48, No. 5, (2022) PP. 284-291.* *https://doi.org/10.5125/kkoams.2022.48.5.284*

[10] Douglas E. Morse, Walter J. Psoter, Deborah Cleveland, et. al., "***Smoking and drinking in relation to oral cancer and oral epithelial dysplasia****"*, Cancer Causes Control, Vol. 18, No. 9, (2007) PP. 919-929. https://doi.org/10.1007/s10552-007-9026-4

[11] Kyoung Ah Kim, Joon Beom Seo, Kyoung Hyun Do, et. al., "***Differentiation of Recently Infracted Myocardium from Chronic Myocardial Scar: The Value of Contrast-Enhanced SSFP-Based Cine MR Imaging***", Korean J Radiol, Vol. 7, No. 1, PP. 14-19. https://doi.org/10.3348/kjr.2006.7.1.14

[12] Ross Gruetzemacher, David Paradics, Kang-bok Lee, "***Forecasting extreme labor displacement: A surver of AI practitioners***", Technology Furecasting and Social Change, Vol. 161, (2020) PP. 120323 https://doi.org/10.1016/j.techfore.2020.120323

[13] Heang-Ping Chan, Lubomir M Hadjiiski, Ravi K Samala, "***Computer-aided diagnosis in the era of deep learning****"*, Med Phys, Vol. 47, No. 5, (2020) PP. e218-e227 https://doi.org/10.1002/mp.13764

[14] Dartmouth workshop - Wikipedia

[15] https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

[16] Sarah Friedrich, et al., "***Is there a role for statistics in artificial intelligence?***", Advanced in Data Analysis and Classification, (2022) 16: PP. 823-846. https://doi.org/10.1007/s11634-021-00455-6

[17] IBM Is Counting on Its Bet on Watson, and Paying Big Money for It - The New York Times (nytimes.com)

[18] Jiyeon Yu, Angelica de Antonio, Elena Villalba-Mora, "***Deep Learning(CNN, RNN) Application for Smart Homes: A Systematic Review***", Computers 2022, 11(2), 26

https://doi.org/10.3390/computers11020026

[19] Liviu Ciortuz, "***Machine learning***", Department of CS, University of Iasi, Romania https://profs.info.uaic.ro/~ciortuz/SLIDES/2017s/ml0.pdf

[20] Koza, J.R., Bennett, F.H., Andre, D., et al., "***Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming***", In: Gero, J.S., Sudweeks, F. (eds) Artificial Intelligence in Design '96 (1996). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-0279-4_9

[21] Junyan Hu, Hanlin Niu, Joaquin Carrasco, Barry Lennox, Senior Member, "***Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning***", IEEE Transactions on Vehicular Technology, Vol. 69, No. 12, (2020) PP. 14413-14423. IEEE Xplore Full-Text PDF:

[22] ***Machine Learning for Beginners: An Introduction to Neural Networks*** | by Victor Zhou | Towards Data Science

[23] Stuart J. Russell, Peter Norvig, "***Artificial Intelligence: A Modern Approach***", *Third Edition*, (2010) Prentice Hall ISBN 9780136042594,

[24] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar. "*Foundations of Machine Learning*", (2012) The MIT Press ISBN 9780262018258.

[25] Hinton, G., "***A Practical Guide to Training Restricted Boltzmann Machines***" (PDF). Neural Networks: Tricks of the Trade, Lecture Notes in Computer Science . Springer, (2012) Vol. 7700, ISBN 978-3-642-35289-8 https://doi.org/10.1007/978-3-642-35289-8_32.

[26] L. P. Kaelbling, M. L. Littman, A. W. Moore, "***Reinforcement Learning: A Survey***", Journal of Artificial Intelligence Research, Vol. 4, (1996) PP. 237-285, https://doi.org/10.48550/arXiv.cs/9605103

[27] Martijn van Otterlo, Marco Wiering, "***Reinforcement Learning and Markov Decision Processes***", ALO, Volume 12, PP. 3-42. ISBN 973-3-642-27644-6. https://doi.org/10.1007/978-3-642-27645-3_1

[28] Hardesty, Larry (14 April 2017), "***Explained: Neural networks***", MIT News Office, *Explained: Neural networks | MIT News | Massachusetts Institute of Technology* , Retrieved 2 June 2021.

