Computers in Biology and Medicine 180 (2024) 108978

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Insights from explainable AI in oesophageal cancer team decisions

Navamayooran Thavanesan [a,*], Arya Farahi [b], Charlotte Parfitt [c], Zehor Belkhatir [d], Tayyaba Azim [d], Elvira Perez Vallejos [e], Zoë Walters [a], Sarvapali Ramchurn [d], Timothy J. Underwood [a], Ganesh Vigneswaran [a]

[a] School of Cancer Sciences, Faculty of Medicine, University of Southampton, UK
[b] Department of Statistics and Data Science, University of Texas at Austin, United States
[c] University Hospitals Southampton NHS Foundation Trust, UK
[d] School of Electronics and Computer Science, University of Southampton, UK
[e] School of Computer Science, Horizon Digital Economy Research, University of Nottingham, UK

ARTICLE INFO

Keywords:
Machine learning
Oesophageal cancer
Multidisciplinary teams
Decision-making

ABSTRACT

Background: Clinician-led quality control into oncological decision-making is crucial for optimising patient care. Explainable artificial intelligence (XAI) techniques provide data-driven approaches to unravel how clinical variables influence this decision-making. We applied global XAI techniques to examine the impact of key clinical decision-drivers when mapped by a machine learning (ML) model, on the likelihood of receiving different oesophageal cancer (OC) treatment modalities by the multidisciplinary team (MDT).

Methods: Retrospective analysis of 893 OC patients managed between 2010 and 2022 at our tertiary unit, used a random forests (RF) classifier to predict four possible treatment pathways as determined by the MDT: neo-adjuvant chemotherapy followed by surgery (NACT + S), neoadjuvant chemoradiotherapy followed by surgery (NACRT + S), surgery-alone, and palliative management. Variable importance and partial dependence (PD) analyses then examined the influence of targeted high-ranking clinical variables within the ML model on treatment decisions as a surrogate model of the MDT decision-making dynamic.

Results: Amongst guideline-variables known to determine treatments, such as Tumour-Node-Metastasis (TNM) staging, age also proved highly important to the RF model (16.1 % of total importance) on variable importance analysis. PD subsequently revealed that predicted probabilities for all treatment modalities change significantly after 75 years ($p < 0.001$). Likelihood of surgery-alone and palliative therapies increased for patients aged 75–85yrs but lowered for NACT/NACRT. Performance status divided patients into two clusters which influenced all predicted outcomes in conjunction with age.

Conclusion: XAI techniques delineate the relationship between clinical factors and OC treatment decisions. These techniques identify advanced age as heavily influencing decisions based on our model with a greater role in patients with specific tumour characteristics. This study methodology provides the means for exploring conscious/subconscious bias and interrogating inconsistencies in team-based decision-making within the era of AI-driven decision support.

1. Introduction

As with all cancers managed within the UK, Oesophageal cancer (OC) treatment plans are determined by a multidisciplinary team (MDT). Since their introduction in the mid-1990s, they have been shown to improve cancer outcomes, especially within OC, which remains the 6th leading cause of cancer-related death globally and is still characterised by dismal 5 & 10-year survival rates [1,2]. With MDTs, the incidence of futile surgical procedures, operative mortality, and incomplete disease burden assessment ("cancer staging") dropped significantly [3–5]. However, this same framework which centralises a diverse group of domain experts in a single place and time is also

* Corresponding author. School of Cancer Sciences, Faculty of Medicine, University of Southampton, South Academic Block, University Hospitals Southampton, Tremona Road, Southampton, SO16 6YD, UK.
E-mail address: N.Thavanesan@soton.ac.uk (N. Thavanesan).
@TimTheSurgeon (T.J. Underwood), @ganesh_vignes (G. Vigneswaran)

https://doi.org/10.1016/j.compbiomed.2024.108978
Received 15 April 2024; Received in revised form 31 July 2024; Accepted 31 July 2024
Available online 5 August 2024

vulnerable to challenges stemming from increased caseload pressure, reduced preparation time, interpersonal dynamics. Perhaps most importantly they continue to experience inadequate time for reflection or self-audit for the decisions they make, thereby limiting experiential growth, and in some cancer types leading to a growing pursuit to pre-select MDT cases by complexity as a means of improving workflow [6–13].

Objective, data-driven insight into oncological decision-making allows clinicians to interrogate, validate and ensure the appropriateness of their treatment choices over time which directly impacts patient outcomes and quality of life, something exemplified in OC [14]. OC treatment decisions are highly complex; heavily influenced by primary tumour characteristics, metastatic spread and the physiological robustness of the patient [15,16]. Peak incidence is between 85 and 89 years, with this cohort often experiencing polypharmacy, poor nutrition, frailty and disability, all of which impact clinical outcomes [17,18]. Almost 80 % of patients over the age of 85 years have two or more co-morbidities (increasing co-morbidity is known to be a negative prognostic marker of 90-day mortality post-surgical resection [19]) leading to age historically acting as a barometer of perceived risk for intensive therapeutic interventions [20–23]. Judicious patient selection is critical; surgery alone is a monumental physiological stressor, further compounded by toxicity associated with neoadjuvant therapies (NAT), while even eligibility for palliative oncological therapies necessitates significant physiological reserve [24].

Additionally, while it is established that treatment decisions should be based on “physiological” over “chronological” age [25], it has recently been shown through ML that age plays a disproportionate role in treatment choice for curative OC patients at MDT. This bias is particularly evident when determining eligibility for multimodal versus unimodal therapy even when chronological age is not necessarily a guarantee of a negative outcome [26,27]. It is not yet clear whether this is a conscious or unconscious bias nor if there is interplay between age and a patient’s performance status (an oncological surrogate measure of baseline physical activity and thus a marker of resilience to otherwise deconditioning therapies). Implicit bias is a recognised aspect of healthcare, and while such bias has been reported for OC treatment allocation based on gender, race and socioeconomic status previously, how it manifests within more clinical parameters is currently unknown [28–31].

OC decision-making clearly carries high stakes, and yet while many of the clinical variables considered at MDT may be known or derived from guidelines, experience and current oncological doctrine [16,32,33], the relative weighting of these factors within the final decision is not currently known. This is salient when we consider the well-established literature surrounding vulnerabilities of cancer MDTs to inefficiency and sub-optimal decision-making in surgical oncology [7–9,11,13,34].

