Development of radiomics and machine learning model for predicting occult cervical lymph node metastasis in patients with tongue cancer

Katsumaro Kubo, MD[1], Daisuke Kawahara, PhD[1], Yuji Murakami, MD, PhD[1]*, Yuki Takeuchi, MD, PhD[1], Tsuyoshi Katsuta, MD[1], Nobuki Imano, MD, PhD[1], Ikuno Nishibuchi, MD, PhD[1], Akito Saito, PhD[1], Masaru Konishi, DDS, PhD[2], Naoya Kakimoto, DDS, PhD[3], Yukio Yoshioka, DDS, PhD[4], Shigeaki Toratani, DDS, PhD[4], Shigehiro Ono, DDS, PhD[5], Tsutomu Ueda, MD, PhD[6], Sachio Takeno, MD, PhD[6], and Yasushi Nagata, MD, PhD[1]

[1]Department of Radiation Oncology, Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima, 734-8553, Japan.

[2]Department of Oral and Maxillofacial Radiology, Hiroshima University Hospital, 1-2-3 Kasumi, Minami-ku, Hiroshima, 734-8551, Japan.

[3]Department of Oral and Maxillofacial Radiology, Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima, 734-8553, Japan.

[4]Department of Molecular Oral Medicine and Maxillofacial Surgery, Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima, 734-8553, Japan.

[5]Department of Oral and Maxillofacial Surgery, Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima, 734-8553, Japan.

[6]Departments of Otolaryngology and Head and Neck Surgery, Graduate School of Biomedical and

Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima, 734-8553, Japan.

*Corresponding author: Yuji Murakami

Department of Radiation Oncology, Graduate School of Biomedical and Health Sciences, Hiroshima University

1-2-3 Kasumi Minami-ku Hiroshima-shi, Hiroshima 734-8553, Japan

Telephone: +81-257-1545

Fax: +81-257-1546

E-mail: yujimura@hiroshima-u.ac.jp

Declarations of interest: None

Word count

Abstract: 195

Manuscript: 3437

References: 36

Figures: 5

Tables: 5

Abstract

Objective

We aimed to develop a predictive model for occult cervical lymph node metastasis (OCLNM) in patients with tongue cancer using radiomics and machine learning from pretreatment contrast-enhanced computed tomography (CT).

Study Design

This study included 161 patients with tongue cancer who received local treatment. CT images were transferred to a radiomics platform. The volume of interest was the total neck node level, including levels Ia, Ib, II, III, and IVa at the ipsilateral side, and each neck node level. The dimensionality of the radiomics features was reduced using least absolute shrinkage and selection operator (LASSO) logistic regression analysis. We compared five classifiers with or without the synthetic minority oversampling technique (SMOTE).

Results

For the analysis at the total neck node level, random forest with SMOTE was the best model with an accuracy of 0.85 and an area under the curve (AUC) score of 0.92. For the analysis at each neck node level, support vector machine with SMOTE was the best model with an accuracy of 0.96 and an AUC score of 0.98.

Conclusions

Predictive models using radiomics and machine learning have potential as clinical

decision-support tools in the management of patients with tongue cancer for prediction of OCLNM.

Introduction

Surgical operation and brachytherapy are standard local treatments for early stage tongue cancer.[1] Patients who do not have poor prognostic factors and undergo treatment with these modalities for tongue lesions that are small and have adequate treatment margins, are placed under observation if their neck is clinically and radiographically cancer-free.[1, 2] However, this observation period may risk the development of occult cervical lymph node metastasis (OCLNM), which may have been present during initial treatment but escaped detection even after careful examination[3].

OCLNM has been reported in 20%–50% of early stage tongue cancer cases, and the prognosis is poor.[3, 4, 5] Despite advances in imaging modalities, it is difficult to detect OCLNM in high-risk patients. Elective neck dissection (END) is an option that could potentially manage OCLNM.[6, 7] However, for patients without metastasis, END is an overtreatment.

Radiomics is the study of systemically handling large amounts of imaging information in radiology.[8] Radiomics has been successfully applied in the screening, diagnosis, treatment, and evaluation of multiple tumor types.[9, 10] Machine learning, which is the study of computer algorithms that improve automatically through experience, has also been applied in medical imaging analysis. In recent years, the introduction of radiomics and machine learning into medicine has become a major topic of discussion. Indeed, many studies have created

predictive models using radiomics and machine learning from medical imaging of head and neck cancers.[11, 12, 13]

Hence, it might be possible to develop a predictive model for OCLNM of tongue cancer through radiomics and machine learning using pretreatment computed tomography (CT). Detecting OCLNM with high accuracy could help avoid unnecessary surgery, such as END, in patients who are free of metastasis. Indeed, several studies have reported predictive models for OCLNM in patients with tongue cancer using radiomics and machine learning.[14, 15, 16] The performance is good, but the data regarding these modalities is still limited.

Therefore, we aimed to develop a predictive model for OCLNM in patients with tongue cancer using radiomics and machine learning from pretreatment contrast-enhanced CT.

Materials and methods

Patients

Data from patients diagnosed with tongue cancer who underwent local treatment without additional therapy, such as END and chemotherapy, at Hiroshima University Hospital (Hiroshima, Japan) between October 2008 and July 2019 were retrospectively analyzed. Individuals who fulfilled the following criteria were included in the present study: received a histological diagnosis of squamous cell carcinoma of the tongue; underwent contrast-enhanced CT before treatment; were negative for cervical lymph node metastasis

before treatment; and underwent surgical operation or brachytherapy without additional treatment such as END, external beam radiation therapy to the neck lymph node, and chemotherapy as initial treatment. OCLNM was defined as cervical lymph node metastasis that was not detected by careful examinations including ultrasonography and radiographic studies before treatment but was evident after initial treatment without recurrence of the primary tumor within one year after initial treatment. Analysis was limited to one year after initial treatment because it was unlikely that lymph node metastases that were not identified at initial therapy would become apparent more than one year after treatment.

All patients with OCLNM underwent salvage surgery and were histologically proven to have metastatic squamous cell carcinoma. Patients with OCLNM who did not undergo salvage surgery and were not histologically proven to have metastatic squamous cell carcinoma were excluded. Patients without OCLNM were defined as those with no recurrence of cervical lymph node metastasis for at least 2 years after surgical operation or brachytherapy for tongue cancer. Among the patients without OCLNM, those without enough follow-up were excluded. Therefore, patients with a follow-up duration of less than 2 years without cervical lymph node metastasis were excluded from this study. The clinical TNM stage was defined according to the Tumor Node Metastasis classification (Union for International Cancer Control, 7th Edition). The study was approved by the Human Ethics Review Committee of Hiroshima University Hospital (E-1656). The need for informed consent was

waived owing to the retrospective nature of the research.

