Abstract. Background/Aim: To investigate survival outcomes and recurrence patterns using machine learning in patients with salivary gland malignant tumor (SGMT) undergoing adjuvant chemoradiotherapy (CRT). Patients and Methods: Consecutive SGMT patients were identified, and a data set included nine predictor variables and a dependent variable [disease-free survival (DFS) event] was standardized. The open-source R software was used. Survival outcomes were estimated by the Kaplan–Meier method. The random forest approach was used to select the important explanatory variables. A classification tree that optimally partitioned SGMT patients with different DFS rates was built. Results: In total, 54 SGMT patients were included in the final analysis. Five-year DFS was 62.1%. The top two important variables identified were pathologic node (pN) and pathologic tumor (pT). Based on these explanatory variables, patients were partitioned in three groups, including pN0, pT1-2 pN+ and pT3-4 pN+ with 26%, 38% and 75% probability of recurrence, respectively. Accordingly, 5-year DFS rates were 73.7%, 57.1% and 34.3%, respectively. Conclusion: The proposed decision tree algorithm is an appropriate tool to partition SGMT patients. It can guide decision-making and future research in the SGMT field.

Salivary gland malignant tumors (SGMTs) accounts for less than 5% of all head and neck cancers (1). Despite improvement in SGMT diagnosis and management in recent years, the risk of loco-regional recurrence and development of distant metastasis in patients treated with curative intent remains relatively high (up to 35%) (2). It is important to understand what drives SGMT recurrence onset. Different predictors have been found to be relevant, including demographic factors (age and gender) health factors (general condition and co-morbidities) and tumor factors (stage, site, margin status, skin/bone invasion, facial nerve dysfunction, lymph-vascular invasion, histology and grade) (2-4). But no definitive factors have been established to predict recurrence outcomes. Recently the application of artificial intelligence has progressed in the medicine field. Machine learning, including decision tree algorithms, is now considered a valid predictive technique, but its use is sparse in SGMTs (5-6). Based on our clinical data, we applied machine learning approaches to analyze survival outcomes and predict recurrence rate in high risk SGMT patients. The hope is to be useful in future clinical trials design.
Patients and Methods

Patient population. Patients at Policlinico Umberto I, Sapienza University of Rome were included in this study following institutional ethical committee approval (ref. 5975). All clinical data were anonymized by the researchers and all potential patient identifiers were removed prior to data analysis. Data for consecutive patients with histologically proven SGMT were reviewed in this retrospective analysis. All patients with metastatic disease at diagnosis, those who had received previous radiation to the head and neck region, those who did not necessitate adjuvant treatment, and those treated with palliative intent were excluded. Diagnosis was based on the clinical presentation, imaging and cytology/histology results. Fine-needle aspiration cytology (FNAC), followed by fine-needle aspiration biopsy (FNAB), when suggestive of the diagnosis, was performed to obtain preliminary histologic information. Magnetic resonance imaging (MRI) was performed with intravenous contrast in the head and neck region. Total-body contrast-enhanced computed tomography (CT) was recommended to exclude distant metastases. In the case of an uncertain diagnosis, patients underwent positron emission tomography (PET)-CT imaging. For the tumor (T)/lymph node (N)/metastasis (M) classification, all cases were re-staged according to the 8th American Joint Committee on Cancer Staging System (AJCC) (7).

Treatment. All patients were referred to the multidisciplinary head and neck tumor board to define treatment strategy. Written informed consent was obtained from all patients before treatment initiation. The detailed treatment process has been previously described (8) and is briefly summarized below. Wide excision with adequate clear surgical margins was the standard. Some form of neck dissection was combined, depending on primary tumor site and tumor histology. Neck dissection as well as adjuvant treatment indications were driven by clinical stage at diagnosis and adverse pathologic features. Post-operative (chemo)radiotherapy [(C)RT] was performed within 6 weeks after surgery. A careful dental-oral evaluation was recommended before adjuvant (C)RT.Intensity modulated radiotherapy (IMRT) technique was used to irradiate the entire surgical tumor bed and the anatomic site of possible disease spread, including lymphatic drainage. Target volume delineation depended on primary tumor site. A total dose of 66 Gy (2 Gy/fraction) up to 70 Gy (2 Gy/fraction; in case of positive margins or macroscopic residual disease or extracapsular nodal spread) was delivered.