Machine Learning (ML), a branch of Artificial Intelligence (AI), is rapidly evolving within this aspect of healthcare, offering huge potential in multiple avenues relevant to OC. ML techniques can characterise complex patterns within current decision-making paradigms, inform future decision-making within human-AI and Group-AI collaborative (HAIC) processes, theoretically transforming multi-disciplinary team (MDT) efficiency [35–37]. Over the last decade, AI-based decision--support has also developed within MDT-type use cases with a view to changing the narrative from one of “human-versus-AI” to “human--and-AI” [38]. The architectures being tested within oncology have ranged from traditional tree-based ML models and neural networks, through complex natural-language decision-support systems aiming to assimilate up-to-date clinical knowledge such as IBM’s Watson, to more recently still, conversation-style, Large Language Model-based (LLM) architectures such as ChatGPT [39–42]. This utility of AI however must be balanced with sufficient transparency and explainability to preserve clinician-AI trust within the recommendations and insights generated [43–45].

Within OC there clearly remains a research gap in how clinicians routinely utilise clinical variables in for oncological decision-making. The aim of this study was therefore to demonstrate a viable approach to leveraging eXplainable AI (XAI) in order to characterise in-detail, the influence these clinical variables exert (of which some may have sub-conscious impact) on OC treatment decisions. Combining explainable ML techniques, our goal is to offer clinicians a clearer perspective into decision-making variation for OC patients in a trustworthy and explainable fashion. This in turn sets the foundations for trust in future Human-AI collaborations within the inevitable clinical decision-support space and represents a novel application of XAI in OC surgical oncology to date.

2. Materials and methods

This study was a retrospective complete-case analysis of oesophageal cancer patients at a single specialist cancer centre (University Hospitals Southampton) under the ethical approval of IRAS 233065.

2.1. Patient selection and data collection

OC patients who underwent MDT discussion from 2010 to 2022 were identified from a prospectively maintained oesophagectomy database combined with unit-submission records for the UK National Oesophagogastric Audit (NOGCA). Patients selected underwent either a curative pathway (surgery ± NAT) or a non-curative (palliative) pathway (best supportive care, palliative stenting, palliative chemotherapy, palliative radiotherapy or a combination thereof). Definitive chemoradiotherapy was excluded as this strategy occurred too infrequently for adequate model training. Clinical staging was assessed on baseline imaging, computer tomography (CT) and/or Positron Emission Tomography (PET), and tissue biopsies in accordance with the American Joint Committee on Cancer (AJCC) Tumour-Node-Metastasis (TNM) staging system.

2.2. Statistical analysis

Data analyses and model training were conducted using R (version 4.2.2) and Python (version 3.10.11). Sub-group comparison of continuous variables was made by Kruskal-Wallis analysis (adjusted with the Benjamini-Hochberg correction).

2.3. Data pre-processing and feature selection

Clinicopathological data within this study were analysed as structured tabular data. ‘Label encoder’ was employed within python to encode categorical variables for analysis. Features were selected through a combination of a priori domain expertise and established features form current UK clinical guidelines for OC management [16,19,27,32,33,46].

2.4. Treatment classification model development and performance

MDT treatment-decisions were modelled using a Random forests (RF) classifier in Python (“Ranger” Library, sklearn v1.2.2) using variables consistently available to the MDT prior to a final treatment decision (Table 1 & Supplemental Table 1). Using $k = 5$ cross validation, optimal max depth was determined as 6 which was used to train the final model on the whole dataset. The remaining hyper-parameters were set as default as RF models are not sensitive to small variations in these. The Random Forests algorithm is well-established and capable of handling higher-order interactions within classification tasks using both numerical and categorical features to produce strong predictive performance [47,48]. It has been utilised in numerous healthcare settings [49–51] and has already been shown to perform well in classification tasks as related to MDT treatment plans [48]. As this pilot study aimed to test

Table 1
Patient demographics and model predictor variables by sub-group. Referral unit statistics are provided in Supplementary Table. Performance status is measured as per the Eastern Cooperative Oncology Group (ECOG) Performance status scale.