Segmentation

Delineation of neck node levels

Pretreatment contrast-enhanced CT images were available for 161 eligible patients who had had scans performed on an Aquilion One CT unit (Canon Medical Systems, Otawara, Tochigi, Japan). CT scans were of diagnostic quality and performed using tube voltage of 120 kV, tube current of auto exposure control (noise index 6), 1.3–5 mm slice thickness, and intravenous administration of non-ionic iodinated contrast material (500 mgI/kg at a rate of 1.5 mL/s). All CT images were imported into the Pinnacle Radiotherapy Treatment Planning System (Phillips Medical Systems, Fitchburg, WI, USA).

For all CT scans, each neck node level was contoured slice-by-slice in the axial plane on the Pinnacle Radiotherapy Treatment Planning System, referring to the 2013 update of the guidelines on neck nodal level delineation for head and neck tumors.[17] Contouring was performed by two radiation oncologists without the use of software. Neck node levels included levels Ia, Ib, II, III, and IVa on the ipsilateral side of the neck. Contouring of neck node levels was divided between the two radiation oncologists for each case. Review after contouring was performed by the two initial radiation oncologists plus another radiation oncologist. All physicians had 10 years or more of professional experience.

Contouring was performed at each neck node level instead of the primary tumor

because the objective was to detect evidence of OCLNM. In addition, it was sometimes difficult to contour the primary tumor accurately due to artifacts or small tumor size. Supplementary Table S1 shows anatomical level boundaries and Figure 1 depicts the contouring of neck node levels in this study. We analyzed the total neck node levels, including levels Ia, Ib, II, III, and IVa, and each neck node level, separately. For the analysis of total neck node level, 161 segments (1 for each case) were examined for the subsequent process. For the analysis at each neck node level, a total of 805 segments (1 at each level for a total of 5 for each case) were used. Analysis of each neck node level was added to analysis of the total neck in the expectation that it would improve the accuracy in the diagnosis of OCLNM.

The total neck node level for each patient was labeled "positive" or "negative" for OCLNM based on the presence of at least one metastasis in any level (Figure 2a). Each neck node level was labeled individually as "positive" or "negative" for OCLNM based on the presence of at least one metastasis in that level (Figure 2b).

Image processing

Radiomics features were standardized using z-score normalization because of the variation in CT image protocols.[18]

Feature extraction

The creation of radiomics features was performed using an open-source package in Python (Pyradiomics software).[19] The following features were created: first-order statistical features;

shape-based features; and texture analysis features including gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM), and neighborhood gray tone difference matrix (NGTDM). The higher order statistical features used a wavelet imaging filter. The wavelet filter has low-pass (L) and high-pass (H) filters. The decompositions were constructed in the x, y, and z directions. For example, "wavelet-HLL" was interpreted as a wavelet subband image resulting from directional filtering with a high-pass filter along the x-direction (H), a low-pass filter along the y-direction (L), and a low-pass filter along the z-direction (L).

Dimension reduction

The predictive model was established by combining the extracted features through feature de-redundancy and dimensionality reduction, preconditioning, and machine learning-based classification. The least absolute shrinkage and selection operator (LASSO) logistic regression analysis with MATLAB code[20, 21] was used to reduce dependency and redundancy.

Model development

We used various machine learning techniques to develop five models to predict the occurrence of OCLNM in patients with tongue cancer: k-nearest neighbor (kNN),[22] support vector machine (SVM),[23] classification and regression trees (CART),[24] random forest (RF),[25] and Ada Boost (Ada).[26] A 10-fold cross-validation was used to avoid overfitting. The prepared datasets incurred a data imbalance problem. To address this concern, we applied the synthetic

minority oversampling technique (SMOTE).[27]

k-nearest neighbor

kNN is a simple machine learning method. The kNN classifier classifies unlabeled observations by assigning them to the class of the most similar labeled examples. The characteristics of observations are collected for both training and test datasets.[22]

Support vector machine

SVM is a widely used, supervised learning approach for classification or regression analysis. It can be applied to transform training data into a high-dimensional feature space and determine a linear optimal solution by separating a hyperplane that provides the smallest distance between the hyperplane points and the largest margin between the classes.[23]

Classification and regression trees

CART is a rule-based method that generates a binary tree through binary recursive partitioning that splits a subset (called a leaf) of the data set into two subsets (called sub-leaves) according to the minimization of a heterogeneity criterion computed on the resulting sub-leaves. Each split is based on a single variable; some variables may be used several times, while others may not be used at all.[24]

Random forest

RF is an ensemble supervised learning method composed of multiple decision trees corresponding to various subdatasets. Each tree calculates the results and obtains the average

of the prediction outcomes. This approach allows for the reduction of variance in decision trees.[25]

Ada Boost

Ada is an ensemble learning algorithm used to elevate a weak classifier into a strong one. First, it trains a base classifier and assigns higher weights to the misclassified samples; thereafter, it is applied to the next process. This iterative process continues until a stop condition is reached, or the error rate becomes sufficiently small.[26]

Model evaluation

Accuracy, precision, and recall are significant indices for evaluating the performance of each model. These parameters are defined as follows:

$$Accuracy = \frac{True\ positives + True\ negatives}{True\ positives + True\ negatives + False\ positives + False\ negatives}$$

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}$$

Receiver operator characteristic curves were created to compare the performance of each model. The optimal cutoff of significant variables for each model was calculated by the Youden index in the ROC curve analysis. The model with the highest accuracy was designated as the best model. The data analysis workflow is shown in Figure 3.

Results

Clinical characteristics

Patient and tumor characteristics of the 161 patients enrolled in this investigation are summarized in Table I. Of the total number of patients, 46 (28.6%) had OCLNM, with metastasis detected at level Ia in 3 patients (6.5%), level Ib in 24 (52.2%), level II in 25 (54.3%), level III in 8 (17.4%), and level IVa in 3 (6.5%). The median time of neck node recurrence was 3.5 months (range, 0–12 months) after initial treatment. The median follow-up at the time of evaluation was 56 months (range, 9–131 months) in the survivors.

Model performance

Radiomic sets were built after LASSO analysis. Seven features were selected for the analyses at both the total neck node level and each neck node level (Table II). For the analysis of the total neck node level, there were two first-order statistical features and five texture analysis features (GLCM, $n = 5$). For the analysis at each neck node level, there were one first-order statistical feature, two shape-based features, and four texture analysis features (GLRLM $n = 1$; GLSZM $n = 3$). Table III and Figures 4 and 5 demonstrate the performance of the considered predictive models. Five classifiers, with or without SMOTE, were used to obtain the best model based on the feature set. For the analysis at the total neck node level before applying SMOTE, the SVM model achieved the best performance with the accuracy of 0.71. The precision, recall, and an area under the curve (AUC) score were 0.77, 0.69, and 0.72 respectively. After applying SMOTE, the performance of all models increased, and the RF

model was the best model with the accuracy of 0.85. The precision, recall, and an AUC score were 0.88, 0.82, and 0.92 respectively. For the analysis at each neck node level before applying SMOTE, the RF model achieved the best performance with the accuracy of 0.70. The precision, recall, and an AUC score were 0.84, and 0.66, and 0.65 respectively. After applying SMOTE, the SVM model was the best model with the accuracy of 0.96. The precision, recall, and an AUC score were 0.96, 0.95, and 0.98 respectively.

