Journal Pre-proof

Automatic discrimination of Yamamoto-Kohama classification by machine learning approach for invasive pattern of oral squamous cell carcinoma using digital microscopic images: A retrospective study

Kunio Yoshizawa , Hidetoshi Ando , Yujiro Kimura , Shuichi Kawashiri , Hiroshi Yokomichi , Akinori Moroi , Koichiro Ueki

PII: S2212-4403(21)00683-0
DOI: https://doi.org/10.1016/j.oooo.2021.10.004
Reference: OOOO 4736

To appear in: *Oral Surg Oral Med Oral Pathol Oral Radiol*

Received date: 15 June 2021
Revised date: 2 September 2021
Accepted date: 6 October 2021

Please cite this article as: Kunio Yoshizawa , Hidetoshi Ando , Yujiro Kimura , Shuichi Kawashiri , Hiroshi Yokomichi , Akinori Moroi , Koichiro Ueki , Automatic discrimination of Yamamoto-Kohama classification by machine learning approach for invasive pattern of oral squamous cell carcinoma using digital microscopic images: A retrospective study, *Oral Surg Oral Med Oral Pathol Oral Radiol* (2021), doi: https://doi.org/10.1016/j.oooo.2021.10.004

Automatic discrimination of Yamamoto-Kohama classification by machine learning approach for invasive pattern of oral squamous cell carcinoma using digital microscopic images: A retrospective study

Kunio Yoshizawa[a], Hidetoshi Ando[b], Yujiro Kimura[c], Shuichi Kawashiri[d], Hiroshi Yokomichi[e], Akinori Moroi[f], Koichiro Ueki[g]

[a]Associate Professor, Department of Oral Maxillofacial Surgery, Division of Medicine, Interdisciplinary Graduate School, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi, 409-3898, Japan.

[b] Associate Professor, Department of Media Engineering, Graduate School of University of Yamanashi, 4-4-37 Takeda, Kofu, Yamanashi, 400-8510, Japan

[c] Graduate student, Department of Oral Maxillofacial Surgery, Division of Medicine, Interdisciplinary Graduate School, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi, 409-3898, Japan.

[d] Professor, Department of Oral and Maxillofacial Surgery, Kanazawa University Graduate School of Medical Science, 13-1 Takara-machi, Kanazawa, Ishikawa, 920-8641, Japan

[e] Department of Health Sciences, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi, 409-3898, Japan. Japan

[f] Lecture, Department of Oral Maxillofacial Surgery, Division of Medicine, Interdisciplinary Graduate School, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi, 409-3898, Japan.

[g] Professor, Department of Oral Maxillofacial Surgery, Division of Medicine, Interdisciplinary Graduate School, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi, 409-3898, Japan.

[*]Correspondence to: Kunio Yoshizawa

Tel.: +81-55-273-2387; Fax: +81-55-273-9673

E-mail: yoshizawak@yamanashi.ac.jp (K. Yoshizawa)

Statement of Clinical Relevance: 39 words

We have introduced machine learning to automatically determine invasion patterns using medical image processing based on digital images of the invasive front of oral cancer. This has made it possible to accurately determine the invasion pattern of oral cancer.

Abstract.

Objective. The Yamamoto–Kohama criteria are clinically useful for determining the mode of tumor invasion, especially in Japan. However, this evaluation method is based on subjective visual findings and has led to significant differences in determinations between evaluators and facilities. In this retrospective study, we aimed to develop an automatic method of determining the mode of invasion based on the processing of digital medical images. **Study Design.** Using 101 digitized photographic images of anonymized stained specimen slides, we created a classifier that allowed clinicians to introduce feature values and subjected the cases to machine learning using a random forest approach. We then compared the Yamamoto–Kohama grades (1, 2, 3, 4C, 4D) determined by a human oral and maxillofacial surgeon with those determined using the machine learning approach. **Results.** The input of multiple test images into the newly created classifier yielded an overall F-measure value of 87%, (Grade 1: 93%, Grade 2: 67%, Grade 3: 89%, Grade 4C: 83%, Grade 4D: 94%). These results suggest that the output of the classifier was very similar to the judgments of the clinician. **Conclusions.** This system may be valuable for diagnostic support to provide an accurate determination of the mode of invasion.

Keywords: mode of invasion, oral squamous cell carcinoma, random forest, medical image processing, machine learning

1. Introduction

Oral squamous cell carcinoma (OSCC) accounts for approximately 90% of all cases of oral cancer. Despite improvements in treatment options over the past few decades, the 5-year survival rates have remained fairly low (50–60%) among OSCC patients [1-4]. Treatment failure in a case of OSCC is mainly ascribed to the highly invasive nature of the tumor [5, 6]. As the tumor becomes more invasive, the invasion front progresses from the epithelium through the stroma to infiltrate the lymphatic and blood vasculature [7]. This phenomenon is directly linked to an increased likelihood of metastasis and a poor survival prognosis. Particularly, an accurate diagnosis of the invasiveness of an OSCC is a very important component of treatment planning and prognostic predictions [8, 9]. To improve the diagnosis and prognosis of OSCC, previous researchers have reported various histopathological classifications.[5, 9-12] The histopathological classification of OSCC by the World Health Organization (WHO) is based on the original description by Broders.[10] In this system, squamous cell carcinoma is a malignant epithelial neoplasm exhibiting squamous differentiation, as characterized by the formation of keratin and/or the presence of intercellular bridges representing desmosomes.[13] The Yamamoto-Kohama (YK) classification proposed by Yamamoto et al.,[9, 14, 15], which subdivides Grade 4 that represents the worst prognosis based on Jakobson's criteria.[12],

into Grade 4C (cord-like type) and Grade 4D (Diffuse-type), classifies OSCC according to invasive capacity. Japanese oral surgeons often use the YK system to predict metastasis and prognosis (Table 1). Understanding the invasive ability of OSCC is essential in providing appropriate treatment. YK system classifies invasion into the stroma as type 3 or higher, and prognosis worsens as the carcinoma progresses to type 4C and 4D. We have previously shown that intercellular adhesions, such as E-cadherin, weaken with worsening invasive patterns.[6, 16] In the modification of the Eighth Edition of the American Joint Committee on Cancer (AJCC),[17] depth of invasion (DOI) was incorporated into the criteria for T classification.[17] Aaboubout et al. recommend elective neck dissection if the DOI is greater than 4 mm, considering the possibility of occult lymph node metastasis in early-stage cancer.[18] These results suggest that YK classification and the DOI values are closely correlated with the prognosis. This classification system appears to be a powerful predictor of regional metastasis in a patient with clinically node-negative OSCC. A YK-classification-based evaluation mainly involves biopsy and excised tissues, and the results are used to determine the prognosis and select treatments. However, this evaluation method is based on subjective visual findings and has led to significant differences in determinations between evaluators and facilities. Therefore, the YK classification is not a sufficiently objective

index. Furthermore, no report has described the relationship between the visual aspects of infiltration on images of pathological tissues and the results of an objective image evaluation based on samples from patients with OSCC.