[29] Yang Z. R., Yang, Z. "***Comprehensive Biomedical Physics***", Elsevier, (2014) Volume 1, ISBN 978-0-444-53633-4. Comprehensive Biomedical Physics - 1st Edition (elsevier.com)

[30] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, "***Deep learning***", PMID: 26017442, https://doi.org/10.1038/nature14539

[31] M. V. Valueva, N. N. Nagornov, P. A. Lyakhov, et al., "***Application of the residue number system to reduce hardware costs of the convolutional neural network implementation***", Mathematics and Computers in Simulation, Vol. 177, (2020), PP. 232-243, https://doi.org/10.1016/j.matcom.2020.04.031

[32] Zhang, Wei, "***Parellel distributed processing model with local space-invarient interconnection and its optical architecture***", Applied Optics, Vol. 29, No. 32, PP. 4790-4799. https://doi.org/10.1364/AO.29.004790.PMID20577468

[33] Coenraad Mouton, Johannes C. Myburgh, Marelie H. Davel, "***Stride and Translation Invariance in CNNs***", Communications in Computer and Information Science,(2020) Vol. 1342, PP. 267-281. https://doi.org/10.1007/978-3-030-66151-9_17

[34] Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen, "***Deep content-based music recommendation***(PDF)", Curran Associates Inc, PP. 2643-2651. Deep content-based music recommendation (neurips.cc)

[35] Ronan Collobert, Jason Weston, "***A unified archtecture for natural language processing: deep neural networks with multitask learning***", ICML'08: Proceeding of the 25th international conference on Machine learning, (2008) PP.160-167. https://doi.org/10.1145/1390156.1390177

[36] Avraam Tsantekidis, Nikolaos Passalis, Anastasios Tefas, et al., "***Forecasting Stock Prices from the Limit Order Book Using Convolutional Neural Networks***", 2017 IEEE 19th Conference on Business Information(CBI), https://doi.org/10.1109/CBI.2017.23

[37] D. H. Hubel, T. N. Wiesel, "***Receptive fields and functional architecture of monkey striate cortex***", J. Physiol. (1968) Vol. 195. No. 1, PP. 215-243. https://doi.org/10.1113/jphysiol.1968.sp008455

[38] Emily Y. Chew, "***Age-related Macular Degeneration: Nutrition, Genes and Deep Learning-The LXXVI Edward Jackson Memorial Lecture***", American Journal of Ophthalmology, (2020) Vol. 217, P 335-347. https://doi.org/10.1016/j.ajo.2020.05.042

[39] Noel Codella, Quoc-Bao Nguyen, S. Pankanti, et al., "***Deep Learning Ensembles for Melanoma Recognition in Dermoscopy Images***". Ibm Journal of Research and Development, (2017) Vol. 61 No. 4. PP.

[40] Yifan Peng, Shazia Dharssi, Qingyu Chen, et al., "***DeepSeeNet: A deep learning model for automated classification of patient-based age-related macular degeneration severity from color fundus***

photographs", Ophthalmology, (2019) Vol. 126, No. 4, PP 565-575. https://doi.org/10.1016/j.ophtha.2018.11.015

[41] Feng Li, Zheng Liu, Hua Chen, et al., "***Automatic Detection of Diabetic Retinopathy in Retinal Fundus Photographs Based on Deep Learning Algorithm***", Transl Vis Sci Technol, (2019) 8(6):4. https://doi.org/10.1167/tvst.8.6.4

[42] Rishab Gargeya, Theodore Leng, "***Automated Identification of Diabetic Retinopathy Using Deep Learning***", Opthalmology, (2017) 127(7), PP. 962-969. https://10.1016/j.ophtha.2017.02.008

[43] Andre Esteva, Brett Kuprel, Roberto A. Novoa, et al., "***Dermatologist-level classifiaction of skin cancer with deep neural netwirk***", Nature, (2017) 542(7639). PP. 115-118. https://doi.org/10.1038/nature21056

[44] Maria Garcia-Pola, Eduardo Pons-Fuster, Carlota Suarez-Fernandez, et al., "***Role of Artificial Intelligence in the Early Diagnosis of Oral Cancer. A Scoping Review***", Cancer(Basel), (2021) 13(18): 4600, https://doi.org/10.3390/cancers13184600

[45] Jerry Helzner, "***SUBSPECIALTY NEWA: FDA approves first AI device to detect DR, brolucizumab reliable for 12-week dosing, and more***", Retinal PHYSICAL, Article. https://www.retinalphysician.com/issues/2018/june-2018/subspecialty-news-fda-approves-first-ai-device-to

[46] Y Q Jiang, J H Xiong, H Y Li, et al., "***Recognizing basal cell carcinoma on smartphone Capured difital histopathology images with a deel neural network***", Br J Dermatol, (2020) 182(3), PP. 754-762. https://doi.org/10.1111/bjd.18026.

[47] Nazneen N. Sultane, Bappaditya Mandal, N. B. Puhan, "***Deep residual network with regularised fisher framework for detection of melanoma***", IET Computer Vision in Cancer Data Analysis, (2018) Vol. 12, Issue. 8, PP. 1067-1227. https://doi.org/10.1049/let-CVI.2018.5238