Pre-treatment variables	"NACT + S" (N = 209) (%)	"NACRT + S" (N = 196) (%)	"Surgery-only" (N = 102) (%)	"Palliative" (N = 386) (%)	Total (N = 893) (%)
Gender					
Male	179 (85.6 %)	137 (69.9 %)	80 (78.4 %)	280 (72.5 %)	676 (75.7 %)
Female	30 (14.4 %)	59 (30.1 %)	22 (21.6 %)	106 (27.5 %)	217 (24.3 %)
Median Age, Years (Range)	65.7 (21–81.8)	66.6 (40.0–81.0)	73.4 (33.7–83.0)	74.8 (32.0–96.7)	69.1 (21.0–96.7)
ECOG Performance status					
0	120 (57.4 %)	138 (70.4 %)	34 (33.3 %)	68 (17.6 %)	360 (40.3 %)
1	84 (40.2 %)	54 (27.6 %)	56 (54.9 %)	122 (31.6 %)	316 (35.4 %)
2	5 (2.4 %)	3 (1.5 %)	12 (11.8 %)	124 (32.1 %)	144 (16.1 %)
3	0 (0 %)	1 (0.5 %)	0 (0 %)	69 (17.9 %)	70 (7.8 %)
4	0 (0 %)	0 (0 %)	0 (0 %)	3 (0.8 %)	3 (0.3 %)
cT stage					
1	0 (0 %)	0 (0 %)	8 (7.8 %)	1 (0.3 %)	9 (1.0 %)
2	35 (16.7 %)	44 (22.5 %)	49 (48.0 %)	40 (10.4 %)	168 (18.8 %)
3	149 (71.3 %)	138 (70.4 %)	43 (42.2 %)	211 (54.7 %)	541 (60.6 %)
4	25 (12.0 %)	14 (7.1 %)	2 (2.0 %)	134 (34.7 %)	175 (19.6 %)
cN stage					
0	40 (19.1 %)	64 (32.7 %)	53 (52.0 %)	82 (21.2 %)	239 (26.8 %)
1	138 (66.0 %)	112 (57.1 %)	42 (41.2 %)	131 (33.9 %)	423 (47.4 %)
2	31 (14.8 %)	19 (9.7 %)	6 (5.9 %)	121 (31.3 %)	177 (19.8 %)
3	0 (0 %)	1 (0.5 %)	1 (1.0 %)	52 (13.5 %)	54 (6.0 %)
cM stage					
0	209 (100 %)	196 (100 %)	102 (100 %)	162 (42.0 %)	669 (74.9 %)
1	0 (0 %)	0 (0 %)	0 (0 %)	224 (58 %)	224 (25.1 %)
Tumour location					
Oesophagus					
Proximal	0 (0 %)	3 (1.5 %)	0 (0 %)	18 (4.7 %)	21 (2.4 %)
Middle	5 (2.4 %)	22 (11.2 %)	7 (6.8 %)	59 (15.3 %)	93 (10.4 %)
Distal	103 (49.3 %)	148 (75.5 %)	64 (62.7 %)	235 (60.9 %)	550 (61.6 %)
GOJ					
GOJ Siewert 1	24 (11.5 %)	8 (4.1 %)	4 (3.9 %)	20 (5.2 %)	56 (6.3 %)
GOJ Siewert 2	39 (18.7 %)	10 (5.1 %)	19 (18.6 %)	54 (14.0 %)	122 (13.7 %)
GOJ Siewert 3	23 (11.0 %)	1 (0.5 %)	5 (4.9 %)	0 (0 %)	29 (3.2 %)
GOJ Siewert Undefined	15 (7.2 %)	4 (2.0 %)	3 (2.9 %)	0 (0 %)	22 (2.5 %)
Tumour Histology					
Adenocarcinoma	197 (94.3 %)	134 (68.4 %)	93 (91.2 %)	274 (71.0 %)	698 (78.1 %)
Squamous Cell (SCC)	12 (5.7 %)	62 (31.6 %)	9 (8.8 %)	112 (29.0 %)	195 (21.8 %)
Co-morbidities					
History of MI (MI)	9 (4.3 %)	11 (5.6 %)	10 (9.8 %)	34 (8.8 %)	64 (7.2 %)
Chronic heart failure (CHF)	1 (0.5 %)	1 (0.5 %)	2 (2.0 %)	17 (4.4 %)	21 (2.4 %)
Chronic pulmonary disease (CPD)	26 (12.4 %)	28 (14.3 %)	19 (18.6 %)	48 (12.4 %)	121 (13.5 %)
Connective tissue disease	2 (1.0 %)	5 (2.6 %)	1 (1 %)	0 (0 %)	8 (0.9 %)
Peripheral vascular disease (PVD)	6 (2.9 %)	7 (3.6 %)	5 (4.9 %)	21 (5.4 %)	39 (4.4 %)
Cerebrovascular disease (CVD)	8 (3.8 %)	6 (3.1 %)	7 (6.7 %)	65 (16.8 %)	86 (9.6 %)
Dementia	0 (0 %)	0 (0 %)	0 (0 %)	10 (2.6 %)	10 (1.1 %)
History of Peptic Ulcer Disease (XPUD)	8 (3.8 %)	7 (3.6 %)	5 (4.9 %)	14 (3.6 %)	34 (3.8 %)
Uncomplicated diabetes (DM uncomp)	21 (10.0 %)	20 (10.2 %)	16 (15.7 %)	60 (15.5 %)	117 (13.1 %)
Complicated diabetes (DM comp)	0 (0 %)	1 (0.5 %)	1 (1.0 %)	3 (0.8 %)	5 (0.6 %)
Leukaemia	0 (0 %)	0 (0 %)	3 (2.9 %)	1 (0.3 %)	4 (0.5 %)
Lymphoma	1 (0.5 %)	2 (1.0 %)	3 (2.9 %)	4 (1.0 %)	10 (1.1 %)
Mild liver disease	2 (1.0 %)	0 (0 %)	0 (0 %)	4 (1.0 %)	6 (0.7 %)
Hemiplegia	0 (0 %)	0 (0 %)	0 (0 %)	2 (0.5 %)	2 (0.2 %)
Renal failure	0 (0 %)	1 (0.5 %)	3 (2.9 %)	33 (8.5 %)	37 (4.1 %)
AIDS	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)	0 (0 %)

whether XAI techniques could enhance complex decision-making processes, unlike linear models, random forest models can capture interactions, non-linear relationships, is recognisable and accessible for the analysis and future reproducibility.

Year of diagnosis was incorporated into model training within defined time-periods (termed "Epoch" for the purposes of this study) relative to the dissemination of key randomised clinical trials to account for, (and assess changes in) clinical practice over time. Treatment outcomes were classified into neoadjuvant chemotherapy prior to surgery (NACT + S), neoadjuvant chemoradiotherapy prior to surgery (NACRT + S), surgery-only (Surgery) or palliative therapy (Palliative). Model performance was assessed via multi-class area-under-the-curve (AUC)/Receiver Operator Characteristic (ROC), balanced accuracy and calibration. For evaluating initial model-generalisability we used a 5-fold cross validated approach, following which hyperparameters were fixed allowing for training on the whole dataset. This preserves statistical power during partial dependence analysis as we discover what the model has learned [52,53]. For assessment of different algorithmic performances see Thavanesan et al., 2023 [48].

2.5. Variable importance analysis

Variable importance analysis of the final model was undertaken using all variables included for model-training (Table 1, Supplemental Table 1). The 'sk-learn' library function was employed; for RF, the significance of a feature is determined by averaging its value over all trees in the forest. Each characteristic gains greater significance as the impurity lowers. The total of these normalised importance values is 1.

2.6. Partial-dependence analysis

Partial-dependence (PD) analysis visualises how given predictor variables may influence predicted probabilities of a specified outcome across a range of values within the trained ML model including tree-based algorithms and allows for causal interpretations [53]. PD has been utilised previously to evaluate and explain predictive models in a wide array of use-cases [43,54–56]. PD selectively perturbs variables of interest incrementally while preserving the remaining variables to generate new predicted probabilities from the model after each perturbation. These may then be plotted either for individual patients (individualised conditional expectation plots), as an averaged curve, or as probability contours providing an intuitive, visual, model-agnostic approach to global interpretability of the ML model and so was chosen for this study especially as it offers causal interpretations.