Many recent studies have demonstrated the effectiveness of pathological image analysis methods that incorporate artificial intelligence (AI). One experimental study that compared the diagnostic accuracy of a pathologist and AI with respect to pathological images of breast cancer lymph node metastasis determined that the latter was more time-efficient [19]. In another example, reports from various countries have reported that the Gleason score, an index of prostate cancer malignancy, is poorly reproducible among pathologists. In that context, Arvantii et al. demonstrated the use of AI to match the accuracy rate of the Gleason index with its reproducibility among pathologists [20].

Several recent reports have described various approaches that have used machine learning to detect various grades of carcinomas from photographic images of lesions, radiological images and pathological specimens [21-29]. In the field of oral cancer and dental surgery, there are many reports on the application of AI systems, such as a study using deep learning to determine lymph node metastasis from CT findings and color Doppler ultrasonography[30-32], and classification of dental restorations from panoramic

findings[33]. However, few reports have described an automatic method for determining the invasion activity based on the computer processing of a digital image of the invasion front in oral cancer[34]. On the other hand, while few reports describe a method to automatically determine invasive activity based on computer processing of digital images of the invasive front of oral cancer, shan et al. reported that invasive patterns such as "tumor budding" and DOI predominantly predict lymph node metastasis[34].

Therefore, in this study, we aimed to develop a method for medical image processing to automatically determine the mode of invasion based on digital images of the invasive front of an OSCC.

2. Materials and Methods

2.1. Specimens

Sixty-seven primary OSCC biopsy specimens were obtained from patients who underwent surgical resection at the Department of Oral and Maxillofacial Surgery, Kanazawa University Hospital between 1989 and 2009. The patients (38 male and 29 female subjects) ranged in age from 32 to 91 years (mean age, 60 years). Informed consent for the experimental use of the samples was obtained from the patients according to the hospital's ethical guidelines. The engineering department of Yamanashi

University performed the imaging analysis of the pathological specimens as a third-party assessment organization to eliminate evaluator bias. Prof. Yamamoto, the proponent of the Yamamoto-kohama classification, helped us (clinicians) discriminate the Grade for YK classification from each HE specimen as supervised images. A total of 101 specimens were evaluated and assigned the following YK grades: Grade 1, 23 specimens; Grade 2, 12 specimens; Grade 3, 27 specimens; Grade 4C, 21 specimens and Grade 4D, 18 specimens. The retrospective study protocol was approved by the ethics committees of Yamanashi University (approval number: 1267) and Kanazawa University (approval number: 1647-1). This study was conducted in accordance with the Declaration of Helsinki.

2.2. Staining methods

Immunohistochemistry (IHC) of deparaffinized and rehydrated sections was performed according to the labeled streptavidin-biotin (LSAB) method as described by Nozaki et al. [14]. To clearly detect tumor cells at the borderline, the sections were reacted overnight at 4 °C with each primary monoclonal antibody specific for urokinase-Grade plasminogen activator/receptor (uPA/uPAR) (American Diagnostica Inc., USA; 200-fold dilution in phosphate-buffered saline [PBS]) and claudin-7 (Invitrogen Corp., Camarillo, CA, USA; 200-fold dilution in PBS). uPA/uPAR was proven to distinguish

OSCC with higher invasive Grades (Grade 4C and 4D), while claudin-7 with tight junction component was proven to distinguish OSCC with lower invasive Grades (Grade 1, 2, and 3) in an immunohistochemical analysis of pathological tissue specimens according to the YK classification [15, 35]. The sections were then reacted with a secondary antibody (biotin-labeled goat anti-rabbit immunoglobulin polyclonal antibody; Dako Japan, Kyoto, Japan) at room temperature for 60 min. Sections treated with PBS instead of the primary antibody were used as the negative controls.

2.3. Yamamoto–Kohama (YK) classification

In Japan, the departments of oral–maxillofacial surgery at many institutions use the YK classification. This method is used for the histological evaluation of malignant tissues and is focused on the invasion pattern at the tumor–host tissue border. The YK classification was previously shown to be strongly correlated with the risk of lymph node metastasis and prognosis [9]. The YK evaluation criteria are presented in Table 1.

2.4. Overview of the machine learning methods

Two approaches to the automatic determination of the OSCC invasion pattern were applied in this study. First, we searched the region of interest (ROI) of the invasive front using whole slide images (WSI) and then the same ROI site was expanded to ×100 to extract the features and determine the mode of invasion (Figure 1). The ROI was

selected with reference to the depth of invasion (DOI) recommended by the American Joint Committee on Cancer (AJCC) eight edition as staging system of T-category[36]. To select ROI as point of deepest invasion, we establish the horizon that is at the level of the basement membrane relative to the closest intact squamous mucosa, and select ROI from the deepest invasion by dropping a "plumb line" from the horizon, as shown in Fig1. Second, machine learning was applied to cases for which a clinician had previously evaluated the mode of invasion based on the YK classification, and the images were classified by random forest. Here, we considered a shape characterization of the invasive front in the image to be effective for discriminating the mode of invasion. The characterized shape of the OSCC invasion patterns was then extracted by experienced pathologists to create feature vectors that were suitable for the classifier of the mode of invasion.

The proposed processing method was performed as described by Inoue et al. [37] and is summarized in Figure 2. A schematic summary of the series of image processing is shown in Figure 2. We evaluated two approaches for extracting the feature vector of YK classification. First, we extracted the color features for binarization from the original immunohistochemical images and made a classifier for binarization. Second, we extracted the shape features for discrimination of YK classification and made a

classifier for the discrimination of YK classification. The invasion mode (i.e., YK classification) was determined automatically using machine learning according to the following methods, which are presented in the following order [37]: 1. extraction of color features for binarization; 2. creation of classifiers for binarization; 3. binarization of unknown color data; 4. extraction of shape features for the discrimination of YK classification; 5. creation of classifiers for YK classification; and 6. discrimination of the YK classification of the binary image. The microscope images were digitalized using a whole slide scanner (KEYENCE BZ-9000, KEYENCE, Osaka) at a resolution of 680 x 512 pixels/inch. When the original image was binarized, the image was compressed to 320×240 pixels/inch, which was sufficient to capture the features of color and texture. In this research, we demonstrate the use of local binary patterns (LBP) [38] in combination with random decision tree classifiers, which can be used to divide the tumor epithelium and the stromal region of OSCC. Furthermore, scikit-learn, which is famous for Python, was used for numerical science and technology calculation of machine learning [39].

