**Machine Learning in Clinical Diagnosis of Head and Neck Cancer**

**Abstract**

Objective

Machine learning has been effective in other areas of medicine, this study aims to investigate this with regards to HNC and identify which algorithm works best to classify malignant patients.

Design

An observational cohort study.

Setting

Queen Elizabeth University Hospital.

Participants

Patients who were referred via the USOC pathway between January 2019 and May 2021.

Main outcome measures

Predicting the diagnosis of patients from three categories, benign, potential malignant and malignant, using demographics and symptoms data.

Results

The logistic regression-based models with a penalty term worked best on the data, ridge achieving an AUC of 0.7081. The demographic features describing living alone and recreational drug use history were the most important variables alongside the red flag symptom of a neck lump.

Conclusion

Further studies should aim to collect larger samples of malignant and pre-malignant patients to improve the class imbalance and increase the performance of the machine learning models.

**Key Points**

* This observational cohort study's aim is to identify the machine learning model which best predicts head and neck cancer, through factors such as demographics, red flag symptoms or associated symptoms.
* After up-sampling was conducted on the imbalanced dataset, the models evaluated were ordinal regression, lasso, elastic net, ridge, random forest, classification trees and linear discriminant analysis.
* Ranking was based on the multiclass area under the receiver operating characteristic (ROC) curve (AUC) gave the following results: ridge (0.7081), elastic net (0.7044), lasso (0.7044), classification trees (0.6172), linear discriminant analysis (0.5877), ordinal logistic regression (0.5716), and random forest (0.5001).
* The three variables deemed to be most important were found to be the patient’s living situation, drug use and having the symptom of a neck lump.
* Further studies should aim to collect more data on malignant and pre-malignant cases or use different forms of up-sampling to remove the class imbalance of the data presented in this study.

**Keywords**

head and neck cancer, machine learning, up-sampling, area under the receiver operator curve

**1 Introduction**

Currently the number of patients referred to Urgent Suspicion of Cancer (USOC) diagnostic clinics are rising. Less than 10% of people referred to these clinics are diagnosed with cancer (1). Within the Head and Neck clinic, malignant diagnosis pick-up rates are even lower where the cancer pick-up rate is between 3-8% (2, 3). This high volume of patients attending for diagnoses has created a significant burden on the USOC head and neck referral pathway, making it challenging to meet the 31-day diagnostic target created by the Scottish government.

The head and neck risk calculator has created a classification system which can identify the probability of a patient having cancer based on their demographics and symptoms. The study obtained results with an AUC of 88.6% (4). This study aims to review machine learning models and identify whether these algorithms can better predict head and neck diagnosis of cancer, to support USOC clinics.

**2 Methodology**

**2.1 Data**

There were 1045 patients eligible for inclusion in this observational cohort study. The reporting of this study adhered to the EQUATOR reporting guidelines for cohort studies. These patients were referred via the USOC pathway between January 2019 and May 2021. All patients included in the study agreed to anonymised data collection and analysis. The patients were categorised into three groups: patients with a benign diagnosis, diagnosis with malignant potential, and malignant diagnosis. There were 885 patients diagnosed with a benign condition, 61 patients with a diagnosis of malignant potential, and 99 malignant diagnoses.

For the purposes of subgroup analysis, benign, potential malignant and malignant diagnoses were subclassified and are shown in Table 1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 1**. **Classification and frequency of diagnoses into benign, malignant potential and malignant outcome classification.** | | | | | |
| **Benign Classification** | **Frequency** | **Malignant Potential Classification** | **Frequency** | **Malignant Classification** | **Frequency** |
| **No abnormality** | 301 | **Malignant potential salivary** (pleomorphic adenoma) | 29 | **Malignant oropharyngeal** | 26 |
| **Benign neck** (sebaceous cyst, lipoma, thyroglossal duct cyst, branchial cyst, goitre, U2 nodules, reactive lymph nodes | 298 | **Malignant potential thyroid**  (Thy3 follicular lesions) | 18 | **Lymphoma** | 21 |
| **Benign pharynx** (globus, benign oropharynx, reflux, benign oesophageal stricture) | 146 | **Malignant potential laryngeal**  (laryngeal dysplasia) | 11 | **Malignant laryngeal** | 23 |
| **Benign salivary** (sialadenitis) | 57 | **Malignant potential oral**  (leukoplakia) | 3 | **Malignant thyroid** – (Thy4 or 5) | 13 |
| **Benign laryngeal** (Reinke’s oedema, presbyphonia) | 54 |  |  | **Malignant hypopharyngeal** | 6 |
| **Musculoskeletal pathology** (C-spine pathology, sternoclavicular joint arthritis) | 13 |  |  | **Metastatic SCC unknown primary** | 5 |
| **Benign oral** | 8 |  |  | **Synchronous H&N primaries** | 3 |
| **Granulomatous neck infection** | 7 |  |  | **Malignant salivary** | 1 |
| **Benign lateral skull base** | 1 |  |  | **Malignant oral** | 1 |
| Total | 885 |  | 61 |  | 99 |