Tools such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive exPlanations (SHAP) offer insight into predictions at the instance-level (although SHAP values can be aggregated over predictions to provide global insights too) [57,58]. Such local explanation tools are however principally beneficial in explaining predictions for individual patients once a clinical decision-support tool has already been deployed. PD, (as with variable importance) by comparison offers clinicians value earlier in the development of such HAIC processes by conveying global model interpretability as a surrogate microcosm of their team's decision-making paradigm and increasing trust in the validity of the underlying model as a result. While PD allows for causal interpretation, LIME creates new hyper-localised models for a given instance and is thus inappropriate for this, while SHAP has been shown to be unreliable in causal interpretations [59].

3. Results

3.1. Clinical cohort demographics

Of 938 initially identified cases, 13 were excluded as relating to patients who underwent failed endoscopic resection prior to salvage oesophagectomy. A further 32 cases with cT stages "cT0" (N = 4), "cTis" (N = 3) and "cTX" (N = 25) were excluded for low numbers and to allow examination of any ordinal relationships. The final cohort of 893 cases are summarised by predictor variable in Table 1 with additional referral unit data presented in Supplementary Table 1. Treatment-allocation over time was plotted to visualise general trends within the context of the landmark CROSS (NACRT + S) and FLOT4 (NACT + S) trials as well as assessed on PD by epoch for effect on treatment probabilities (Supplemental Figs. 1 and 2 respectively) [60,61].

3.2. Model performance

Classification performance for the RF classifier model using multi-class ROC AUCs is illustrated in Fig. 1. All classes were separable with excellent AUCs, (NACT + S 0.90, NACRT + S 0.88, Surgery-only 0.88, Palliative therapies 0.99) with reasonable calibration (Supplemental Fig. 3) and mean balanced accuracy (0.795 ± 0.008). This again aligns with previous experience of the use of random forest models classifying curative OC treatment plans in a smaller dataset [48].

3.3. Variable importance

Variables such as clinical TNM stage and tumour characteristics (location & histology) comprise standard criteria for treatment planning within national guidelines with cM stage and performance status key differentiators for curative versus palliative pathways [16]. On relative variable importance however, age notably ranked third when trained on the full cohort (Fig. 2a) after cM stage and performance status, followed by, epoch, cN stage, cT stage, referring location, tumour site and histological subtype. In view of its consistently high ranking, we focussed on age in PD analysis both in isolation and in combination with these variables to examine their interrelations further. Within a second 'curative-only' model age ranked first, further validating its focus within this study (Fig. 2b).

3.4. Influence of age on treatment decisions

Variation in treatment probability due to age alone was investigated using individual conditional expectation plots (Fig. 3). In all groups, a noticeable change in probabilities occurs after 75 years. Patients predicted for surgery-alone experience a probability rise between 75 and 85 years after which they return to pre-75-year baselines. For NAT, probabilities fall sharply after 75 years, however this decline starts as early as 70 years in the NACT + S group. Palliative pathway probabilities are largely consistent prior to 75 years however a clear upshift is seen beyond this time point.

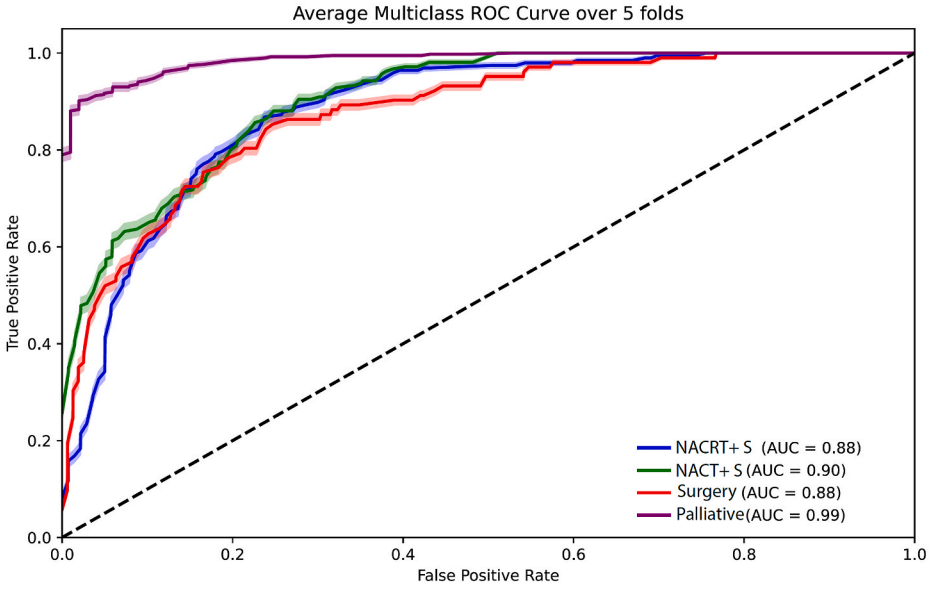

Fig. 1. Multiclass ROC curve for random forests treatment classifier representing a "one vs others" class-prediction performance. K = 5 Cross-validation was conducted using an 80:20 split. Mean ROC is presented ±1*x* Standard Error of the Mean.