2.5. Binarization

The histopathological image of each tumor was divided into epithelial and stromal regions to extract the invasion front from the image. First, a borderline was created to

divide the tumor epithelium and stromal regions. Binarization was then performed to distinguish the epithelial and stromal sides. In this process, the color pixels on the tumor side were converted to black and those on the stromal side to white. Initially, clinical experts performed the binarization processing series and used the resulting human analyst-generated images as a training dataset for machine learning. The binarization process is summarized in the upper panel of Figure 2. Next, a local binary patterns operator [38] was used to construct a binary code from feature vectors that extracted the three color data of RGB values (red, green, blue) and the local texture features from 49 pixels within a square area of which one side comprised 7 pixels centered on the pixel of ROI. Color deconvolution is used extensively in histopathology image analysis to separate an RGB image into three channels (red, green, and blue), each corresponding to the actual colors of IHC staining [40, 41]. At this point, the number of feature vectors was 49 pixels with a feature vector of ×3 (the three colors of RGB: Red, Green, Blue), resulting in 147 dimensions per pixel. We applied the local binary patterns (LBP) operator to capture the spatial representation of the color image and thus enable the classification [38]. We created a classifier that could use machine learning to determine whether a training data image should be classified as black or white based on image-specific stain information with the local RGB pixel information in a supervised

classification framework according to the methods used past researcher [42-44]. In these protocols, we attained 101 images of the binary images from the original IHC images as shown in Fig. 3.

2.6. Design of the feature extractor

The clinician applied the following five features to automatically classify the binarized image into YK-classification by machine learning: 1. number of epithelial areas, 2. borderline disturbance, 3. cord-shaped epithelial area, 4. size of the epithelial area and 5. borderline length.

2.7. Features and extraction

2.7.1. Number of tumor (epithelium) areas

The inputted binary image data were subjected to labeling on the tumor side consisting of epithelium tissue (i.e., black-colored side). The areas surrounded by continuous black lines were counted to determine the number of tumor islands. Figure 3 demonstrates that the number of tumor islands, which was defined as feature value 1 (Number) [22], increased as the YK grade increased. The data serial numbers were then ordered from YK Grade 1 to 4D, such that Grade 4D data sets had the highest serial numbers.

2.7.2. Disturbance of the borderline

The inputted binary image was then vectorized with respect to the pixels that

represented the tumor side of the borderline, which corresponded to the basal cell layer of OSCC. For a labeled object, i, if the length of the borderline is Li and the number of division points used for vectorization is Ni, the curvature factor (R) of the average borderline in the image can be expressed as Formula (1) [37]:

$$R = \frac{1}{n}\sum_{i=1}^{n}\frac{N_i}{L_i} \quad (1).$$

Consequently, the curvature of the borderline increases as the number of division points increases, even when the lengths of the borderlines are identical. This curvature is defined as feature value 2 (Curvature). Feature value 3 (Smooth) and feature value 4 (Sharp) were extracted from the protrusion that occurred from the angle formed by a vector from the middle division point to the front and rear division points according to the threshold value. Furthermore, to classify cases where there was a difference in the variation in the directionality of the protrusions, the directionality of the protrusions was extracted and set as feature value 5 (Direction).

2.7.3. Cord shape of the epithelial area

For an inputted binary image, the highest numerical value yielded by dividing the square root of the labeled object size by the length of the corresponding contour line, was extracted and set as feature value 6 (Cord).

2.7.4. Size of the epithelial area

The number of labeled objects with a size below a certain threshold was extracted and set as feature value 7.

2.7.5. Length of the borderline

The length of the borderline was set as feature value 8 (Length) because this parameter was expected to facilitate the distinction between YK Grades 1 and 2.

2.8. Performance evaluation test using a random forest approach

Next, we experimentally analyzed the resulting discriminant performance when we performed an evaluation based on the YK classification and the extracted features. Here, we used the leave-one-out (LOO) evaluation method [45-47] and the random forest machine learning algorithm to create a classifier [48]. Specifically, at each iteration of LOO process, which is performed over the 101 cases of the entire dataset, one sample is reserved for testing and all other samples are used to provide as training data of random forest. Table 2 summarizes the main hyperparameters of the random forests used in the experimental analysis of the image data subjected to machine learning. This hyperparameter was determined after being tuned by random search for optimization [49, 50]. The F-measure was used as an indicator of precision–recall and was calculated from a confusion matrix that summarized the discrimination analysis of each YK classification.

2.9. Comparison of survival rates determined by the machine learning approach and a clinician in accordance with the YK classification

We constructed Kaplan–Meier survival estimates to illustrate the 5-year overall survival rates by YK classification. We performed a log-rank test to detect the statistical significance between the estimates for the inter-group difference.

2.10 Statistical Analysis

Data analyses were performed using the statistical software SPSS 27.0 for Windows (SPSS, Inc., Chicago, IL, USA). One-way ANOVA and *t*-test with Bonferroni adjustment were used to compare the means of feature amount among modes of invasion group. The Kaplan–Meier estimate compared the survival rates in modes of invasion between classifications made by machine learning and clinicians. We performed a log-rank test for the estimates. $P < 0.05$ was considered statistically significant.

3. Results

3.1 Association between DOI and YK-classification

The mean and SD for DOI of YK-1, YK-2, YK-3, YK-4C and YK-4D is 814 ± 680, 873 ± 361, 3930 ± 3623, 7677 ± 5955 and 12450 ± 2490, respectively. The DOI was higher as the mode of invasion for YK classification increased.

3.2 Binary image processing

101 cases, which is deposited as open source-access figures, depict representative binarized images used to perform the YK classifications in this research. The borderlines between the tumor and stromal tissues could be distinguished clearly up to a fine point and are clearly binarized on the IHC image corresponding to each YK classification.

3.3 Parallel coordinates

Fig. 4 shows that the parallel coordinates indicate the relative ratio of feature amount on the Y axis and indicate each feature domain (features No 1–8) on the X axis based on each mode of invasion (five types: Grade 1–4D). The parallel coordinates form a statistical graph that is useful for visualizing the multivariate data [51].

3.4 Distributional observation of the number of epithelial areas (feature value 1)

In Grades 1 and 2, most of the feature value 1 results were distributed near 1, which was consistent with the single tumor masses observed on the images. In contrast, most of the data sets for Grade 3, 4C and 4D cases yielded values greater than 1, which was consistent with the appearance of multiple tumor masses on the images. Particularly, the feature value 1 for Grade 4D specimens was 15 or higher at a half ratio. This

phenomenon was not observed in the other Grades. Feature value 1 provided a good distinction of Grades 1 and 2 from Grade 4C and 4D (Appendix Fig. 1).

3.5 Distributional observation of the disturbance of borderline" (feature values 2 -5)

Most Grade 4D specimens yielded a feature value 2 of 0.08 or greater. Accordingly, this feature 2 value could effectively discriminate Grade 4D tumors (Appendix Fig. 2). Feature value 3 shows that Grade 4C is higher than the other grades, which is suitable for distinguishing Grade 4C (Appendix Figure 3). The amount of feature 4 was the lowest in Grade 1 (Appendix Figure 4). Feature 5 did not show a significant difference between any mode of invasion and was unsuitable for discrimination (Appendix Figure 5).

3.6 Distributional observation of a cord-shaped epithelial area (feature value 6)

Notably, this parameter yielded large values for Grade 4C, intermediate values for Grades 2, 3 and 4D and small values for Grade 1 tumors. Accordingly, feature value 6 could successfully distinguish the cord-like Grade 4C tumors from the other tumors other than Grade 2 (Appendix Fig. 6).