Follow-up. As previously reported, a follow-up program, including a complete head and neck exam and diagnostic imaging exams, was routinely planned (9). The clinical examination was performed every 3 months during the first and second year, and every 6 months thereafter. Whereas, after a post-treatment baseline exam (3 months after treatment), MRI imaging and/or CT with contrast was performed every 6 months for the first 2 years and annually thereafter.

Statistical analysis. Statistical analysis was performed using R-Studio 0.98.1091 software (Boston, MA, USA). Standard descriptive statistics were used to evaluate the distribution of each variable. Continuous variables are reported as median and categorical variables as frequencies or percentages. Disease-free survival (DFS) and overall survival (OS) were calculated in months from the date of diagnosis to the first event, including date of the last follow-up or death (OS) and/or relapse (DFS). Survival distributions were estimated by the Kaplan-Meier method and compared with log-rank tests. p-Values <0.05 were considered significant. In addition to these standard statistical methods, a machine learning-based methodology was applied to define significant clinical predictors of recurrence rate. The randomForest package was used to define important explanatory variables. In the model, continuous variables were dichotomized. SGMTs were classified in two major categories: high grade and low grade tumor, based on their higher and lower risk of nodal metastasis, respectively (10). The following variables were investigated: age at diagnosis (<65 years versus ≥65 years), gender (male versus female), type of salivary gland (minor versus major), histologic category (low grade versus high grade), pathologic (p) T (pT1-2 versus pT3-4), pN (pN0 versus pN+), surgical margin (negative versus positive), lymphovascular invasion/perineural invasion (LVI/PNI) (negative versus positive) and type of adjuvant treatment (RT versus CRT). The randomForest algorithm was applied to build a random forest of a fixed number of classification trees based on the investigated variables. The dependent variable referred to recurrence event (no or yes). Then, using the importance() function, we evaluated the importance of each variable. Variables associated with a mean decrease in accuracy >1% were then included to construct the classification tree. The rpart packages were used to identify a corresponding optimal decision tree. The rpart algorithm splits a group into two groups that are as different from each other as possible. It was used to decide which variables to split and which splitting value to take at each step of the tree’s construction. To define the optimal tree size, the tree was pruned using the cross-validation error criterion. The minimum error rule (size producing the minimum cross-validation error) was applied.

Results

Description of patients cohort. Patient and tumor characteristics are listed in Table I. Median age was 60 years (range=29-86 years). All patients underwent radical surgery at Policlinico Umberto I, Sapienza University of Rome between January 2002 and October 2019 and received a pathological SGMT diagnosis. Most tumors were pT3-4 (n=28; 51.6%) and/or pN positive (n=20; 37.0%). Based on high histological grade or adenoid cystic histology, microscopic or macroscopic residual disease, perineural or lymphovascular invasion, T3-4 stage disease and/or positive lymph nodes, adjuvant RT was planned. Approximately 22% of the patients (n=12; 22.2%) received concurrent cisplatin-based chemotherapy.

Recurrence patterns. During the study period, 21 events (recurrence or death) were observed in the entire cohort. Median DFS time and OS time were, respectively, 123 months and 136 months. The most common recurrence pattern was hematogenous distant metastasis (n=10; 47.6%) followed by local relapse (n=6; 28.6%). Details are presented in Table II.