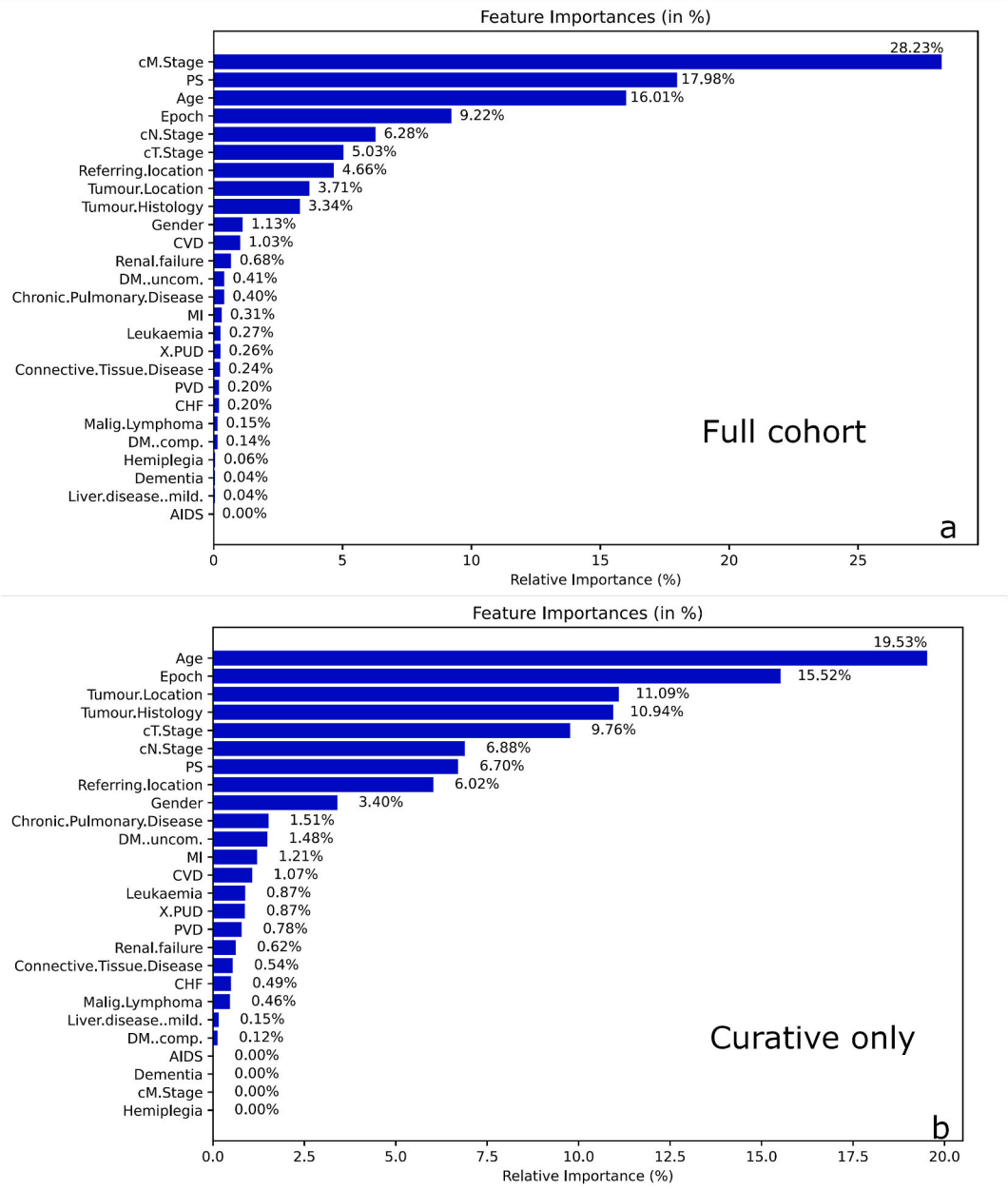

Fig. 2. Variable importance plot of relative importance for each predictor variable contributing to the Random Forests classifier model. Importance values are plotted for all patients (a) and curative patients only (b) in rank order with most important at the top. "Epoch" is a time variable split into three key time periods: "Pre-CROSS trial", "Cross-to-FLOT4" and "Post-FLOT4 trial".

The patient cohort was segregated into two subgroups (<75 years vs 75+ years) to statistically test for age-related differences between treatment classes (Supplemental Table 2). No significant difference was found between treatment groups within the younger subgroup or between NACT vs NACRT within the older cohort. A significant difference is seen between the palliative cohort against curative treatments as well as between Surgery and both NAT modalities within the older cohort.

3.5. Age vs tumour staging

The relationship between age and tumour staging (cT/cN) stage was assessed (Fig. 4a & b, purple regions represent low probability, yellow regions represent high probability). For surgery-alone strategies, age proved minimally influential under 75, directed instead by disease-stage. From 75 to 85yrs however, probabilities increase independently of staging. The probability contours demonstrated most variation for the surgery-alone group at approximately cT2 N0 indicating this group may experience significant variability in treatment plans. For NACT + S, highest likelihood (yellow) was focussed on cT3-4 N1 for under 75s after which likelihood dropped in line with advancing age. A similar pattern was observed for NACRT + S however the high probability zone is comparatively larger, extending from cT1-3 and cN0-1. For palliative therapies, advancing age acted synergistically with stage. As cM stage only applies to non-curative patients it could not be meaningfully assessed across pathways, however it demonstrates a binary influence across all treatments (Supplemental Fig. 4).

3.6. Age vs tumour characteristics

Tumour location demonstrated a hierarchical influence, conferring greater likelihood for surgery-alone strategies with progressively more distal tumours (Fig. 4c). A similar, exaggerated effect is seen in NACT +

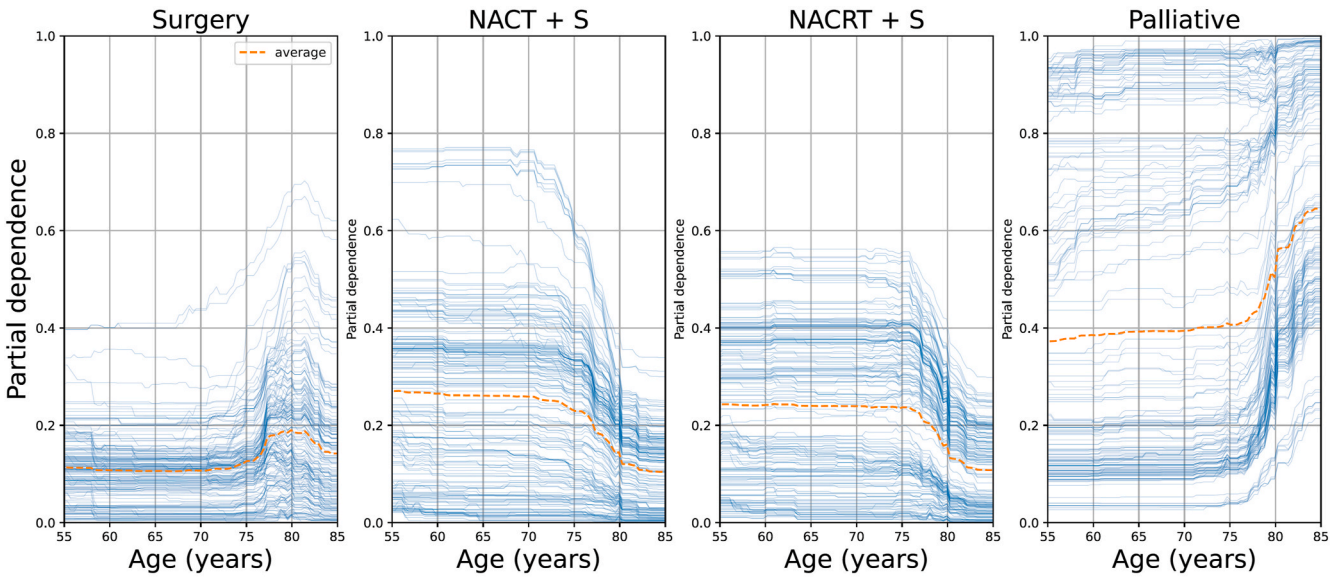

Fig. 3. Individual conditional expectation plots for predicted probability of treatment decision against age. Predicted probability (y axis) of each treatment pathway is plotted against the age range of the cohort (x axis) for each patient (blue lines). The averaged curve is also provided (orange dotted line).