3.7 Distributional observation of the size of the epithelial area (feature value 7)

Grade 4D tumors accounted for most cases with a feature value 7, greater than 13. The

data suggest that feature value 7 is very effective for discriminating Grade 4D (Appendix Fig 7)

3.8 Distributional observation of the borderline length (feature value 8)

Grade 1 tumors tended to yield low values for feature value 8, and most cases with values less than 1000 met the criteria for this grade. Therefore, feature value 8 can effectively discriminate Grade 1 tumors (Appendix Fig. 8).

3.9 Confusion matrix-based performance evaluation

As shown in Table 3, the test data of Grades 1 and 4D yielded high classification accuracy values, whereas the data of Grade 2 yielded a low value. Among Grade 1 cases, only 2 of 23 specimens were misjudged as Grade 2. Among Grade 4D cases, only 1 of 18 specimens was misjudged as Grade 4C. However, 4 of 12 Grade 2 specimens were misjudged as other Grades.

The test data show the original correct results of YK classification determined by a clinician. The discrimination data are the results of the mode of invasion determined by machine learning. The overall sensitivity/specificity was 87.1% (88/101)/96.8% (391/404), and the sensitivity/specificity for each mode of invasion was 91.3%/97.5% in Grade 1, 66.7%/95.5% in Grade 2, 92.6%/97.2% in Grade 3, 81.0%/95.1% in Grade 4C, and 94.4%/98.8% in Grade 4D. Importantly, the sensitivity is clearly lower in Grade 2.

3.10 Precision–recall

The precision–recall was calculated using a confusion matrix and reported using F values, as shown in Table 4. The overall F value was 0.87. In an analysis stratified by classification, Grade 2 received the lowest F value of 0.67, whereas Grades 1 and 4D received the highest F values of 0.93 and 0.94, respectively.

3.11 Comparison of survival rates according to the YK classifications assigned by a clinician or the machine learning method

A comparison of the Kaplan–Meier survival curves calculated for each YK classification revealed a significant difference between the rates associated with the machine learning and clinician classifications only in Grade 2 cases (Fig. 5). Specifically, a Grade 2 classification via machine learning was associated with a lower survival rate, compared to the same classification when assigned by a clinician ($p < 0.05$). No other significant differences in classification accuracy were observed for the other YK grades (Fig. 5).

4. Discussion

OSCC is characterized by a high degree of invasion into the surrounding tissues, as well as a high incidence of lymph node metastasis [52]. In this research, the DOI was higher as the mode of invasion increased. Therefore, it is essential to determine the mode of

tumor invasion in each case.

From the viewpoint of radiomics, it can be predicted that the phenotype of medical images includes genetic information of tumors and prognostic information of cases [53, 54]. Radiomics research includes brain tumor, lung cancer, and breast cancer as typical diseases [55-59]; however, there are still very few reports on oral cancer. Romeo et al. reported a radiomic machine learnig approach employing texture analysis features extracted from primary tumor lesions and CT images applied to primary tumor lesions could predict tumor grade and nodal status in OSCC.[60] In this study, we developed an automatic machine-learning-based method for differentiating OSCC cases according to the YK classification through digital images from histopathological specimens. Overall, this system yielded relatively accurate results, as indicated by a high F value of 0.87. However, a further analysis of individual grades yielded a relatively low F value for Grade 2. The number of specimens in Grade 2 was 13, the lowest compared to the other grades, and this imbalance classification may have led to poor predictive performance.

When we analyzed the survival rates according to the YK grade, the survival rate decreased as the grade determined by the clinician increased. In contrast, however, the machine learning-determined YK Grade 2 cases had the second-worst survival rate after Grade 4D. Imbalanced classification is a predictive modeling challenge because most

machine learning algorithms used for classification are designed with the same number of examples in each class[61]. Therefore, it is conceivable that the lower survival rate for Grade 2 is calculated based on cases that were misclassified by machine learning than the actual value of Grade2. Moreover, only two-thirds of Grade 2 cases (8/12 cases) were correctly assigned by the machine learning system, and three-quarters of the mismatched cases (3/4 cases) actually met the criteria of a higher grade. Grade 2 may be particularly easy to misjudge via machine learning because these lesions have an unclear borderline and a cord-like shape and are easily misclassified as more invasive tumors (e.g., Grade 3, Grade 4C), even during a subjective clinician-based analysis. Grade 2 cases also comprised the smallest subpopulation in this study. Consequently, machine learning became inadequate, and many cases were misinterpreted.

Twelve cases of Grade 2, which as classified by machine learning was incorrectly interpreted in 4 out of 12 cases as follows: one case in Grade 1, two cases in Grade 3, and one case in Grade 4C. The reason why Grade 2 was misinterpreted by the other mode of invasion was considered to be as follows based on a comparison of the feature amount possessed by each mode of invasion using the multiple comparison method of analysis of one-way ANOVA with Bonferroni correction. First, the reason why Grade 2 was misinterpreted as Grade 1 may be that the number of epithelial regions (feature 1)

was extremely close between Grade 1 and Grade 2, making it difficult to distinguish between them. Second, the reason why Grade 2 was misinterpreted as Grade 3 may be that not only the number of epithelial regions, represented by Feature 1, but also Feature 4, which indicates the sharpness of the protrusion of the tumor area, and Feature 8, which indicates the length of the borderline, were similar and were therefore difficult to distinguish. Third, the reason why Grade 2 was misinterpreted as Grade 4C may be that Feature 6, which indicates the cord-like feature, was not significantly different only between Grade 2 and Grade 4C, but was also significantly different between the other mode of invasion (Grade1, Grade3, Grade4D), making it impossible to distinguish between Grade 2 and Grade 4C.

This study was limited because hematoxylin- and eosin-stained (HE) images were primarily not used, despite the desirability of such an approach from the perspectives of cost and convenience. However, as this research involved the challenge of a first approach to this technology, we performed IHC to detect claudin-7 and uPA/uPAR, which specifically stains OSCC tumor cells, to further clarify the borderline between the tumor and the stroma and ensure clear binary images [15, 62]. The use of HE specimens alone would have made it particularly difficult to capture the sparsely scattered tumor cells in the stroma tissue of Grade 4D specimens. However, the inclusion of these IHC

analysis better facilitated the detection of tumor cells even in these Grade 4D cases [15]. In the future, we needed to ensure that machine learning can detect bivalence using more straightforward and more useful HE samples. To improve the classification accuracy using deep learning, it is necessary to include a substantially high number of cases; however, we did not have the required number of cases. Therefore, it was necessary to create a classifier in a limited number of cases captured by the expert clinicians. The good overall F value suggests that good feature values were extracted by them. Moreover, this approach might also be useful for constructing an automatic YK classification discrimination method, although the accuracy must be improved.