Discussion

OCLNM is one of the most serious problems after local treatment for early tongue cancer. Various studies have reported occult lymph node metastasis as a predictive factor for poor outcomes.[3, 4, 5] Despite advances in imaging modalities, high-risk patients with OCLNM are unidentifiable. END is one treatment option that has the potential to manage occult lymph node metastasis. However, over half of the patients with early tongue cancer do not need END, which would result in overtreatment and treatment-related toxicities. Published investigations of END have yielded both positive and negative results.[6, 7, 28] Hence, the efficacy of END remains controversial.

Recently, radiomics and machine learning have been applied to medical research, particularly in the context of head and neck cancers.[11, 12, 13] Several studies reported machine learning models to predict occult lymph node metastasis. The best model performances, listed

in Table IV, resulted in accuracy ranging from 0.74 to 0.95 and AUC scores ranging from 0.80 to 0.96.[14, 15, 16, 29, 30, 31] In this study, we developed predictive models for OCLNM in patients with tongue cancer using radiomics and machine learning from pretreatment contrast-enhanced CT. RF with SMOTE for the analysis of the total neck node level showed the best prediction accuracy of 0.85 with an AUC score of 0.92. SVM with SMOTE for the analysis at each neck node level showed the best prediction accuracy of 0.96 with an AUC score of 0.98. Although our results demonstrated similar efficacy compared to previous reports, the analysis at each neck node level was probably more useful than that of the total neck node level because of the high accuracy and the possibility that it could identify the level where occult lymph node metastasis occurred.

The major difference between previous reports and our study was that we chose the neck node level instead of the primary tumor as the volume of interest. In tongue cancer, when the primary tumor itself is the volume of interest, artifacts or small tumor size often prevent accurate delineation of the tumor. On the other hand, in the case of neck node level, the contouring method has been established, and there seemed to be fewer differences among individuals. Therefore, this study was meaningful in indicating that the predictive model for OCLNM using radiomics and machine learning has potential as a clinical decision-support tool.

Many studies have shown the relationship between occult lymph node metastasis and

tumor depth and thickness. However, the cutoff depth of invasion (DOI) for the prediction of occult lymph node metastasis varies widely among reports, ranging from 3 to 10 mm. [32, 33, 34] The National Comprehensive Cancer Network guidelines recommend the positive consideration of END when the thickness is greater than 4 mm and the evaluation of indications for END when the thickness is 2-4 mm.[1] D'Cruz et al. reported that the percentage of patients with metastasis was 5.6% when the DOI in pathological specimens was 3 mm and 16.9% when it was 4 mm.[7] Indeed, there is a relationship between tumor thickness and occult lymph node metastasis; if the thickness is above a certain level, the recommendation for END may be considered. However, if all patients with tumor thickness greater than 4 mm underwent END for suspected cervical lymph node metastases, there would be a risk of overtreatment in most cases due to the high false positive rate. Our model, using radiomics and machine learning, showed high prediction accuracy for cervical lymph node metastasis. Therefore, this model could potentially help avoid unnecessary surgery among patients with tongue cancer and select appropriate cases for END.

When the datasets are imbalanced, accuracy can be misleading. In this study, as reported in other studies, the proportion of cases positive for occult lymph node metastases was lower than that of cases negative for it, resulting in an imbalance. SMOTE is an approach to construct classifiers from imbalanced datasets and is an oversampling technique that allows the generation of synthetic samples for minority categories.[27] Several studies have reported

the validity of SMOTE in unbalanced datasets.[35, 36] We hypothesized that SMOTE could address the problem of the imbalanced dataset, and as a result, the performance of all models was improved with the use of this technique. Thus, SMOTE is a useful method to account for unbalanced data and can improve the capability of the models.

Our study had certain limitations, including its retrospective, single-institution design and limited sample size. In this study, not all patients were evaluated for tumor depth. The models predicted only ipsilateral neck node metastasis that was evident within one year after initial treatment, and did not include contralateral neck node metastasis. In addition, the time frame from acquisition of CT to treatment varied (median, 20 days; range, 1–132 days). Moreover, the imbalanced datasets were subjected to equalization processing using SMOTE. To overcome these limitations, a larger cohort would be needed for further validation of the model. Furthermore, because of this study design, individual factors related to occult lymph node metastasis were unknown. Finally, these limitations may have introduced potential biases. However, while the clinical application of radiomics and machine learning is advancing, reports of predictive models of OCLNM in patients with tongue cancer using these methods are limited. In addition, our models are useful because they may help predict OCLNM through simple methods using pretreatment CT. Compared to DOI, which has no established method of evaluation, CT-based contouring of lymph node levels is likely to result in fewer differences among individuals and institutions. Therefore, despite the limitations, this

study provides useful information regarding the predictive model of OCLNM using radiomics and machine learning. Future research in the form of prospective studies are warranted to test this model in actual clinical practice.

Conclusions

We designed a predictive model for occult cervical lymph node metastasis in patients with tongue cancer using radiomics and machine learning from pretreatment contrast-enhanced CT. The models showed remarkable prediction accuracy, and SVM with SMOTE showed the best prediction accuracy of 0.96 with an AUC score of 0.98 which was comparable to other reports. The predictive model of occult cervical lymph node metastasis using radiomics and machine learning has potential as a clinical decision-support tool.

Acknowledgments: None

References

1. National Comprehensive Cancer Network Inc NCCN Clinical Practice Guidelines in Oncology. Head and Neck Cancers version 2. 2020. Available from URL: https://www.nccn.org/professionals/physician_gls/pdf/head-and-neck.pdf
2. Machiels JP, René Leemans C, Golusinski W, Grau C, Licitra L, Gregoire V; EHNS Executive Board. Electronic address: secretariat@ehns.org; ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org; ESTRO Executive Board. Electronic address: info@estro.org. Squamous cell carcinoma of the oral cavity, larynx, oropharynx and hypopharynx: EHNS-ESMO-ESTRO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020; 31: 1462–1475.
3. Imai T, Satoh I, Matsumoto K, et al. Retrospective observational study of occult cervical lymph-node metastasis in T1N0 tongue cancer. *Jpn J Clin Oncol*. 2017; 47: 130–136.
4. Shibuya H, Hoshina M, Takeda M, Matsumoto S, Suzuki S, Okada N. Brachytherapy for stage I & II oral tongue cancer: an analysis of past cases focusing on control and complications. *Int J Radiat Oncol Biol Phys*. 1993; 26: 51–58.
5. Fujita M, Hirokawa Y, Kashiwado K, et al. Interstitial brachytherapy for stage I and II squamous cell carcinoma of the oral tongue: factors influencing local control and soft tissue complications. *Int J Radiat Oncol Biol Phys*. 1999; 44: 767–775.
6. Fasunla AJ, Greene BH, Timmesfeld N, Wiegand S, Werner JA, Sesterhenn AM. A

meta-analysis of the randomized controlled trials on elective neck dissection versus therapeutic neck dissection in oral cavity cancers with clinically node-negative neck. *Oral Oncol*. 2011; 47: 320–324.