**2.2 Variables**

The variables routinely recorded at clinic were a range of demographic questions, red flag symptoms and a questionnaire regarding other associated symptoms. Demographics included age, gender, employment status, living situation, smoking status, alcohol consumption and drug use. Red flag symptoms were persistent hoarseness, neck lump, persistent throat pain, an oral ulcer/lump, odynophagia or referred otalgia. Also included were associated symptoms reported throughout the questionnaire. These were cough, reflux, unexpected weight loss, dysphagia to solids and globus sensation.

**2.3 Machine Learning**

Due to the unbalanced nature of the data, the first machine learning approach taken was to up-sample the data. Up-sampling, also known as oversampling, is a method to modify the distribution of the data without having to decrease the size of the dataset and lose any important information. The method randomly duplicates rows of data from the class with low observations until the number of observations for this minority class is in line with the majority (5). Before this could be completed, the data was split into a training and testing dataset, 70% training to 30% testing. This increased the training data size from 731 participants to 1884, increasing the groups of malignant potential and malignant in line with the benign patients to 628 patients per class.

The second machine learning approach taken was the modelling of the up-sampled data. Seven models were created using four logistic regression-based models (ordinal logistic regression, lasso, ridge and elastic net), two tree-based models (random forest and classification trees) and lastly, a discriminant analysis model (linear discriminant analysis). Cross-validation was also used within lasso, ridge and elastic net to obtain the optimal value of theta, a parameter within the model which represents the weighting given to the penalty term.

All analysis was conducted within R. To up-sample the data the caret package was utilised. For the ML models the MASS, randomForest, rpart, and glmnet packages were used.

**2.4 Model Comparison**

To compare the models' predictive power, the multiclass area under the receiver operating characteristic (ROC) curve (AUC) was used. The pROC package was used to obtain the multi-class AUC scores for each of the models. Multiclass AUC is the mean of the one-to-one AUC scores (6). A higher AUC score is considered better, with 1 representing perfect classification. If the score is 0.5 this means that the model predicts no better than a guess (7). The model with the highest multiclass AUC was chosen to be the best model for the data.

Each of the machine learning classification models chooses the most important variables, that have the most impact on the model. The logistic regression-based and discriminant analysis models have model coefficients which explain the most impactful variables. The tree-based models have a Gini impurity index which tells you the most important variables depending on their Gini score.