The strengths of this study are that the ROI was set using WSI, which is still uncommonly used in oral cancer research. In assessing the mode of invasion, it is problematic to have differences in the evaluation of subjective visual findings among raters. However, in this study, the advocate oneself guided the selection method and characteristics to the supervised images, and thus the discrepancy between raters was minimized. While pathological imaging studies using WSI have become the main method for classifying pathological tissues by machine learning, there are still very few reports of such studies using WSI in oral cancer[63]. A few studies have been reported using WSI to classify whether a tumor is cancerous or not in oral cancer [63, 64], but no

study has yet been reported to classify the nature of cancer as in this study. The reason for this may be that unlike cancers of other tissues, oral cancer contains hard tissues such as teeth and jawbone in the tissue specimen, so the tissue specimen itself is prone to wrinkles, folds, and tears, resulting in a low-quality specimen that is prone to errors when scanning. Similarly, in this study, tongue cancer was easy to evaluate by WSI, while cancer involving jawbone was difficult to evaluate by WSI. It is important to prepare good quality histopathological specimens for oral cancer in order to conduct research using WSI.

The accuracy of machine learning could potentially be improved by dramatically increasing the number of cases. Although many pathological image findings and clinical information can be obtained from The Cancer Genome Atlas database, this information is provided in a pathological image format and the contents are not uniform [65]. Consequently, it is difficult to apply these data in a machine learning setting. In the future, it will be necessary to collect a larger number of cases through a multi-center collaboration. Furthermore, increasing numbers of patients will benefit when clinicians and pathologists use a more effective AI system. In Japan, there are few pathologists and, in many cases, there are no in-hospital pathologists; therefore, intraoperative diagnosis may not be possible, and accurate surgery and treatment may not be possible.

Ideally, we would like to apply this research to clinical practice as soon as possible and utilize the computer-aided diagnosis system for the diagnosis of oral cancer. However, machine learning requires careful judgment and caution because it leaves the decision-making and prediction related to medical ethics to the machine.

In this study, we developed an automatic machine learning-based classifier system to discriminate the mode of invasion of OSCC. Notably, this classifier was confirmed to generate decisions similar to those made by a clinician. Our results suggest that an automatic medical diagnostic imaging system could feasibly and accurately determine the mode of OCC invasion. We should continue to cooperate with the field of AI analysis to develop diagnostic tools under medical ethics.

Acknowledgements

We greatly appreciate Dr Etsuhide Yamamoto, Professor emeritus of Kanazawa University, who gave us valuable advice regarding adapting the mode of invasion to the clinical images used in this study. We also thank Mr. Tatsuo Fukui for supporting this interdisciplinary integration research project of the University of Yamanashi.

Funding

This work was financially supported by a Grant-in-Aid for Scientific Research (JSPS KAKENHI Grand Number # 18K09807) from the Ministry of Education, Science, Sports and Culture of Japan. Funding was received from the interdisciplinary integration research project of the University of Yamanashi.

Availability of data and materials

All code used in the experiments described in the manuscript is available through our GitHub repository (https://github.com/YK-criteria/OSCC/).

Author's contributions: KY planned the study and drafted the paper with the help of AM. KY and SK collected the images and the clinicopathological data. HY conducted statistical analysis of this study. KY, KU, and HA designed the image feature extraction methods. YK and HA executed the experimental work with machine learning. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This study was approved by the ethics committees of Yamanashi University (approval number: 1267) and Kanazawa University (approval number: 1647-1). This study was

conducted in accordance with the Declaration of Helsinki. According to the ethical committee policy, all patient data were anonymized before use. Informed consent was obtained from each patient.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publication history

This manuscript was previously posted as preprint, in 6 August 2020, DOI: 10.21203/rs.3.rs-52099/v1" URL: https://www.researchsquare.com/article/rs-52099/v1.

Reference

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69:7-34.

2. Sano D, Myers JN. Metastasis of squamous cell carcinoma of the oral tongue. *Cancer Metastasis Rev.* 2007;26:645-662.

3. Silverman S, Jr. Demographics and occurrence of oral and pharyngeal cancers. The outcomes, the trends, the challenge. *Journal of the American Dental Association (1939).* 2001;132 Suppl:7s-11s.

4. Bell RB. Oral cancer breaks out at ASCO. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2015;120:421-423.

5. Bryne M, Koppang HS, Lilleng R, Kjaerheim A. Malignancy grading of the deep invasive margins of oral squamous cell carcinomas has high prognostic value. *J Pathol.* 1992;166:375-381.

6. Yoshizawa K, Nozaki S, Okamune A, et al. Loss of maspin is a negative prognostic factor for invasion and metastasis in oral squamous cell carcinoma. *J Oral Pathol Med.* 2009;38:535-539.

7. Michikawa C, Uzawa N, Kayamori K, et al. Clinical significance of lymphatic and blood vessel invasion in oral tongue squamous cell carcinomas. *Oral Oncol.* 2012;48:320-324.

8. Bryne M. Is the invasive front of an oral carcinoma the most important area for prognostication? *Oral diseases.* 2008;4:70-77.

9. Yamamoto E, Kohama G, Sunakawa H, Iwai M, Hiratsuka H. Mode of invasion, bleomycin sensitivity, and clinical course in squamous cell carcinoma of the oral cavity. *Cancer.* 1983;51:2175-2180.

10. Broders A. The microscopic grading of cancer. *Surg Clin North Am.* 1941;21:947-962.

11. Anneroth G, Hansen LS, Silverman S, Jr. Malignancy grading in oral squamous cell carcinoma. I. Squamous cell carcinoma of the tongue and floor of mouth: histologic grading in the clinical evaluation. *J Oral Pathol.* 1986;15:162-168.

12. Jakobsson PA, Eneroth CM, Killander D, Moberger G, Mårtensson B. Histologic classification and grading of malignancy in carcinoma of the larynx. *Acta Radiol Ther Phys Biol.* 1973;12:1-8.

13. Carlile A, Edwards C. Poorly differentiated squamous carcinoma of the bronchus: a light and electron microscopic study. *J Clin Pathol.* 1986;39:284-292.

14. Nozaki S, Endo Y, Kawashiri S, et al. Immunohistochemical localization of a

urokinase-type plasminogen activator system in squamous cell carcinoma of the oral cavity: association with mode of invasion and lymph node metastasis. *Oral Oncol.* 1998;34:58-62.

15. Yoshizawa K, Nozaki S, Kato A, et al. Loss of claudin-7 is a negative prognostic factor for invasion and metastasis in oral squamous cell carcinoma. *Oncol Rep.* 2013;29:445-450.

16. Harada T, Shinohara M, Nakamura S, Shimada M, Oka M. Immunohistochemical detection of desmosomes in oral squamous cell carcinomas: correlation with differentiation, mode of invasion, and metastatic potential. *International journal of oral and maxillofacial surgery.* 1992;21:346-349.

17. Dirven R, Ebrahimi A, Moeckelmann N, Palme CE, Gupta R, Clark J. Tumor thickness versus depth of invasion - Analysis of the 8th edition American Joint Committee on Cancer Staging for oral cancer. *Oral Oncol.* 2017;74:30-33.