Machine learning-based methodology. The following variables were investigated with randomForest: age at diagnosis, gender, type of salivary gland, histologic category, pT, pN, surgical margin, LVI/PNI and type of adjuvant treatment. The dependent variable referred to recurrence event (no or yes).
Table I. Patient and tumor characteristics.

| Characteristics                  | n (%)          |
|----------------------------------|----------------|
| Age                              |                |
| Median (range)                    | 60 (29-86)     |
| Young (<65 years)                | 33 (61.1)      |
| Older (≥65 years)                | 21 (38.9)      |
| Gender                           |                |
| Male                             | 31 (57.4)      |
| Female                           | 23 (42.6)      |
| Race                             |                |
| White                            | 54 (100)       |
| Performance status               |                |
| 0-1                              | 54 (100)       |
| >1                               | 0 (0)          |
| Localization                     |                |
| Parotid gland                    | 33 (61.1)      |
| Nasal cavity and paranasal sinuses | 1 (1.9)     |
| Submandibular gland              | 5 (9.2)        |
| Oral cavity                      | 8 (14.8)       |
| Nasopharynx                      | 4 (7.4)        |
| Sublingual gland                 | 2 (3.7)        |
| Oropharynx                       | 1 (1.9)        |
| Histology                        |                |
| Adenoid cystic carcinoma         | 11 (20.4)      |
| Adenocarcinoma                   | 7 (13.0)       |
| Mucoepidermoid carcinoma         | 9 (16.7)       |
| Squamous cell carcinoma          | 9 (16.7)       |
| Acinic cell carcinoma            | 4 (7.4)        |
| Myoepithelial carcinoma          | 3 (5.5)        |
| Carcinoma ex pleomorphic adenoma | 3 (5.5)        |
| Other                            | 8 (14.8)       |
| Tumor (T) stage                  |                |
| T1                               | 14 (25.9)      |
| T2                               | 12 (22.2)      |
| T3                               | 12 (22.2)      |
| T4                               | 16 (29.7)      |
| Node (N) stage                   |                |
| N0                               | 34 (63.0)      |
| N1                               | 7 (13.0)       |
| N2                               | 6 (11.0)       |
| N3                               | 7 (13.0)       |
| Concomitant chemotherapy          |                |
| No                               | 42 (77.8)      |
| Yes                              | 12 (22.2)      |
| Radiotherapy technique           |                |
| IMRT                             | 54 (100)       |

Table II. Pattern of recurrence.

| Recurrence                             | n (%) |
|----------------------------------------|-------|
| Local                                  | 5 (31.3) |
| Regional                               | 0 (0)  |
| Distant                                | 9 (56.3) |
| Local and regional                     | 1 (6.2)  |
| Local, regional and distant            | 1 (6.2)  |

All predictor variables, as well as their values and proportions are listed in Table I. We applied randomForest using the ntree (number of simulated decision trees)=500 to analyze these data and the top two important predictors were pT and pN involvement, with a mean decrease accuracy of 1.63%, and 1.38%, respectively (Figure 1). These two variables were used in rpart to grow an optimal classification tree. Because of the categorical nature of the dependent variable, the rpart algorithm was applied with the option method="class", which provides a classification tree. To control the length of the tree before pruning, we used the parameter settings complexity parameter (cp)=10^-9 and minbucket (number of observations in any terminal node)=1. The cross-validation error was used to determine the optimal level of tree complexity and the minimum-error rule was applied. The plot of the final classification tree is shown in Figure 2. The decision tree predicts the risk of recurrence of SGMT patients, based on pT and pN involvement. The split at the top of the tree resulted in two large branches: the left-hand branch included patients with pN negative cancer (63% of the overall sample, with 26% probability of recurrence); the right-hand branch corresponded to pN positive cancer (37% of the overall sample, with 60% probability of recurrence). The right branch was further subdivided by pT stage (pT1-2 versus pT3-4). The final tree included two splitting variables and three terminal nodes, which partitioned SGMT patients in three groups: (i) signified as less

%: Percentage; IMRT: intensity modulated radiotherapy.

Figure 1. Important variables. pT: Pathological tumor; N_pos: lymph nodes involvement; salivary gland: type of salivary gland (minor versus major); CRT: chemoradiotherapy; LVI and PNI: lymph-vascular and perineural invasion; R_pos: positive surgical margin.

Figure 2. Classification tree for salivary gland malignancies.
patients with pN negative SGMT cancer (63% of the overall sample, with 26% probability of recurrence); (ii) patients who had pT1-2 disease with pN involvement (15% of the overall sample, with 38% probability of recurrence); (iii) patients who had pT3-4 disease with pN involvement (22% of the overall, 75% recurrence probability).