**3 Results**

**3.1 Variable Descriptive Statistics**

The demographics, red flag symptoms and associated symptoms in addition to descriptive statistics are outlined in Table 2, 3 and 4. From the demographics, a higher mean age is seen with malignancy and potential malignancy compared with benign. There are also more male patients who have malignancies, more retired patients, and a higher rate of patients with consumption of more than 14 units of alcohol per week. Potential malignancy had more smokers than the other categories. Throughout all participants, the rate of drug use is low, meaning the results that come from this variable should be treated cautiously.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 2. Demographics** | | | | |
| **Characteristic** 1 Mean (SD); n (%) | **Overall, N = 1,0451** | **1 Benign, N = 8851** | **2 Malignant Potential, N = 611** | **3 Malignant, N = 991** |
| Age (years) | 52 (17) | 51 (17) | 56 (18) | 61 (15) |
| *Male* | 443 (42%) | 351 (40%) | 29 (48%) | 63 (64%) |
| **Employment Status** |  |  |  |  |
| *Employed* | 513 (50%) | 451 (52%) | 24 (41%) | 38 (40%) |
| *Full Time education* | 42 (4.1%) | 36 (4.1%) | 4 (6.8%) | 2 (2.1%) |
| *No/Retired* | 475 (46%) | 388 (44%) | 31 (53%) | 56 (58%) |
| **Living Situation** |  |  |  |  |
| *Married/living with partner/parents/children* | 746 (75%) | 632 (75%) | 44 (83%) | 70 (74%) |
| *Living with friends* | 18 (1.8%) | 16 (1.9%) | 0 (0%) | 2 (2.1%) |
| *Living alone* | 225 (23%) | 194 (23%) | 9 (17%) | 22 (23%) |
| *Residential care* | 2 (0.2%) | 2 (0.2%) | 0 (0%) | 0 (0%) |
| **Smoking status** |  |  |  |  |
| *Never* | 468 (45%) | 410 (47%) | 23 (38%) | 35 (36%) |
| *Yes* | 255 (25%) | 199 (23%) | 26 (43%) | 30 (31%) |
| *Ex* | 308 (30%) | 265 (30%) | 12 (20%) | 31 (32%) |
| **Alcohol Use** |  |  |  |  |
| *Never* | 341 (35%) | 295 (36%) | 16 (28%) | 30 (33%) |
| *<14 units per week* | 510 (52%) | 438 (53%) | 31 (54%) | 41 (45%) |
| *>14 units per week* | 123 (13%) | 93 (11%) | 10 (18%) | 20 (22%) |
| **Drug Use** |  |  |  |  |
| *Never* | 939 (93%) | 789 (92%) | 58 (97%) | 92 (98%) |
| *Yes* | 31 (3.1%) | 29 (3.4%) | 2 (3.3%) | 0 (0%) |
| *Previously* | 44 (4.3%) | 42 (4.9%) | 0 (0%) | 2 (2.1%) |

Table 3 shows that malignant patients have experienced all six of the red flag symptoms more than benign patients. For potential malignant, these patients also experienced more occurrences of hoarseness and a neck lump as symptoms. However, fewer of them experienced throat pain, pain when swallowing, odynophagia and an oral ulcer/lump.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3. Red Flag Symptoms** | | | | |
| **Characteristic** | **Overall, N = 1,0451** | **1 Benign, N = 8851** | **2 Malignant Potential, N = 611** | **3 Malignant, N = 991** |
| *Hoarseness* | 297 (31%) | 243 (30%) | 18 (33%) | 36 (42%) |
| *Neck Lump* | 634 (64%) | 525 (63%) | 41 (72%) | 68 (72%) |
| Throat Pain | 335 (35%) | 289 (36%) | 12 (23%) | 34 (39%) |
| Oral Ulcer/Lump | 193 (20%) | 163 (20%) | 6 (11%) | 24 (27%) |
| It is painful for me to swallow food (odynophagia) | 191 (19%) | 147 (17%) | 8 (14%) | 36 (38%) |
| The pain travels to my ear (referred otalgia) | 289 (29%) | 240 (28%) | 12 (21%) | 37 (39%) |
| 1 n (%) | | | | |

For the associated symptoms (see table 4), cough, unexpected weight loss, dysphagia to solids and globus were experienced more in patients with malignancy. Less reflux symptoms were recorded in this group compared to benign diagnoses. Additionally, all these associated symptoms were experienced less or around the same overall as benign patients.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 4. Associated Symptoms** | | | | |
| **Characteristic** | **Overall, N = 1,0451** | **1 Benign, N = 8851** | **2 Malignant Potential, N = 611** | **3 Malignant, N = 991** |
| I cough a lot | 213 (21%) | 178 (21%) | 10 (17%) | 25 (27%) |
| I have heartburn or reflux | 418 (41%) | 366 (43%) | 22 (39%) | 30 (32%) |
| I have lost weight unexpectedly | 141 (15%) | 113 (14%) | 9 (15%) | 19 (20%) |
| I find it difficult to swallow solid foods like meats (dysphagia) | 192 (19%) | 154 (18%) | 10 (17%) | 28 (30%) |
| Feeling of something in throat | 291 (33%) | 243 (32%) | 18 (35%) | 30 (38%) |
| 1 n (%) | | | | |

**3.3 Machine Learning**

The performances were ranked by multiclass AUC: ridge (0.7081), elastic net (0.7044), lasso (0.7044), classification trees (0.6172), linear discriminant analysis (0.5877), ordinal logistic regression (0.5716), and random forest (0.5001). Ridge performed best, where the multiclass AUC is greater than the other models.