18. Aaboubout Y, van der Toom QM, de Ridder MAJ, et al. Is the Depth of Invasion a Marker for Elective Neck Dissection in Early Oral Squamous Cell Carcinoma? *Front Oncol.* 2021;11:628320.

19. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *Jama.* 2017;318:2199-2210.

20. Arvaniti E, Fricker KS, Moret M, et al. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Scientific reports.* 2018;8:12054.

21. Kleppe A, Albregtsen F, Vlatkovic L, et al. Chromatin organisation and cancer prognosis: a pan-cancer study. *The Lancet. Oncology.* 2018;19:356-369.

22. Linder N, Taylor JC, Colling R, et al. Deep learning for detecting tumour-infiltrating lymphocytes in testicular germ cell tumours. *Journal of clinical pathology.* 2019;72:157-164.

23. Sapienza LG, Ning MS, Taguchi S, et al. Altered-fractionation radiotherapy improves local control in early-stage glottic carcinoma: A systematic review and meta-analysis of 1762 patients. *Oral Oncol.* 2019;93:8-14.

24. Marka A, Carter JB, Toto E, Hassanpour S. Automated detection of nonmelanoma skin cancer using digital images: a systematic review. *BMC Med Imaging.* 2019;19:21.

25. Basavanhally AN, Ganesan S, Agner S, et al. Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. *IEEE transactions on bio-medical engineering.*

2010;57:642-653.

26. Doyle S, Monaco J, Feldman M, Tomaszewski J, Madabhushi A. An active learning based classification strategy for the minority class problem: application to histopathology annotation. *BMC Bioinformatics.* 2011;12:424.

27. Yu E, Monaco JP, Tomaszewski J, Shih N, Feldman M, Madabhushi A. Detection of prostate cancer on histopathology using color fractals and Probabilistic Pairwise Markov models. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference.* 2011;2011:3427-3430.

28. Ciaccio EJ, Tennyson CA, Bhagat G, Lewis SK, Green PH. Classification of videocapsule endoscopy image patterns: comparative analysis between patients with celiac disease and normal individuals. *Biomed Eng Online.* 2010;9:44.

29. Trivizakis E, Ioannidis GS, Melissianos VD, et al. A novel deep learning architecture outperforming 'off-the-shelf' transfer learning and feature-based methods in the automated assessment of mammographic breast density. *Oncol Rep.* 2019;42:2009-2015.

30. Ariji Y, Fukuda M, Kise Y, et al. Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2019;127:458-463.

31. Ariji Y, Goto M, Fukano H, Sugita Y, Izumi M, Ariji E. Role of intraoral color Doppler sonography in predicting delayed cervical lymph node metastasis in patients with early-stage tongue cancer: a pilot study. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2015;119:246-253.

32. Corbella S, Srinivas S, Cabitza F. Applications of deep learning in dentistry. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2020.

33. Abdalla-Aslan R, Yeshua T, Kabla D, Leichter I, Nadler C. An artificial intelligence system using machine-learning for automatic detection and classification of dental restorations in panoramic radiography. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2020;130:593-602.

34. Shan J, Jiang R, Chen X, et al. Machine Learning Predicts Lymph Node Metastasis in Early-Stage Oral Tongue Squamous Cell Carcinoma. *Journal of Oral and Maxillofacial Surgery.* 2020;78:2208-2218.

35. Yoshizawa K, Nozaki S, Kitahara H, et al. Expression of urokinase-type plasminogen activator/urokinase-type plasminogen activator receptor and

maspin in oral squamous cell carcinoma: Association with mode of invasion and clinicopathological factors. *Oncol Rep.* 2011;26:1555-1560.

36. Lydiatt WM, Patel SG, O'Sullivan B, et al. Head and Neck Cancers-Major Changes in the American Joint Committee on Cancer Eighth Edition Cancer Staging Manual. *Ca-a Cancer Journal for Clinicians.* 2017;67:122-137.

37. Inoue A, Ando H, Ueki K, et al. Automoatic classification of mode of invasion (YK-crieteia) from digital images using machine learning with pathological specimens of oral squamous cell carcinoma. *MEDICAL IMAGING TECHNOLOGY.* 2016;34:279-285.

38. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Ieee Transactions on Pattern Analysis and Machine Intelligence.* 2002;24:971-987.

39. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research.* 2011;12:2825-2830.

40. Reinhard E, Ashikhmin M, Gooch B, Shirey P. Color transfer between images. *IEEE Comput. Graph. Appl.* 2001;21:34-41.

41. Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol.* 2001;23:291-299.

42. Magee D, Treanor D, Chomphuwiset P, Quirke P. Context aware colour classification in digital microscopy. *in Proc. Med. Image Undrstand. Anal.* 2010:1-5.

43. Khan AM, El-Daly H, Rajpoot N. RanPEC: Random projections with ensemble clustering for segmentation of tumor areas in breast histology images. *in Proc. Med. Image Understand. Anal.* 2012:17-23.

44. Martin DR, Fowlkes CC, Malik J. Learning to detect natural image boundaries using local brightness, color, and texture cues. *Ieee Transactions on Pattern Analysis and Machine Intelligence.* 2004;26:530-549.

45. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning.* 2002;46:389-422.

46. Gavin C. Cawley NLCT. Efficient leave-one-out cross-validation of kernel Fisher discriminant classifiers. *Pattern Recognition.* 2003;36:2585-2592.

47. Meijer RJ, Goeman JJ. Efficient approximate k-fold and leave-one-out cross-validation for ridge regression. *Biom J.* 2013;55:141-155.

48. Breiman L. Random forests. *Machine Learning.* 2001;45:5-32.

49. Feurer M, Hutter F. Hyperparameter Optimization: Springer International Publishing; 2019:3-33.

50. James B, Yoshua B. Random search for hyper-parameter optimization. *Journal of Machine Learning Research.* 2012;13:281-305.

51. Inselberg A, Dimsdale B. Parallel coordinates: a tool for visualizing multi-dimensional geometry: IEEE Comput. Soc. Press.

52. Klein Nulent TJW, Noorlag R, Van Cann EM, et al. Intraoral ultrasonography to measure tumor thickness of oral cancer: A systematic review and meta-analysis. *Oral Oncol.* 2018;77:29-36.

53. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging.* 2012;30:1234-1248.

54. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48:441-446.

55. Liu Q, Sun D, Li N, et al. Predicting EGFR mutation subtypes in lung adenocarcinoma using (18)F-FDG PET/CT radiomic features. *Transl Lung Cancer Res.* 2020;9:549-562.

56. Kim M, Jung SY, Park JE, et al. Diffusion- and perfusion-weighted MRI radiomics model may predict isocitrate dehydrogenase (IDH) mutation and tumor aggressiveness in diffuse lower grade glioma. *Eur Radiol.* 2020;30:2142-2151.