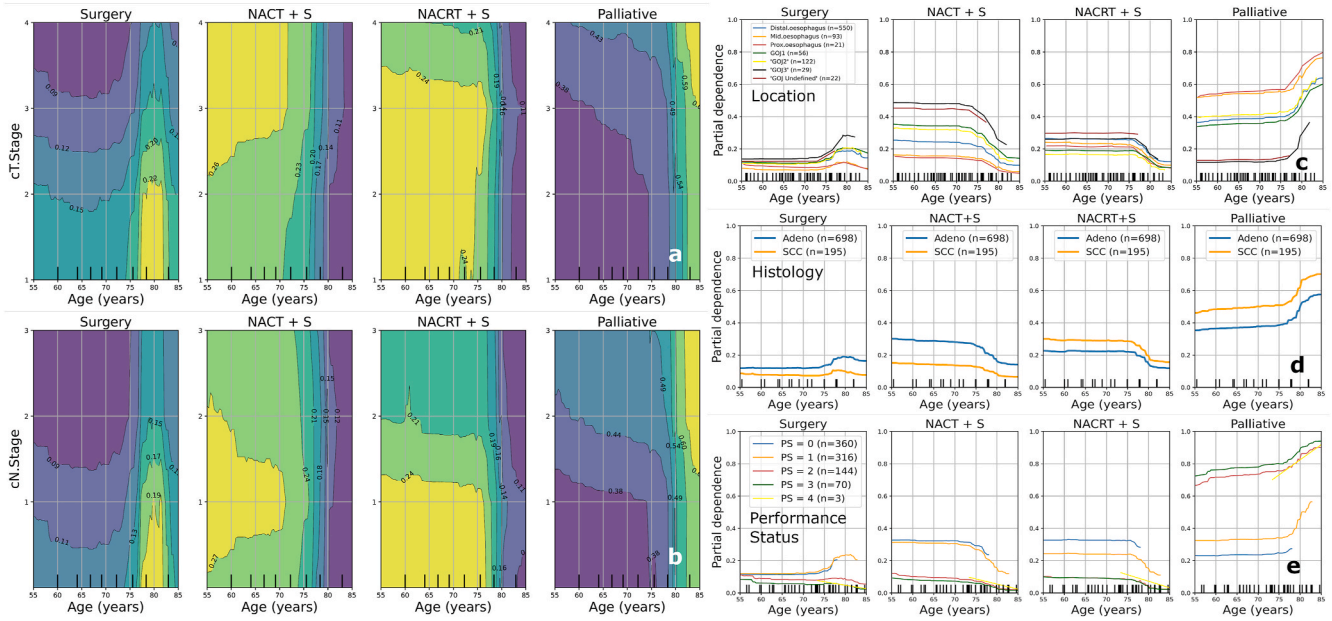

Fig. 4. 2-Dimensional Partial Dependence contour plot of Age vs cT Stage (a) and cN stage (b) on predicted probability of a treatment pathway. Averaged Partial Dependence Plot of Age vs Tumour Location (c), Tumour Histology (d) and Performance status (e) on treatment decision probability to visualise interrelationships between the covariates. Interrelation between disease the co-variate and patient age is mapped against four distinct OC treatment modalities: Surgery (S), NACT + S, NACRT + S and Palliative management. The x-axis delineates the age range of the patient cohort, while the y-axis captures the co-variate levels on a continuous axis. Intensity of the colour gradients within the contour plot signifies the likelihood of selecting a particular treatment, with yellow shades indicating higher probability while purple regions indicate lowest probability and numbered contours equate to that probability (e.g., 0.24 = 24 %).

S cases whereas this grouping is closer for NACRT + S. Mid-distal oesophageal tumours showed higher likelihood for NACRT + S while GOJ type 1–2 and proximal oesophageal tumours exhibited a lower probability. Proximal tumours were associated with highest likelihood for palliative pathways. Across modalities age continued to exert little influence under 75 years.

Histology separated base probabilities for all treatment choices independently of age (Fig. 4d). Irrespective of age, adenocarcinomas were more likely to receive surgery-only and NACT + S over SCCs which were more likely to be assigned NACRT + S or palliative pathways. Palliative therapy likelihood rose in step with advancing ages regardless of histology.

3.7. Age vs performance status

The relationship between age and PS demonstrated clear clustering into two patient cohorts across modalities: PS0-1 and PS2-4 (Fig. 4e). Under 75 years, age exerts minimal influence on surgery-alone probability. In older patients, PS0-1 cases experience a probability rise while PS2-4 patients follow a consistent low-probability trajectory, confirming that advanced age forced selection of the fittest patients for surgery-alone strategies. PS0-1 patients were significantly more likely to get

either NAT modality under 75 years after which probabilities re-converged with the PS2-4 cohort. PS2-4 patients were again much more likely to be assigned palliative pathway designation for across all age groups while PS1 patients only start to converge with the PS2-4 cohort after 75 years.

4. Discussion

4.1. Summary of findings

This study applied XAI techniques to quantify the influence specific clinical variables exert on the probability of a given treatment decision by the OC MDT. The study's findings of a model demonstrating strong AUC, balanced accuracy and calibration show that ML combined with XAI techniques can act as a vehicle to interrogate and analyse team-based decision-making dynamics with a granularity superior to classical statistical approaches. The ability to extract quantifiable objective insights, the majority of which align with observed clinical practice reinforces trust within the underlying model as a microcosm of the human MDT from which it draws inferences. As a proof-of-principle, the modelling in this study was not aimed towards clinical outcomes downstream of the decision (such as survival or quality of life), instead intentionally focussed on the route towards the treatment-decision itself in the first instance.