The variable selection gave the most important variables for each model found in Table 5. The variables selected for lasso, ridge and elastic net were similar, selecting mainly the same variables for all three of the models. Although, ridge selected the variable ‘*I have lost weight unexpectedly’* instead of ‘*smoking’* which were selected by both lasso and elastic net. Both tree-based algorithms choose the continuous variable (age) as the most important over any of the categorical variables; which contributes more data to tree based models than categorical variables. They also both chose ‘*employment status’* and ‘*smoking’* as their top two variables. Ordinal logistic regression and linear discriminant analysis have chosen the same top 5 variables, these are also very similar to those found for lasso, ridge and elastic net. The variable *‘living situation’* is shown to be the top variable in 5 of the models and 4th in one, ‘*drug use’* and *‘neck lump’* were also found to be associated in 5 of the models.

|  |  |  |
| --- | --- | --- |
| **Table 5. Top 5 variables deemed to be important for each model** | | |
| **Model** | **Important Variables** | **Value** |
| **Ordinal Logistic Regression – Variable Coefficients** | Living situation  Drug use  Neck Lump  Odynophagia  Gender | -17.4588  -17.2193  1.6812  1.1808  -1.0157 |
| **Random Forest – Gini Impurity (Feature Importance)** | Age  Smoking  Alcohol  Living situation  Employed | 294.606  92.719  84.634  64.005  59.072 |
| **Classification Trees – Feature Importance** | Age  Smoking  Gender  Odynophagia  Employed | 46  9  7  7  5 |
| **Lasso – Variable Coefficients** | Living situation  Drugs  Neck Lump  Dysphagia  Smoking | 6.0149  5.6528  -2.8768  -2.0847  -0.9516 |
| **Ridge – Variable Coefficients** | Living situation  Drug use  Neck Lump  Dysphagia  Unexpected weight loss | 2.7555  2.5813  -2.111  -1.3428  0.8294 |
| **Elastic Net – Variable Coefficients** | Living situation  Drugs  Neck lump  Dysphagia  Smoking | 4.8827  4.5607  -2.6351  -1.8016  -0.8587 |
| **Linear discriminant analysis – Variable Coefficient** | Living situation  Drugs  Neck lump  Odynophagia  Gender | -3.1102  -2.1132  1.2968  1.0956  -0.9262 |

**4 Discussion**

A number of variables were consistently selected for by multiple machine learning models as being predictive of risk for an underlying malignant diagnosis. This included a patient’s socioeconomic status encompassing *‘living situation’*, *‘employment status’*, *age,* *‘drug use’,* and the red-flag symptoms of ‘*neck lump’, ‘odynophagia*’, and ‘*dysphagia’*. We will discuss the limitations of our results and how we mitigated for the unbalanced dataset which was assembled.

**4.1 Living Situation and Employment Status**

A patient’s living situation is a result of a complex set of social, economic, and environmental factors and can have a significant influence on a patient’s health outcomes.

Our models highlighted *‘living situation’* as a key variable, and identified being in *‘residential care’* as being associated with benign diagnostic classification. Given that residential care home patients commonly have multiple co-morbidities, many will have had previous cancers, and that older age is an independent risk factor for developing HNC (8), this is perhaps quite surprising. This category only contained a very small percentage of patients, with all diagnoses falling within the benign diagnosis classification. It is possible that the low numbers sharing the same outcome have strongly influenced the machine learning models.

Although not borne out in our data, it is possible that other living situations can have significant associations with risk for developing HNC. For example, *‘living alone’* may reflect a degree of social isolation and loneliness. Previous studies examining socioeconomic risk factors for HNC identify being *male*, *single* and *‘never married’* as risk factors for developing HNC (9). There is some research that indicates that the stress of social isolation may potentiate head and neck carcinogenesis and tumour growth (10).

Our results have also identified ‘*employment status’* as a key variable in predicting the risk of underlying malignant diagnosis. Being in work is correlated with an increased likelihood of benign diagnosis. The role of age may confound the association between HNC diagnosis and *‘unemployed/retired’* employment status, as age is an independent risk factor for HNC (16). However, the association between unemployment and HNC is well documented, and has previously been investigated in the West of Scotland (11-13).

**4.2 Drug Use**

The results from this study show that drug use is more prevalent in the benign group and has the opposite effect on what previous research has shown. Woodley et al. suggested that drug use be evaluated as a red flag for those with suspected laryngeal cancer due to its association with increased burden of disease (14). Similarly, Douglas et al. found that drug use is a risk factor for patients with laryngeal cancer, with the study finding a large odds ratio of association with disease (15). It should also be noted that these types of variables are often not reported by patients which could also be affecting the results.