7. D'Cruz AK, Vaish R, Kapre N, et al. Elective versus Therapeutic Neck Dissection in Node-Negative Oral Cancer. *N Engl J Med*. 2015; 373: 521–529.
8. 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.
9. Thawani R, McLane M, Beig N, et al. Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer.* 2018; 115: 34–41.
10. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. *Breast Cancer Res Treat*. 2018; 169: 217–229.
11. Peng Z, Wang Y, Wang Y, et al. Application of radiomics and machine learning in head and neck cancers. *Int J Biol Sci*. 2021; 17: 475–486.
12. Giraud P, Giraud P, Gasnier A, et al. Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers. *Front Oncol*. 2019; 9: 174.
13. Elhalawani H, Lin TA, Volpe S, et al. Machine Learning Applications in Head and Neck Radiation Oncology: Lessons From Open-Source Radiomics Challenges. *Front Oncol*. 2018; 8: 294.
14. Forghani R, Chatterjee A, Reinhold C, et al. Head and neck squamous cell carcinoma:

prediction of cervical lymph node metastasis by dual-energy CT texture analysis with machine learning. *Eur Radiol.* 2019; 29: 6172–6181.

15. Yuan Y, Ren J, Tao X. Machine learning-based MRI texture analysis to predict occult lymph node metastasis in early-stage oral tongue squamous cell carcinoma. *Eur Radiol*. 2021; 31: 6429–6437.
16. Zhong YW, Jiang Y, Dong S, et al. Tumor radiomics signature for artificial neural network-assisted detection of neck metastasis in patient with tongue cancer. *J Neuroradiol*. 2021 Aug 4. doi: 10.1016/j.neurad.2021.07.006.
17. Grégoire V, Ang K, Budach W, et al. Delineation of the neck node levels for head and neck tumors: a 2013 update. DAHANCA, EORTC, HKNPCSG, NCIC CTG, NCRI, RTOG, TROG consensus guidelines. *Radiother Oncol*. 2014; 110: 172–181.
18. Haga A, Takahashi W, Aoki S, et al. Standardization of imaging features for radiomics analysis. *J Med Invest*. 2019; 66: 35–37.
19. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics system to decode the radiographic phenotype. Cancer Res. 2017; 77: e104.
20. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997; 16: 385–395.
21. Zhang JX, Song W, Chen ZH, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol*. 2013;

14: 1295 –1306.

22. Zhang Z. Introduction to machine learning: k-nearest neighbors. *Ann Transl Med*. 2016; 4: 218.

23. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995; 20: 273–297.

24. Bel L, Allard D, Laurent JM, Cheddadi R, Bar-Hen A. CART algorithm for spatial data: application to environmental and ecological data. Comput. *Stat. Data Anal*. 2009; 53: 3082–3093.

25. Breiman L. Random Forests. *Mach Learn*. 2001; 45: 5–32.

26. Hastie T, Rosset S, Zhu J, Zou H. Multi-class adaboost. *Stat. Its Interface.* 2009; 2: 349–360.

27. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002; 16: 321–357.

28. Bulsara VM, Worthington HV, Glenny AM, Clarkson JE, Conway DI, Macluskey M. Interventions for the treatment of oral and oropharyngeal cancers: surgical treatment. *Cochrane Database Syst Rev*. 2018; 12: CD006205.

29. Bur AM, Holcomb A, Goodwin S, et al. Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma. *Oral Oncol*. 2019; 92: 20–25.

30. Shan J, Jiang R, Chen X, et al. Machine Learning Predicts Lymph Node Metastasis in Early-Stage Oral Tongue Squamous Cell Carcinoma. *J Oral Maxillofac Surg*. 2020; 78:

2208–2218.

31. Kwak MS, Eun YG, Lee JW, et al. Development of a machine learning model for the prediction of nodal metastasis in early T classification oral squamous cell carcinoma: SEER-based population study. *Head Neck*. 2021; 43: 2316–2324.
32. Matos LL, Manfro G, Santos RV, et al. Tumor thickness as a predictive factor of lymph node metastasis and disease recurrence in T1N0 and T2N0 squamous cell carcinoma of the oral tongue. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2014; 118: 209–217.
33. Haraguchi K, Yoshiga D, Oda M, et al. Depth of invasion determined by magnetic resonance imaging in tongue cancer can be a predictor of cervical lymph node metastasis. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2021; 131: 231–240.
34. Baba A, Masuda K, Hashimoto K, et al. Correlation between the magnetic resonance imaging features of squamous cell carcinoma of the buccal mucosa and pathologic depth of invasion. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2021; 131: 582–590.
35. Sun T, Zhang R, Wang J, Li X, Guo X. Computer-aided diagnosis for early-stage lung cancer based on longitudinal and balanced data. *PLoS One*. 2013; 8: e63559.
36. Yan S, Qian W, Guan Y, Zheng B. Improving lung cancer prognosis assessment by incorporating synthetic minority oversampling technique and score fusion method. *Med Phys*. 2016; 43: 2694–2703.

Figure legends

Figure 1. Contouring of each neck node level.

Blue, green, red, yellow, and purple lines show levels Ia, Ib, II, III, and IVa, respectively, in the ipsilateral neck side. The total neck node level consisted of these levels.

Figure 2. Data labeling.

a) Analysis in total neck node level. The total neck node level for each patient was labeled "positive" or "negative" for occult cervical lymph node metastasis (OCLNM) based on the presence of at least one metastasis in any level.

b) Analysis in each neck node level. Each neck node level was labeled individually as "positive" or "negative" for OCLNM based on the presence of at least one metastasis in that level.

Figure 3. Data analysis workflow.

CT, computed tomography; LASSO, least absolute shrinkage and selection operator; kNN, k-nearest neighbor; SVM, support vector machine; CART, classification and regression trees; RF, random forest; Ada, Ada Boost; SMOTE, synthetic minority oversampling technique.

Figure 4. Analysis of the total neck node level. Receiver operator characteristic (ROC) curves with accuracy and an area under the curve (AUC) scores for each model with or without the synthetic minority oversampling technique (SMOTE).

a) k-nearest neighbor (kNN)

b) Support vector machine (SVM)

c) Classification and regression trees (CART)

d) Random forest (RF)

e) Ada boost (Ada)

Figure 5. Analysis of each neck node level. Receiver operator characteristic (ROC) curves with accuracy and an area under the curve (AUC) scores for each model with or without the synthetic minority oversampling technique (SMOTE).

a) k-nearest neighbor (kNN)

b) Support vector machine (SVM)

c) Classification and regression trees (CART)

d) Random forest (RF)

e) Ada boost (Ada)

Table 1. Patient and tumor characteristics.