57. Hashido T, Saito S, Ishida T. A radiomics-based comparative study on arterial spin labeling and dynamic susceptibility contrast perfusion-weighted imaging in gliomas. *Scientific reports.* 2020;10:6121.

58. Ma W, Zhao Y, Ji Y, et al. Breast Cancer Molecular Subtype Prediction by Mammographic Radiomic Features. *Acad Radiol.* 2019;26:196-201.

59. Qi Y, Cui X, Han M, et al. Radiomics analysis of lung CT image for the early detection of metastases in patients with breast cancer: preliminary findings from a retrospective cohort study. *Eur Radiol.* 2020;30:4545-4556.

60. Romeo V, Cuocolo R, Ricciardi C, et al. Prediction of Tumor Grade and Nodal Status in Oropharyngeal and Oral Cavity Squamous-cell Carcinoma Using a Radiomic Approach. *Anticancer Res.* 2020;40:271-280.

61. Kaur H, Pannu HS, Malhi AK. A Systematic review on imbalanced data challenges in machine learning: Applications and solutions. *ACM Computing Surveys.* 2019;52:1-36.

62. Lourenço SV, Coutinho-Camillo CM, Buim ME, et al. Claudin-7 down-regulation is an important feature in oral squamous cell carcinoma. *Histopathology.* 2010;57:689-698.

63. Mahmood H, Shaban M, Indave BI, Santos-Silva AR, Rajpoot N, Khurram SA. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: A systematic review. *Oral Oncol.* 2020;110:104885.

64. Halicek M, Shahedi M, Little JV, et al. Head and Neck Cancer Detection in Digitized Whole-Slide Histology Using Convolutional Neural Networks. *Scientific reports.* 2019;9:14043.

65. Weinstein JN, Collisson EA, Mills GB, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013;45:1113-1120.

Figure Legends

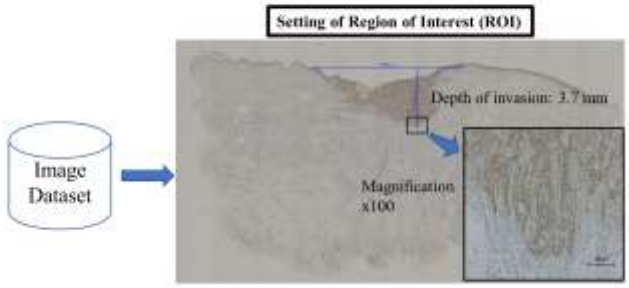

Fig. 1

Figure. 1. Setting of the Region of Interest (ROI) from whole slide image.

We searched the Region of Interest (ROI) of the invasive front at using whole slide images. The ROI site was expanded to ×100 to extract the features and determine the mode of invasion. To select ROI as region of deepest invasion, we establish the horizon (horizontal blue line: length of 9500 μm) that is at the level of the basement membrane relative to the closest intact squamous mucosa, and select ROI from the deepest invasion by dropping a "plumb line" (blue vertical line: length of 3700 μm) from the horizon. The black inserted line in the 100 x magnified image is the 150 μm as a scaler bar.

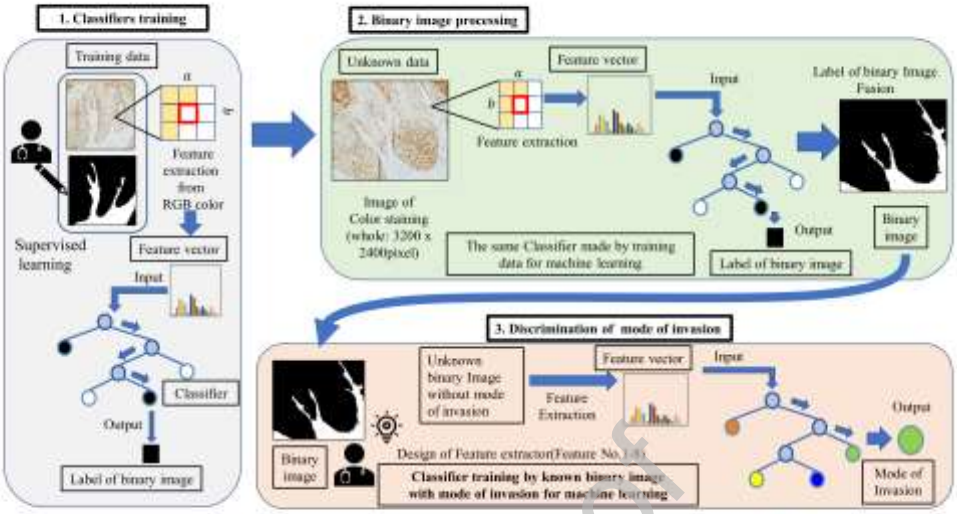

Fig. 2

Figure. 2. Image processing to differentiate the mode of invasion.

1. Classifiers training; 2. The binary image processing procedure; 3. The procedure used to discriminate the mode of invasion (Yamamoto–Kohama [YK] classification).

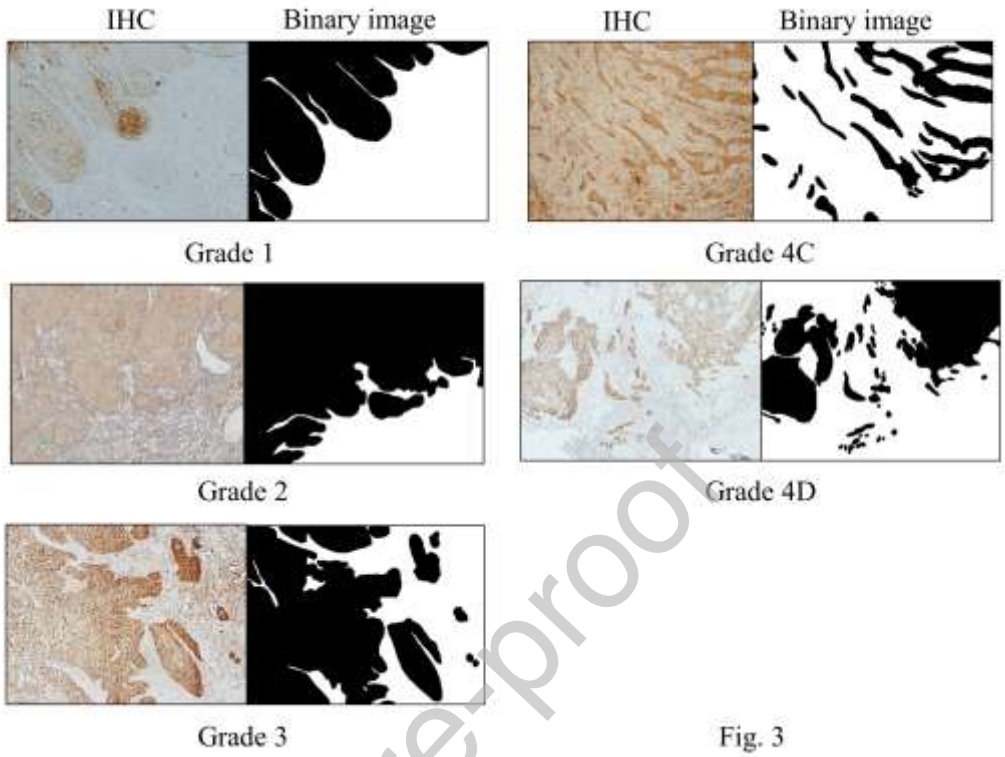

Figure 3. Typical binarized images of each Yamamoto–Kohama (YK) classification used in this research.