Age could well be confounding the seemingly protective factor of drug use. Drug use is likely to be more prevalent in younger age populations. However, between 2016 and 2018 the peak age rate of HNC was between 70 and 74 years (16). Since drug use was very low among the participants, it seems like the significance of this variable could be misinterpreted.

**4.3 Characterising the Significance of Neck Lump, Odynophagia and Dysphagia**

Another variable deemed important in the study was the presence of a neck lump. This is considered one of the main red flag symptoms of HNC (4). That said, the second most common diagnosis in our study population was that of benign neck lumps, including reactive lymph nodes, sebaceous cysts and benign thyroid nodules. Malignant neck lumps behave and present differently to their benign counterparts. Malignant neck lumps will tend to progressively enlarge, while benign neck lumps may both increase and decrease in size, or may be subtle enough to be impalpable on occasion.

Similarly, odynophagia, or *‘painful swallow’* is a well-recognised red flag symptom for HNC (4). However, clinical suspicion for underlying neoplastic process can be guided by certain characteristics of the painful swallow. Odynophagia that is constant, lateralising, and travels to the ear is more concerning for a malignant process than odynophagia that is intermittent and felt in the midline. There are many alternative causes for persistent throat discomfort, including throat clearing, chronic cough, laryngopharyngeal reflux, and inhaler use.

Dysphagia, or *‘difficulty swallowing’*, is another red-flag symptom for HNC (4). Like other alarming symptoms for HNC, features of the dysphagia may make it more or less likely to be associated with an underlying cancer diagnosis. For example, difficulty swallowing liquids that results in coughing or choking is concerning, as it suggests a significantly impaired swallow, perhaps an associated vocal cord palsy. Alternatively, patients who report difficulty swallowing dry foods like tablets can indicate underlying chronic dehydration or a salivary pathology, rather than cancer.

**4.4 Limitations and future work**

Unfortunately, the data was unable to demonstrate the utility of additional questions that attempt to characterise and stratify the significance of the presenting red flag symptom. It is more likely to be related to the dataset than the fundamental associations and is an area that should be develop going forward.

The largest issue faced with modelling this data was the unbalanced nature of the data. While we collected data on nearly 1000 patients, the vast majority of diagnoses were classified as benign. This is in keeping with the 3-8% malignant diagnosis pick up rate at HNC diagnostic clinics (3). The problem with an unbalanced dataset of this nature is that the machine learning tool will have a pick-up rate of 92-97% in simply suggesting a benign diagnosis for every data test. While statistically effective, this is not useful in a clinical context. As such, comparing a comparatively modest cohort of malignant diagnoses limits the ability to develop machine learning diagnostic tool with both high sensitivity and specificity.

The up-sampling technique used helped to counteract this issue as machine learning models work better on balanced datasets. Although not included in the results section, before the up-sampling was conducted the multiclass AUC for ridge (the best-performing model) was significantly lower, specifically 0.4981, compared to the value when up-sampling was used: 0.7081. Ridge was very closely followed by lasso and elastic net both giving a multiclass AUC of 0.7044. These are all regularisation methods which help avoid model overfitting. A key sign of overfitting is when the training data is predicted with low error rates but the testing data is predicted with high error rates (17). This data was continually overfitting with logistic regression which may explain why the linear regularisation methods worked best on the data.

In the future, more data could be collected, and the models could be run again with more cases of malignant and potential malignant diagnoses. Also, an alternative form of up-sampling could also be explored like Synthetic Majority Oversampling Technique (SMOTE). Alternatively, variables which have been found to be significant in other studies could be utilised. For example, in the study conducted by Tikka et al. (4) nose breathing and persistent lesions on head and neck skin were all identified as significantly associated predictors. The inclusion of these predictors may increase the multi-class AUC of the ML models. Lastly, deep learning techniques like Neural Networks (NN) could be used and may be more effective than the machine learning algorithms. Although a previous study reviewing ML models for prediction of outcomes in palatal surgery found that the ML models worked better than NN (18), this type of model may be more effective at diagnosing head and neck cancer since the data presented in this study contains more predictor variables and more patients.

**Ethics Statement**

No ethical approval was sought. Study aims were designed to be in the public interest, with satisfactory methodological quality. High standards of confidentiality and data security were maintained throughout the study. Accordingly, the study was deemed to be low risk by the UK Statistics Authority, Ethics Self-Assessment Tool (19).