	N = 161	100%
Age, years, median (range)	65 (22–91)	–
Sex		
Male	80	49.7
Female	81	50.3
Location		
Right	83	51.6
Left	78	48.4
Clinical T category		
Tis	28	17.4
T1	45	28.0
T2	81	50.3
T3	7	4.3
Ulcer		
Yes	38	23.6
No	122	75.8
Not available	1	0.6
Smoking history		
Yes	75	46.6
No	73	45.3
Not available	13	8.1
Treatment		
Surgery	102	63.4
Brachytherapy	59	36.6
Occult neck metastasis		
Yes	46	28.6
No	115	71.4
Level of occult neck metastasis	*N* = 46	
Ia	3	6.5
Ib	24	52.2
II	25	54.3
III	8	17.4
IVa	3	6.5

Table 2. Extracted features after the least absolute shrinkage and selection of operator logistic regression analysis.

Extracted Features after LASSO logistic regression analysis			
Total neck node level		Each neck node level	
First-order statistical feature	Original_firstorder_Median	First-order statistical feature	Wavelet-HHL_firstorder_Median
	Wavelet-LHH_firstorder_Kurtosis	Shape-based feature	Original_shape_Maximum2DDiameterSlice
Texture analysis feature	Wavelet-HLL_glcm_Imc1		Original_shape_Maximum2DDiameterRow
	Wavelet-HLL_glcm_MaximumProbability	Texture analysis feature	Wavelet-LLH_glrlm_RunLengthNonUniformity
	Wavelet-HHL_glcm_Correlation		Wavelet-HLH_glszm_GrayLevelNonUniformity
	Wavelet-HHH_glcm_ClusterShade		Wavelet-HLH_glszm_GrayLevelNonUniformityNormalized
	Wavelet-LLL_glcm_Imc1		Wavelet-HLH_glszm_SmallAreaLowGrayLevelEmphasis

LASSO, least absolute shrinkage and selection operator; glcm, gray level co-occurrence matrix; Imc1, informational measure of correlation 1; glrlm, gray level run length matrix; glszm, gray level size zone matrix.

Table 3. Prediction performance of all models with or without the synthetic minority oversampling technique.

Total neck node level					Each neck node level				
Model	Accuracy	Precision	Recall	AUC	Model	Accuracy	Precision	Recall	AUC
kNN	0.63	0.79	0.59	0.60	kNN	0.42	0.83	0.45	0.61
SVM	0.71	0.77	0.69	0.72	SVM	0.55	0.85	0.53	0.54
CART	0.59	0.77	0.56	0.61	CART	0.58	0.85	0.55	0.58
RF	0.67	0.78	0.63	0.75	RF	0.70	0.84	0.66	0.65
Ada	0.64	0.79	0.60	0.67	Ada	0.65	0.85	0.60	0.65
kNN + SMOTE	0.71	0.78	0.68	0.78	kNN + SMOTE	0.75	0.79	0.73	0.80
SVM + SMOTE	0.80	0.86	0.76	0.84	SVM + SMOTE	0.96	0.96	0.95	0.98
CART + SMOTE	0.74	0.83	0.70	0.77	CART + SMOTE	0.75	0.77	0.73	0.82
RF + SMOTE	0.85	0.88	0.82	0.92	RF + SMOTE	0.92	0.95	0.89	0.97
Ada + SMOTE	0.81	0.88	0.77	0.90	Ada + SMOTE	0.86	0.91	0.83	0.93

AUC, area under the curve; kNN, k-nearest neighbor; SVM, support vector machine; CART, classification and regression trees; RF, random forest; Ada, Ada Boost; SMOTE, synthetic minority oversampling technique

Table 4. Best prediction performance in other studies

	N	Data	Model	Accuracy	AUC	Sensitivity	Specificity
Bur AM et al[29]	782	Clinicopathologic data	Decision forest	NA	0.84	0.92	0.58
Shan J et al[30]	145	Clinicopathologic data	SVM	NA	0.96	1	0.88
Kwak MS et al[31]	16878	Clinicopathologic data	XGBoost	0.95	0.96	0.91	0.86
Forghani R et al[14]	64	CT (primary tumor)	Random forest	0.88	NA	1	0.67
Yuan Y et al[15]	116	MRI (primary tumor)	Naïve Bayes	0.74	0.80	0.63	0.82
Zhong YW et al[16]	1402	CT (primary tumor)	cNRAD99	0.88	0.94	0.85	0.91
This study	161	CT (neck node level)	SVM + SMOTE	0.96	0.98	0.96	0.95

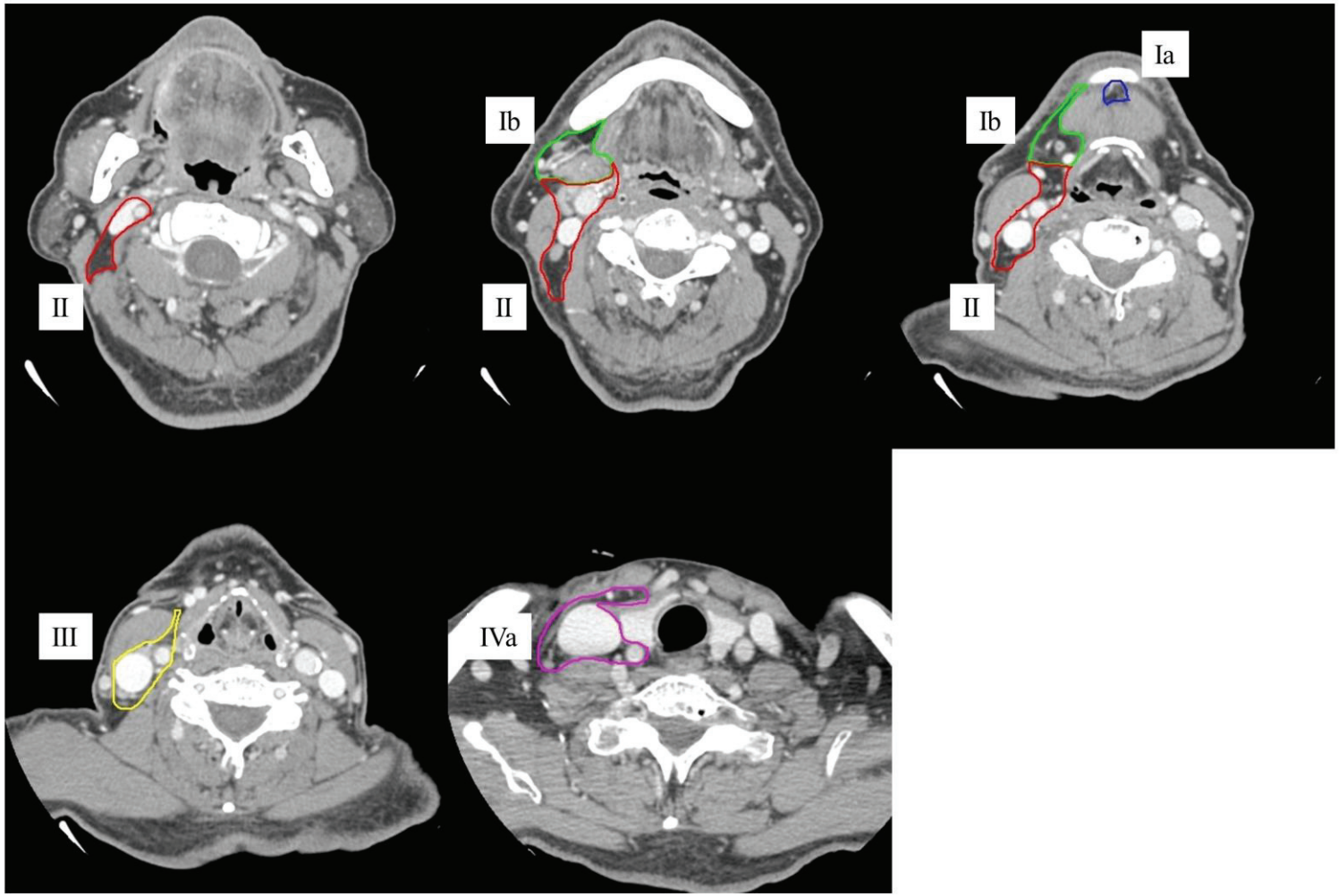
AUC, area under the curve; NA, not available; SVM, support vector machine; CT, computed tomography; MRI, magnetic resonance imaging; SMOTE, synthetic minority oversampling technique