4.2. Age as a potential subconscious bias

Age, while not traditionally a criterion within management guidelines proved significant to OC treatment decisions, a finding consistent with our previous work which we are able to examine in detail here [48]. An important checkpoint within the seventh decade of life is highlighted which splits patients into two cohorts experiencing differing probabilities for treatment pathways. Patients over 75 years remain more likely to receive surgery-alone or palliative strategies and less likely to be offered NAT. As previous studies have historically highlighted a change in risk profile at 75 years this remains in keeping with our findings [62,63]. Furthermore, this study indicates age may act as a surrogate marker of patient fitness even in the presence of functional metrics such as performance status. We shared these findings with our MDT and asked if they recognised chronological age as an influential to their decision making. Initially, members believed that age was not a routine consideration in their decision-making. However, after engaging in reflective feedback sessions, they recognised that age did play a role, albeit subconsciously though they had not initially been able to place a specific age cut-off. This led some members to consider other possible subconscious influences and whether these were biases or simply based on experience [31].

4.3. Variability in treatment decisions

We have explained the relationship within our model between disease-staging co-variates and age, with the former more important in the under-75 group and the latter driving choices thereafter. Of interest, we observed the greatest variability in decision-making (depicted by a broad range of partial dependencies) for surgery-only strategies in those with cT2N0-1 disease. This fits a long-established controversy within the UK regarding the optimal management of this cohort. By definition cT2N0 disease breaches the muscularis propria with further potential for submucosal lymphatic invasion, leading to unpredictable tumoral behaviour within this group [64]. Compounded with historically high rates of under-staging, this cohort poses a therapeutic dilemma – utilise potentially toxic NAT (presuming undetected nodal disease) and risk deconditioning patients out of surgical fitness with potentially no additional survival advantage [65,66].

NAT decisions were mainly influenced by advancing age over staging with NACRT deployed over a wider age and staging range than NACT, but a drop in use of NAT altogether in older patients. This is attributable to a broadly held view that NACT regimes such as FLOT (Fluorouracil, Leucovorin, Oxaloplatin, Docetaxel) may be more toxic or less tolerated than NACRT [24,67–71]. It is worth noting however that while successful completion of all cycles for NACT regimens (e.g., pre- and post-operative FLOT) are lower versus NACRT, a high proportion still manage all pre-operative cycles to reach surgery [68,72]. Furthermore, concern over adverse effects with NACRT on tissue friability and anastomotic leakage rates intra- and post-operatively has prompted some Chinese units to favour NACT, even in OSCC for those with perceived poor treatment tolerance or frailty [69,71,73]. PD analysis suggested that NACT + S use within our unit dropped during epoch 2 (CROSS--FLOT4) but without rebounding post-FLOT4 as NACRT + S did after CROSS. This may be due to slower uptake by those keenly established in using NACRT + S especially while clinical equipoise persists regarding survival advantage. Modelling with trial epochs thus allows for changes in practice over time and requires periodic re-evaluation following future trials [68,74]. Predictably, staging was synergistic with age on palliative pathway prediction reflecting the combination of disease burden and frailty associated with advanced age.

4.4. Age as a surrogate marker of functional fitness

The interrelation between age and performance status is particularly interesting within this study as the former has historically been appropriated as a surrogate marker of frailty, prejudicing older patients away from aggressive treatments [17,63,75]. PD analysis grouped patients into two dominant clusters independently of age: PS0-1 versus PS2-4. The PS2-4 cluster experienced a significantly lower likelihood for NAT and were much more likely to be offered palliative treatments fitting a well-established prognostic significance of pre-treatment patient physical activity. Metabolic Equivalents or METS (measured by oxygen consumption at rest and used in anaesthesia to quantify perioperative functional capacity) are predictive of poor outcomes at scores of 4 or less [76]. Physical activity commensurate with such scores approximate to PS2 or worse, suggesting that this clustering reflects anticipation for treatment-related morbidity in this cohort. While national guidance on stratifying PS in curative OC cases is not currently offered, PD analysis allows for ML-driven benchmarking of observed clinical practice against current recommendations, a concept being explored in other surgical specialties [77].

4.5. Tumour characteristics on neoadjuvant therapy choice

Tumour characteristics also outweighed age in the under-75s in driving treatment probabilities. GOJ tumours were more likely to receive surgery-alone versus oesophageal lesions and significantly more likely to receive NACT than NACRT. This is partly over concern for collateral radiation-induced damage to the planned gastric conduit at surgery, and in part to a historical body of trial data focussed primarily on oesophageal tumours [60,78–81]. Across NAT, these decisions remain consistent until late into the 7th decade, at which point the deleterious effect of age is observed. High oesophageal tumours were additionally more likely to receive palliative outcomes versus distal lesions, in keeping with the significant challenges curative management for such lesions pose, and where resection in particular may be extensive [82]. Histology and age followed a similar pattern with adenocarcinomas more likely to receive surgery or NACT + S independently of age while SCCs were favoured for NACRT + S and palliative outcomes. This fits with the radiosensitivity of SCC subtypes coupled with greater potential for tumour response however a survival benefit from NACRT for adenocarcinomas however remains debateable [83,84].

4.6. Implications of this study

In 2016, Cancer Research UK, demonstrated that MDTs within the

UK were operating under significant strain and resource scarcity [6]. Among many of their key findings was a significant challenge for MDTs finding the time to audit and reflect on their decision-making processes. Although numerous studies have, in recent years, demonstrated the capabilities of AI to support, replicate, or even beat the human clinician in clinical tasks, none to date have considered the benefit of AI in auditing or unpicking the human decision-making process [85]. This study shows that AI may also provide significant benefit as vehicle for early-warnings of subtle shifts in practice, sub-conscious or even unconscious bias, and identifying areas where variability indicates a definitive knowledge gap which may in turn guide research questions downstream. XAI techniques offer the best way forward by championing accurate, capable high-functioning AI while balancing this with the need for transparent, auditable processes. When working in symbiosis with human counterparts within the MDT, this can provide for the ideal of "AI-augmented clinicians" [38].