Survival outcomes. Median follow-up time was 48 months (range=5-161 months). Five-year DFS for the entire population was 62.1% (95%CI=0.458-0.748). Five-year DFS among patients with pN negative SGMT cancer was 73.7% (95%CI=0.088-0.520). For patients with pT1-2 pN positive disease, the 5-year DFS was 57.1% (95%CI=0.187-0.172). Whereas the 5-year DFS for cases with pT3-4 and pN involvement was 34.3% (95%CI=0.153-0.089). These DFS rates according to classification tree branches are shown in Figure 3 (p=0.018).

Overall, the 5-year OS rate was 77.0% (95%CI=0.612-0.870). In case of pN negative tumor, pT1-2 pN+ and pT3-4 pN+, the 5-year OS rates were 85.7% (95%CI=0.067-0.660), 71.4% (95%CI=0.171-0.258) and 57.8% (95%CI=0.170-0.208), respectively, (p=0.160).

Discussion

In this study we used machine learning techniques to build a model to detect and visualize significant prognostic variables of SGMT recurrence probability. Cervical pN involvement and pathologic tumor size/extension were found to be the most prominent variables to predict SGMT recurrence after surgical resection and adjuvant CRT. The decision tree showed that recurrence probability for the three groups pN0, pT1-2 pN+ and pT3-4 pN+, in which our entire sample had been partitioned, was 26%, 38% and 75%, respectively. This partition demonstrated a significant distinction in 5-year DFS rates between the three groups (p-value=0.018). Patients with any pT pN negative disease had a better DFS (73.7%) compared to pT1-2 pN positive (57.1%) and pT3-4 pN positive (34.3%) cases.

To our knowledge, this is the first study to examine predictors of recurrence in SGMT patients after adjuvant CRT using decision tree analysis. All variables that were included in the model have already been identified by classical retrospective multivariate analysis (2-4, 11). Actually, while there is a huge amount of literature aimed to define OS prognostic factors in SGMT, very few publications searched for prognostic factors for DFS. Both historical and recent series, containing multivariate analysis, most often referred to local control, regional control and distant control as single entities (2-4). Therefore, despite there was some evidence to support a prognostic effect of pN and pT (11), resembling data from our series, it was difficult to adequately compare results. When focusing on DFS, the French Network of Rare Head and Neck Tumors (REFCOR) recently published the largest European prospective study on salivary glands (12). It included exclusively mucoepidermoid carcinoma and a total of 292 cases were finally analyzed. With a median follow-up of 26 months, the 5-year DFS was 69% (12). In the multivariate analysis, diabetes, advanced clinical stage and high histological grade were found to have a significant negative impact on DFS rate (12). As stated by the Authors, these results should be interpreted with caution mainly given the difference in the treatment plan according to histological grading and clinical stage. To minimize this difference, our population only included SGMT patients with at least one adverse pathologic feature at the time of surgery. In current clinical practice these high-risk SGMT patients receive adjuvant treatment (8, 13). Our results suggested that a lower
rate of co-morbidities and of histological grade did not improve DFS outcome. The benefit in DFS was limited to pN and pT variables. It is important to note that our pN negative group included pT3pN0 [stage III prognostic group according to the 8th TNM staging system (7)] and pT4pN0 [stage IV prognostic group according to the 8th TNM staging system (7)] patients. This finding is new and to further study the clinical relevance of such a conclusion an external validation is welcomed to determine its reproducibility and validity. A possible reasons for this is that lymph nodes involvement – more than advanced T stage – contributed to worse DFS outcome, considering that most of recurrent events are due to distant metastasis. Accordingly to this assumption, some papers in the literature reflected that metastasis-free survival rate decreased strongly with an increasing N stage, indicating N stage as an independent factor for DFS (2, 14). This finding suggests that, in pN+ SGMT, personalized treatment intensification in the adjuvant phase could have the greatest efficacy in reducing the occurrence of metastatic relapse. The evaluation of epidermal growth factor receptor (EGFR), panRAS, BRAF mutational status and programmed death-ligand 1 (PD-L1) expression represents an attractive research field in the context of SGMT and several studies are exploring these biomarker-driven strategy to individualize the decision for adjuvant systemic therapy (15).