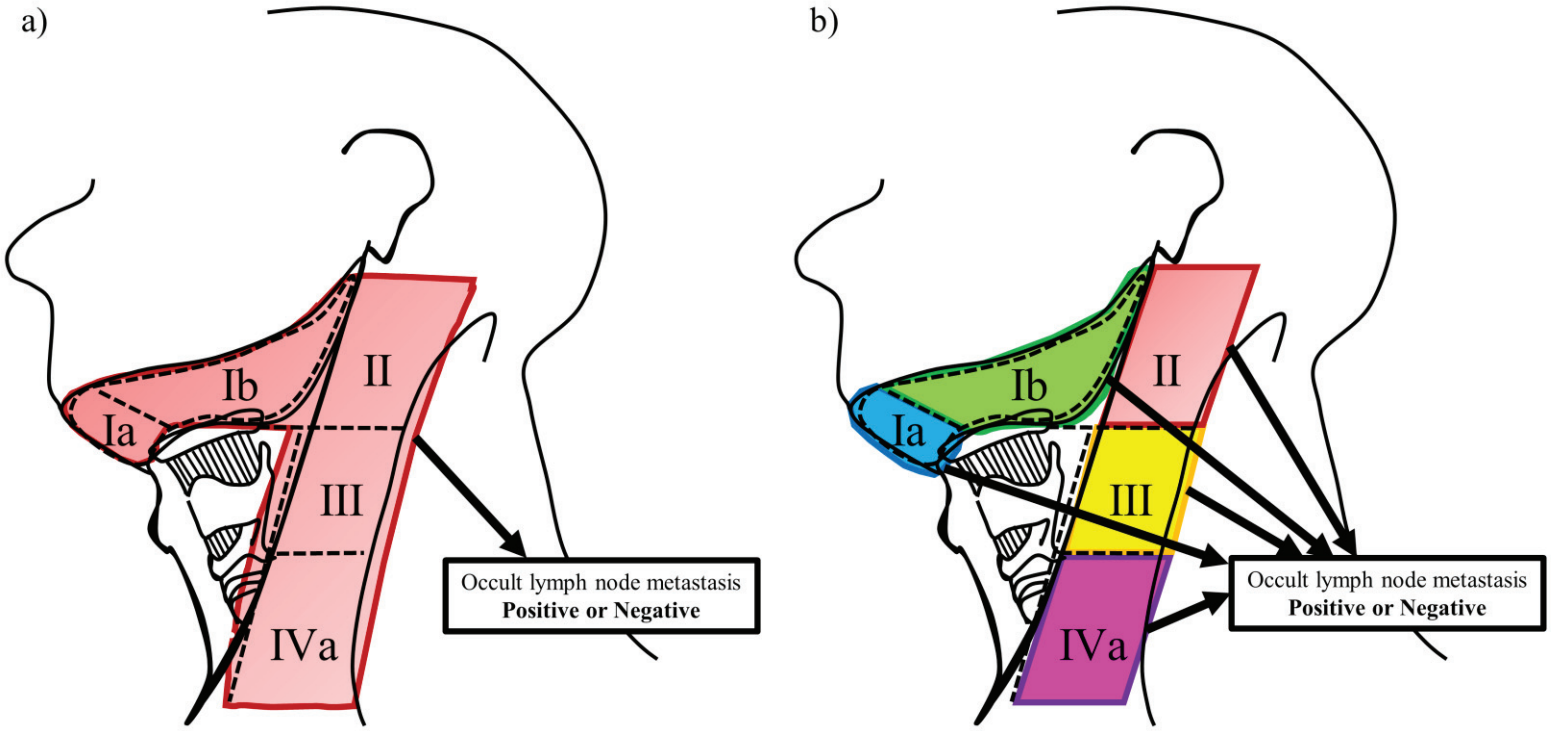
Table 1. Node levels Ia, Ib, II, III, and IVa.

	Cranial	Caudal	Anterior	Posterior	Lateral	Medial
Ia	Mylo-hyoid m.	Platysma m. (caudal edge of the anterior belly of digastric mm)	Symphysis menti	Body of hyoid bone/mylo-hyoid m.	Medial edge of ant. belly of digastric m.	NA
Ib	Cranial edge of submandibular gland; anteriorly, mylo-hyoid m.	Plane through caudal edge of hyoid bone and caudal edge of mandible; alternatively caudal edge of submandibular gland (whichever is more caudal)/platysma m.	Symphysis menti	Posterior edge of submandibular gland (caudally)/posterior belly of digastric m. (cranially)	Medial aspect (innerside) of mandible down to caudal edge/platysma m. (caudal)/medial pterygoid m. (posteriorly)	Lateral edge of ant. belly of digastric m. (caudally)/posterior belly of digastric m. (cranially)
II	Caudal edge of the lateral process of C1	Caudal edge of the body of the hyoid bone	Posterior edge of the submandibular gland/posterior edge of posterior belly of digastric m.	Posterior edge of sternocleidomastoid m.	Deep (medial) surface of sternocleidomastoid m./platysma m./parotid gland/posterior belly of digastric m.	Medial edge of internal carotid artery/scalenius m.
III	Caudal edge of the body of the hyoid bone	Caudal edge of cricoid cartilage	Anterior edge of sternocleidomastoid m./posterior third of thyro-hyoid m.	Posterior edge of sternocleidomastoid m.	Deep (medial) surface of sternocleidomastoid m.	Medial edge of common carotid artery/scalenius mm.
IVa	Caudal edge of cricoid cartilage	2 cm cranial to sternal manubrium	Anterior edge of sternocleidomastoid m. (cranially)/body of	Posterior edge of sternocleidomastoid m. (cranially)/scalenius mm.	Deep (medial) surface of sternocleidomastoid m. (cranially)/lateral edge of	Medial edge of common carotid artery/lateral edge of thyroid gland/scalenius mm.