**References**

1. Your urgent suspected cancer referral: Cancer Research UK; 2022 [Available from: <https://www.cancerresearchuk.org/cancer-symptoms/what-is-an-urgent-referral>.

2. Lim AE, Douglas CM, Montgomery J. An open structure questionnaire on reasons for delay in presentation: a study of patients attending clinic with red flag symptoms of head and neck cancer. Eur Arch Otorhinolaryngol. 2020;277(6):1801-6.

3. Douglas C, Carswell V, Montgomery J. Outcomes of urgent suspicion of head and neck cancer referrals in Glasgow. Annals of The Royal College of Surgeons of England. 2018.

4. Tikka T, Kavanagh K, Lowit A, Jiafeng P, Burns H, Nixon IJ, et al. Head and neck cancer risk calculator (HaNC-RC) – v.2. adjustments and addition of symptoms and social history factors. 2020.

5. Thanker T. What is upsampling and downsampling with examples? : Kaggle; 2021 [Available from: <https://www.kaggle.com/general/262007>.

6. Multi-class AUC R Documentation [Available from: <https://search.r-project.org/CRAN/refmans/pROC/html/multiclass.html>.

7. Zach. What is Considered a Good AUC score? Statology2021 [Available from: <https://www.statology.org/what-is-a-good-auc-score/#:~:text=The%20answer%3A%20There%20is%20no%20specific%20threshold%20for,better%20than%20a%20model%20that%20performs%20random%20guessing>.

8. Hashibe M, Brennan P, Chuang S. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. Cancer Epidemiol Biomarkers Prev. 2009;18(2):541-50.

9. Johnson S, McDonald J, Corsten M. Socioeconomic factors in head and neck cancer. . J Otolaryngol - Head Neck Surg. 2008;37(4):597-601.

10. Verza F, Valente V, Oliveira L. Social isolation stress facilitates chemically induced oral carcinogenesis. PLoS One. 2021;16(1).

11. Conway D, Hovanec J, Ahrens W. Occupational socioeconomic risk associations for head and neck cancer in Europe and South America: individual participant data analysis of pooled case–control studies within the INHANCE Consortium. J Epidemiol Community Health. 2021;75(8):779.

12. Conway D, McMahon A, Smith K. Components of socioeconomic risk associated with head and neck cancer: a population-based case-control study in Scotland. Br J Oral Maxillofac Surg. 2010;48(1):11-7.

13. Greenwood M, Thomson P, Lowry R, Steen I. Oral cancer: material deprivation, unemployment and risk factor behaviour—an initial study. Int J Oral Maxillofac Surg. 2003;32(1):74-7.

14. Woodley N, Slim MAM, Ton T, Montgomery J, Douglas C. Does recreational drug use influence survival and morbidity associated with laryngeal cancer. Addict Health. 2022;14(2):115-26.

15. Douglas CM, Middleton C, Sim P, Wight M, Young D, MacKenzie K, et al. Patterns of urgent hoarseness referrals to ENT—When should we be suspicious of cancer? Clin Otolaryngol. 2021;46(3):562-9.

16. Head and neck cancer incidence statistics Cancer Research UK2018 [Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/incidence>.

17. Overfitting IBM Cloud Education2021 [Available from: <https://www.ibm.com/cloud/learn/overfitting>.

18. Yang SJ, Kim JS, Chung SK, Song YY. Machine learning‐based model for prediction of outcomes in palatal surgery for obstructive sleep apnoea. Clin Otolaryngol. 2021;46(6):1242-6.

19. Ethics Self-Assessment Tool UK Statistics Authority [Available from: <https://uksa.statisticsauthority.gov.uk/the-authority-board/committees/national-statisticians-advisory-committees-and-panels/national-statisticians-data-ethics-advisory-committee/ethics-self-assessment-tool/>.

**List of Tables**

Table 1. Classification and frequency of diagnoses into benign, malignant potential and malignant outcome classification……………………………………………………………………………………………………………………………………………………3

Table 2. Demographics…………………………………………………………………………………………………………………………………….6

Table 3. Red Flag Symptoms…………………………………………………………………………………………………………………………….7

Table 4. Associated Symptoms………………………………………………………………………………………………………………………….8

Table 5. Top 5 variables deemed to be important for each model…………………………………………………………………….9