The border lines between the tumor and stromal tissues can be distinguished clearly up to a fine point and are clearly binarized on the immunohistochemistry images corresponding to tumors in each YK class.

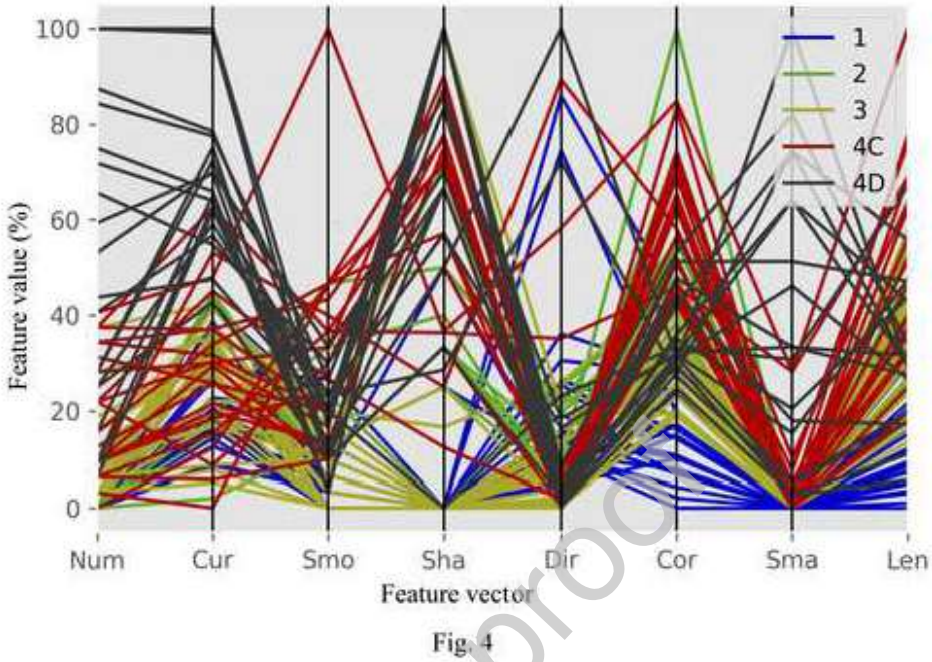

Fig. 4

Figure 4. Parallel coordinates based on Yamamoto–Kohama classification.

The parallel coordinates indicate the relative ratio of feature amount on the Y axis and indicate each of the feature domains (features No 1–8) on the X axis, based on each of mode of invasion (five types: Grades 1–4D).

Num: Number of tumor area as Future value 1; Cur: Curvature of tumor area as Future value 2; Smo: Smooth protrusion as Future value 3; Sha: Sharp protrusion as Future vale 4; Dir: Direction of tumor area as Future value 5; Cor: Cord shape of tumor area as Future value 6; Sma: Small area of tumor as Future value 7; Len: Length of the borderline as Future vale 8.

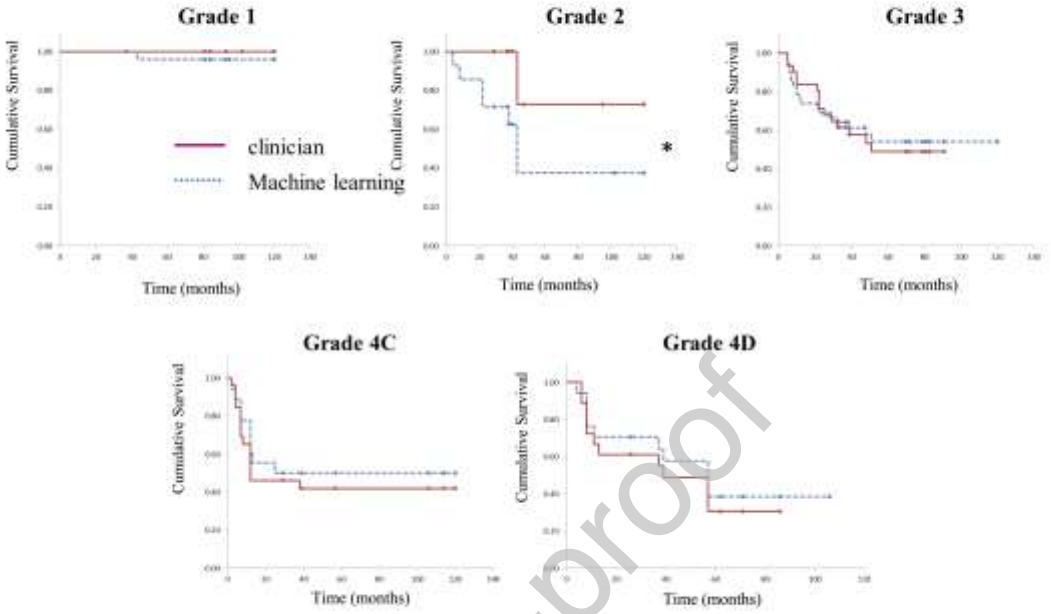

Fig. 5

Figure 5. Kaplan–Meier analysis of survival based on Yamamoto–Kohama criteria.

The survival rates based on the modes of invasion determined by the clinician and by machine learning are compared. * $P < 0.05$

Table 1. Yamamoto-Kohama classification.

Grade	Histologic grading
1	Well-defined borderline
2	Cords, less marked borderline
3	Groups of cells, no distinct borderline
4C	Diffuse invasion, Cord-like type
4D	Diffuse invasion, Widespread type

[a]Yamamoto-Kohama classification

Table 2. Hyperparameters of the random forests

Items of parameter	Numerical Value
Number of Trees	300
Maximum depth	10
Feature-number of by Random-selection	3
Number of minimum samples at leaf	2
Minimum information-gain	0.01

Table 3. Discrimination result with confusion matrix

		Discrimination result				
		Grade 1	Grade 2	Grade 3	Grade 4C	Grade 4D
Test data	Grade 1	21	2	0	0	0
	Grade 2	1	8	2	1	0
	Grade 3	0	1	25	1	0
	Grade 4C	0	1	2	17	1
	Grade 4D	0	0	0	1	17

Table 4. Precision–Recall

	Number of correct answers	Discrimination number	Number of matches	Precision	Recall	F-measure
Grade 1	23	22	21	0.95	0.91	0.93
Grade 2	12	12	8	0.67	0.67	0.67
Grade 3	27	29	25	0.86	0.93	0.89
Grade 4C	21	20	17	0.85	0.81	0.83
Grade 4D	18	18	17	0.94	0.94	0.94