	sternocleidomastoid m. (caudally)	(caudally)	sternocleidomastoid m. (caudally)	(cranially)/medial edge of sternocleidomastoid m. (caudally)

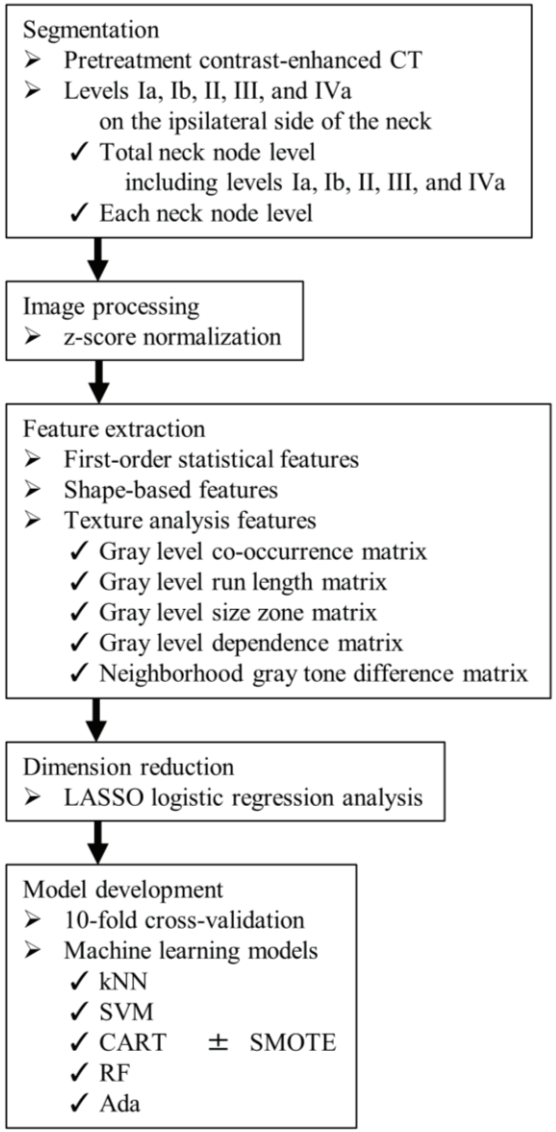
NA, not available.

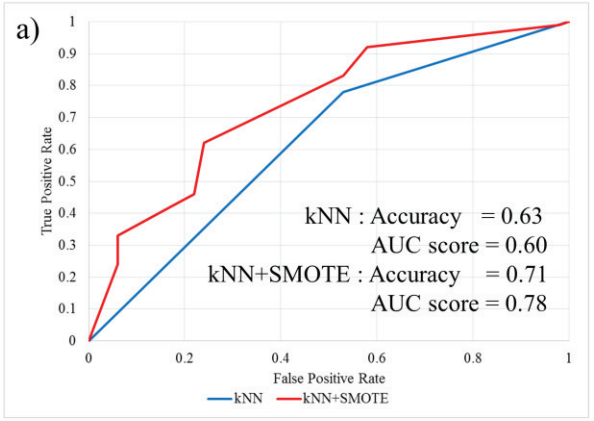

II
Ib
II
Ia
Ib
II
III
IVa

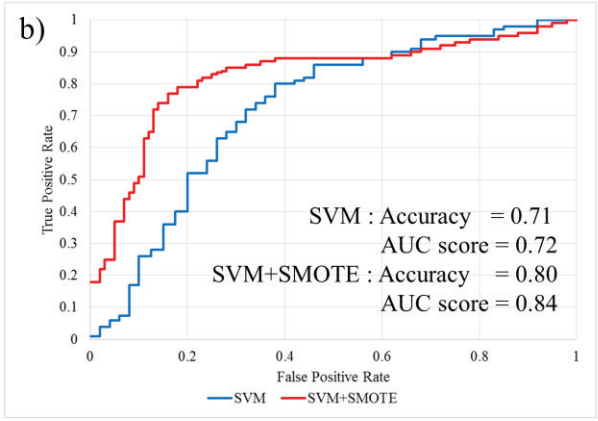


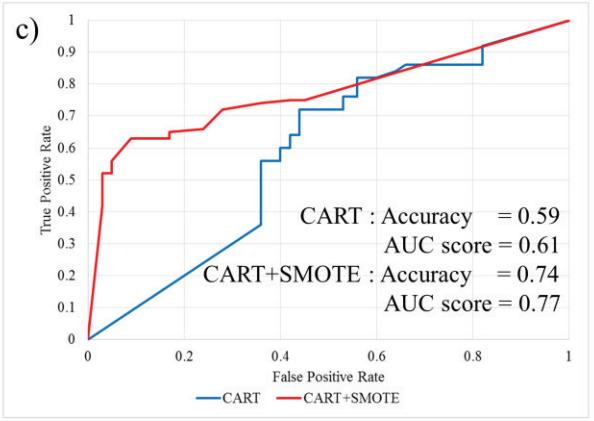
1 label per each case

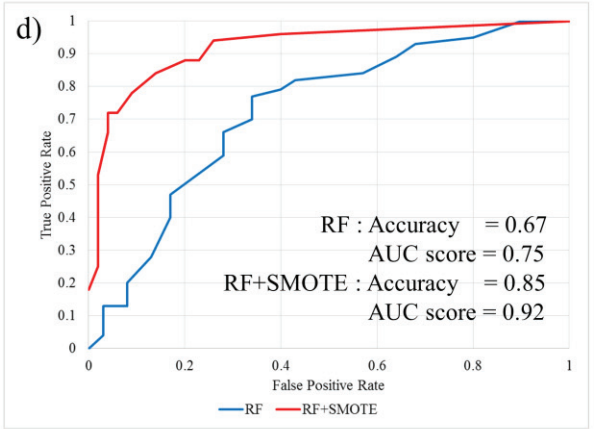
5 labels per each case

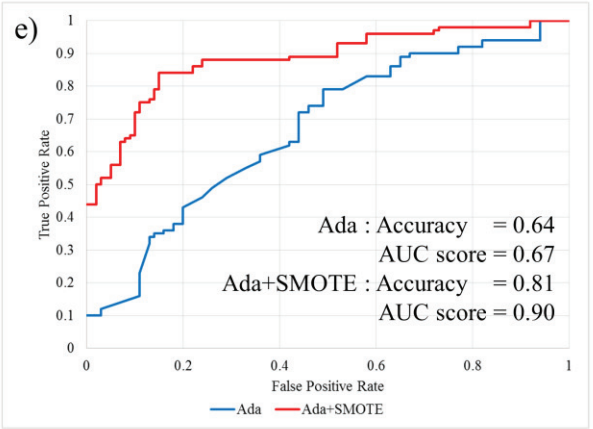

Segmentation
➢ Pretreatment contrast-enhanced CT
➢ Levels Ia, Ib, II, III, and IVa
on the ipsilateral side of the neck
✓ Total neck node level
including levels Ia, Ib, II, III, and IVa
✓ Each neck node level
Image processing
➢ z-score normalization
Feature extraction
➢ First-order statistical features
➢ Shape-based features
➢ Texture analysis features
✓ Gray level co-occurrence matrix
✓ Gray level run length matrix
✓ Gray level size zone matrix
✓ Gray level dependence matrix
✓ Neighborhood gray tone difference matrix
Dimension reduction
➢ LASSO logistic regression analysis
Model development
➢ 10-fold cross-validation
➢ Machine learning models
✓ kNN
✓ SVM
✓ CART ± SMOTE
✓ RF
✓ Ada