Finally, the heterogeneity of histology subtypes must be discussed. Because histology was not associated with a DFS difference, one could jeopardize the efficacy of the system. Certainly the low number of patients is responsible for the heterogeneity observed and can reduce the power of the analysis, but we believe that the exclusive inclusion of patients with high-risk indications to adjuvant treatment do not bias it.

The strengths of this study included the use of a homogenous sample and the novel use of the decision tree analysis to examine predictors of cancer recurrence. Compared to more traditional methods such as logistic regression, decision tree analysis is able to better handle non-linear relationships between outcome and variables. No literature exists regarding the application of machine learning algorithms in SGMT management and identifying risk groups should have important implications for potential treatment strategies and follow-up definition. However satisfying, the applicability of our partitioned groups needs independent validation. Other limitations of the study included its retrospective design, which introduces the risk of potential selection bias, and the relative small number of patients. For
sure, random forest algorithm performs best when dealing with big data, and the accuracy of the decision tree algorithm is better. Waiting for the implementation of SGMT multigene profiles, it is expected that the decision tree can be translated into decision support tool in SGMT management. It can help to provide necessary information and knowledge required by both clinicians and patients for accurate prediction of SGMT recurrence and better decision-making.

To conclude, we introduced for the first time the decision tree approach to analyze SGMT data. The proposed classification tree confirmed the importance of pN and pT as recurrence predictors in this patient population.

Conflicts of Interest

The Authors have no conflicts of interest to declare in relation to this study.

Authors’ Contributions

FDF, DM, VT designed and supervised the study, AAT, LIR, VC, SM, DM, GT, MDM, FV collected data. FDF, AAT and LIR did the statistical analyses and wrote the draft, with revisions from the other Authors. All Authors approved the final version.