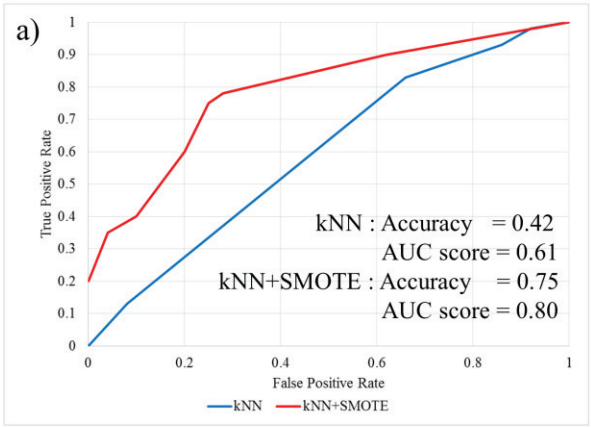

a)
True Positive Rate
False Positive Rate
kNN : Accuracy = 0.63
AUC score = 0.60
kNN+SMOTE : Accuracy = 0.71
AUC score = 0.78
kNN
kNN+SMOTE

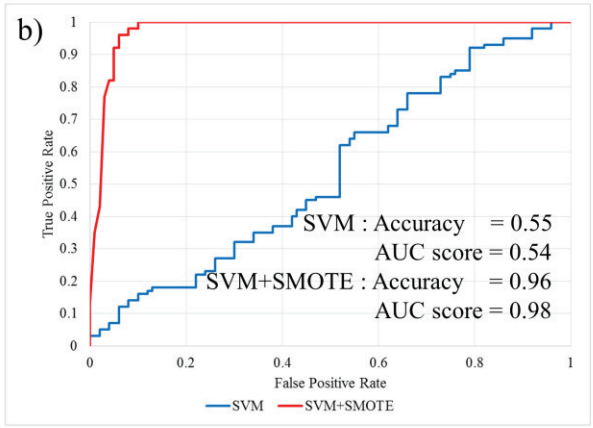

b)
True Positive Rate
False Positive Rate
SVM : Accuracy = 0.71
AUC score = 0.72
SVM+SMOTE : Accuracy = 0.80
AUC score = 0.84
SVM
SVM+SMOTE

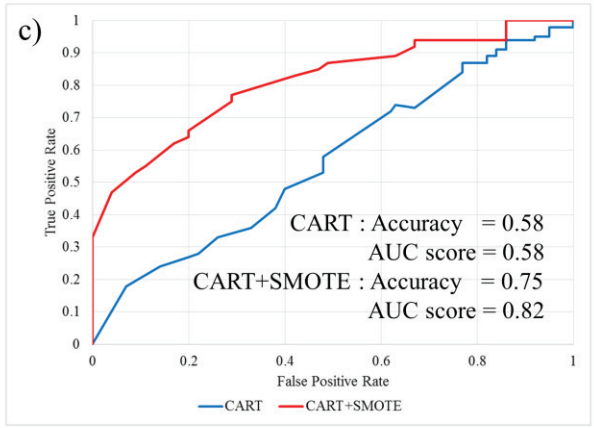

c)
True Positive Rate
False Positive Rate
CART : Accuracy = 0.59
AUC score = 0.61
CART+SMOTE : Accuracy = 0.74
AUC score = 0.77
CART
CART+SMOTE

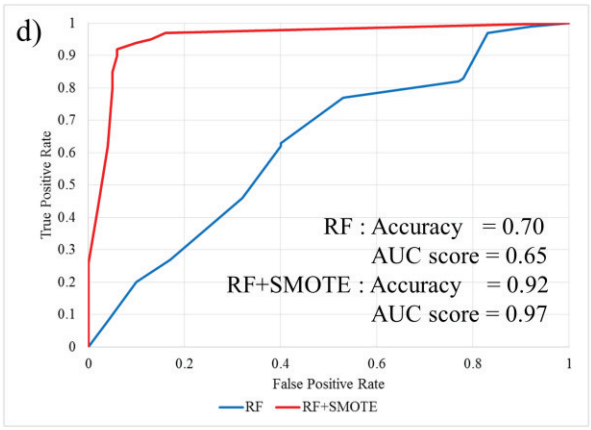

d)
True Positive Rate
False Positive Rate
RF : Accuracy = 0.67
AUC score = 0.75
RF+SMOTE : Accuracy = 0.85
AUC score = 0.92
RF
RF+SMOTE

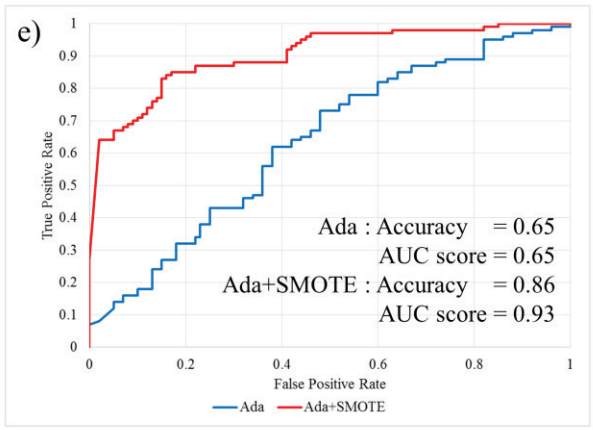

e)
True Positive Rate
False Positive Rate
Ada : Accuracy = 0.64
AUC score = 0.67
Ada+SMOTE : Accuracy = 0.81
AUC score = 0.90
Ada
Ada+SMOTE


a)
True Positive Rate
False Positive Rate
kNN : Accuracy = 0.42
AUC score = 0.61
kNN+SMOTE : Accuracy = 0.75
AUC score = 0.80
kNN
kNN+SMOTE


b)
True Positive Rate
False Positive Rate
SVM : Accuracy = 0.55
AUC score = 0.54
SVM+SMOTE : Accuracy = 0.96
AUC score = 0.98
SVM
SVM+SMOTE


c)
True Positive Rate
False Positive Rate
CART : Accuracy = 0.58
AUC score = 0.58
CART+SMOTE : Accuracy = 0.75
AUC score = 0.82
CART
CART+SMOTE


d)
True Positive Rate
False Positive Rate
RF : Accuracy = 0.70
AUC score = 0.65
RF+SMOTE : Accuracy = 0.92
AUC score = 0.97
RF
RF+SMOTE


e)
True Positive Rate
False Positive Rate
Ada : Accuracy = 0.65
AUC score = 0.65
Ada+SMOTE : Accuracy = 0.86
AUC score = 0.93
Ada
Ada+SMOTE