References

1 Misfud MJ, Burton JN, Trotti AM and Padhya TA: Multidisciplinary management of salivary gland cancers. Cancer Control 23(3): 242-248, 2016. PMID: 27556664. DOI: 10.1177/10732748166300307
2 Terhaard CH, Lubsen H, Van der Tweel I, Hilgers FJ, Eijkenboom WM, Marres HA, Tjho-Heslinga RE, de Jong JM, Roodenburg JL and Dutch Head and Neck Oncology Cooperative Group: Salivary gland carcinoma: independent prognostic factors for locoregional control, distant metastases, and overall survival: results of the Dutch head and neck oncology cooperative group. Head Neck 26(8): 681-92; discussion 692-3, 2004. PMID: 15287035. DOI: 10.1002/hed.10400
3 Holtzman A, Morris CG, Amdurd RJ, Dziegielewski PT, Boyce B and Mendenhall WM: Outcomes after primary or adjuvant radiotherapy for salivary gland carcinoma. Acta Oncol 56(3): 484-489, 2017. PMID: 27846763. DOI: 10.1080/028418 6X.2016.1253863
4 Li Y, Ju J, Liu X, Gao T, Wang Z, Ni Q, Ma C, Zhao Z, Ren Y and Sun M: Nomograms for predicting long-term overall survival and cancer-specific survival in patients with major salivary gland cancer: a population-based study. Oncotarget 8(15): 24469-24482, 2017. PMID: 28160551. DOI: 10.18632/ oncotarget.14905
5 Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV and Fotiadis DI: Machine learning applications in cancer treatment and prediction. Comput Struct Biotechnol J 13: 8-17, 2014. PMID: 25750696. DOI: 10.1016/j.csbj.2014.11.005
6 Deist TM, Dankers FJWM, Valdes G, Wijsman R, Hsu IC, Oberije C, Lusberg T, van Soest J, Hoeben F, Jochems A, El Naqa I, Wee L, Morin O, Raleigh DR, Bots W, Kaanders JH, Belderbos J, Kwint M, Solberg T, Monshouwer R, Bussink J, Dekker A and Lambin P: Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers. Med Phys 45(7): 3449-3459, 2018. PMID: 29763967. DOI: 10.1002/mp.12967
7 Lydiatt WM, Patel SG, O’Sullivan B, Brandwein MS, Ridge JA, Migliaccia JC, Loomis AM and Shah JP: Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. CA Cancer J Clin 67(2): 122-137, 2017. PMID: 28128848. DOI: 10.3322/caac.21389
8 De Felice F, de Vincentis M, Valentina V, Musio D, Mezi S, Lo Mele L, Della Monaca M, D’Aguanno V, Terenzi V, Di Brino M, Brauner E, Bulzonetti N, Tenore G, Pomati G, Cassoni A, Tombolini M, Battisti A, Greco A, Pompa G, Minni A, Romeo U, Cortesi E, Polimeni A and Tombolini V: Management of salivary gland malignant tumor: the Policlinico Umberto I, “Sapienza” University of Rome Head and Neck Unit clinical recommendations. Crit Rev Oncol Hematol 120: 93-97, 2017. PMID: 29198342. DOI: 10.1016/j.critrevonc.2017.10.010
9 De Felice F, de Vincentis M, Valentina V, Musio D, Mezi S, Lo Mele L, Terenzi V, D’Aguanno V, Cassoni A, Di Brino M, Tenore G, Bulzonetti N, Battisti A, Greco A, Pompa G, Minni A, Romeo U, Cortesi E, Polimeni A and Tombolini V: Follow-up program in head and neck cancer. Crit Rev Oncol Hematol 113: 151-155, 2017. PMID: 28427504. DOI: 10.1016/j.critrevonc.2017.03.012
10 PDQ Adult Treatment Editorial Board: Salivary gland cancer treatment (Adult) (PDQ®): Health professional version. 2019. PMID: 26389389.
11 Pooten VV, Hart A, Vauterin T, Jeunen G, Schoenaers J, Hamoir M, Balm A, Stennert E, Gunitinas-Lichius O and Delaere P: Prognostic index for patients with parotid carcinoma: international external validation in a Belgian-German database. Cancer 115(3): 540-550, 2009. PMID: 19137571. DOI: 10.1002/cncr.24015
12 Dahan LS, Giorgi R, Vergez S, Le Taillandier de Gabory L, Costes-Martineau V, Herman P, Poissonnet G, Mauvais O, Malard O, Garrel R, Uro-Coste E, Barry B, Bach C, Chevalier D, Mouawad F, Merol JC, Bastit V, Thariat J, Glinain L, Dufour X, Righini CA, Moya-Plana A, Even C, Radulesco T, Michel J, Bautjat B, Fakhry N and REFCOR members: Mucoepidermoid carcinoma of salivary glands: A French Network of Rare Head and Neck Tumors (REFCOR) prospective study of 292 cases. Eur J Surg Oncol 47(6): 1376-1383, 2021. PMID: 33248902. DOI: 10.1016/j.ejso.2020.11.123
13 National Comprehensive Cancer Network Guidelines in Oncology Head and Neck Cancers Version 1.2021. Available at: http://www.nccn.org/ [Last accessed on August 4, 2021]
14 Hocwald E, Korkmaz H, Yoo GH, Adsay V, Shibuya TY, Abrams J and Jacobs JR: Prognostic factors in major salivary gland cancer. Laryngoscope 118(8): 1434-1439, 2001. PMID: 11568581. DOI: 10.1097/00005537-200108000-00021
15 Mezi S, Pomati G, Botticelli A, De Felice F, Musio D, Della Monaca M, Amirhassankhani S, Vullo F, Cerbelli B, Carletti R, Di Gioia C, Catalano C, Valenitin L, Tombolini V, Della Rocca C and Marchetti P: Primary squamous cell carcinoma of major salivary gland: “Sapienza Head and Neck Unit” clinical recommendations. Rare Tumors 12: 2036361320973526, 2020. PMID: 33282162. DOI: 10.1177/2036361320973526

Received July 22, 2021

Revised August 7, 2021

Accepted August 9, 2021