4.7. Study limitations and strengths

This was a single-centre retrospective analysis of 893 OC patients over a 13-year period during which a number of shifts in oncological practice have undoubtedly occurred in both NACT regimens and emerging immunotherapies. However, little clarity has been achieved even now in optimal NAT regimens or management of cT2N0 patients. The strength of this study is in its novel use of XAI on a large single-centre cohort of nearly 900 patients evaluating both curative and non-curative treatment pathways which broadens its generalisability. We have demonstrated how transparency can be introduced for team-based oncological treatment decisions, detailing clear shifts in human decision-making when faced with specific clinicopathological scenarios in oncological settings known to suffer chaotic leadership styles [86]. This approach allows us to examine and re-examine the robustness of our decision-making to standardise practice for OC patients (especially given that there is evidence to indicate heterogeneity of decision-making even between OC MDTs [87]) and can be applied to other MDTs regionally, nationally or internationally in future for direct comparison of MDTs as well as being translatable to MDTs from other cancer types. Where MDTs have little time across cancer types for audit, self-reflection or learning [6,7,11,12,88], global XAI could be integrated into MDT workflows as part of annual departmental audits for quality control, sense-checking shifts in practice. Training data drift can be tracked to ensure models remain appropriate and true to the local population [89]. As CDSS tools the evolve, local XAI techniques such as LIME and SHAP may be integrated within the user-interface to offer additional instance-level explanations in real-time tailored to the individual patient [57,58]. While the present study is not designed to determine the clinical justification for decisions influenced by variables such as age, it highlights scenarios for MDTs to focus upon during clinical governance processes while introducing clinicians to the capabilities of AI-derived decision support.

However, while XAI techniques explain recommendations, this does not automatically guarantee clinician uptake of that recommendation. A recent study testing clinicians' fluid-prescriptions when offered additional advice from simple AI or XAI noted little difference on self-reporting, in outcome whether explanations were provided or not, questioning whether explanations were of material benefit above an AI recommendation [90]. The study faced some methodological challenges, namely the reliability of self-reporting, sample size and the generalisability of the clinical scenario. The question it raises however is valid, engagement often depends on the user's level of technical understanding, the effectiveness of communication methods for explanations, and whether clinicians perceive the explanations as beneficial beyond ML experts [37,91]. Despite this, the prevailing wind within healthcare remains a need for trustable AI solutions which open the "black box". Bridging the gap to non-technical clinicians must the occur through education programs to ensure they can critically appraise not only AI models but the explanations which may accompany their outputs [92].

4.8. Future work

Future work will include applying the technique to external centres to compare and contrast our findings both within OC but also in other cancer-types. Testing other well-known ML algorithms from more inherently interpretable options such as decision-tree models and more complex ensemble learners such as eXtreme Gradient Boost may also be useful in evaluating insights across algorithms. Furthermore, ongoing work within this space will inevitably lead to the co-development of ML-derived decision-support tools trained on human-led MDT decisions. By applying Responsible Research and Innovation (RRI) frameworks we are currently engaging with and including stakeholders' opinions (oncologists, radiologists, psychologists, computer scientists, and patient representatives) to identify strategies for optimal implementation, acceptability and usability of such an ML-derived tool [93]. We are incorporating principles to support multidisciplinary scientific collaboration, anticipate key future challenges and reflect on better practices for responsible data governance. The need for explainable and preferably interpretable models built on RRI principles is paramount, especially now with the potential for AI in healthcare to transform the practice of medicine in general. The approach presented here represents a route towards trust within these frameworks by first offering global insight into team-level decision-making when mirrored by ML. While the model used in this study is tailored to our local MDT, the process can be performed either on a population-level for scalability or targeted to a specific unit, to enhance data diversity and representativeness of underrepresented demographic groups or geographical areas). Understanding decision-drivers, some of which we argue are sub-conscious in practice, is invaluable for the pursuit of personalised medicine for OC patients and essential in building clinician-patient trust in future implementations of AI within OC.

5. Conclusion

This study applied XAI methods to highlight how significant, yet sometimes subconscious factors like age may influence treatment decisions for OC patients. While treatment choices are often framed by clinical factors, age remains salient even with functional metrics like performance status delineating patients into a fitter cohort, more likely to undergo all curative treatments, versus an unfit less-eligible group. The uniformity in predicted probabilities for curative treatments persists only until the 7th decade of life. After this, a notable rise in the probability for surgery-alone and palliative options is juxtaposed against a decline in neoadjuvant therapy (NAT) prospects. Our analysis not only emphasizes age's pivotal role amidst traditional clinical drivers but also showcases the clarity and insight achievable with ML in navigating complex treatment landscapes. This explainability is crucial for clinician-engagement and trust within future AI-based decision support tools.

Funding support acknowledgement

NT receives a joint studentship from the Institute for Life Sciences (University of Southampton) and University Hospitals Southampton. The project receives funding from the UKRI Trustworthy Autonomous Systems Hub (TASHub) Pump Priming Fund.

7. Conflicts of interests to declare

None.

Data availability

The data included within this study relates to sensitive clinical patient data and so cannot be shared.

CRediT authorship contribution statement

Navamayooran Thavanesan: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Arya Farahi:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Charlotte Parfitt:** Writing – review & editing, Methodology, Investigation, Data curation. **Zehor Belkhatir:** Writing – review & editing, Methodology. **Tayyaba Azim:** Writing – review & editing, Methodology. **Elvira Perez Vallejos:** Writing – review & editing, Methodology, Investigation. **Zoë Walters:** Writing – review & editing, Supervision, Investigation. **Sarvapali Ramchurn:** Writing – review & editing, Supervision, Methodology, Investigation. **Timothy J. Underwood:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition. **Ganesh Vigneswaran:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

.

All authors declare that they have no potential conflicts of interest.

Acknowledgements

For the purpose of open access, the authors have applied a Creative Commons attribution license (CC-BY) to any Author Accepted Manuscript version arising from this submission.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2024.108978.

Glossary

Co-morbidity the simultaneous presence of two or more clinical conditions within a patient

CROSS trial A landmark clinical trial published in 2012 which demonstrated the survival benefit of providing neoadjuvant chemoradiotherapy prior to surgery for oesophageal cancer patients

FLOT4 trial A landmark clinical trial published in 2017 which demonstrated the efficacy of the neoadjuvant chemotherapy regimen of 5-Fluoro Uracil (5FU), Oxaloplatin, Leucovorin and Docetaxel over the previous gold-standard chemotherapy regimen of the day for OC

Multidisciplinary Teams This is a clinical framework for shared decision-making in cancer care defined by the presence of multiple separate domain experts at the time of determining a course of treatments. MDTs typically comprise cancer surgeons, oncologists, radiologists, pathologists, specialist nurses, palliative care physicians, administrative and clerical team members as well as many other allied health care professions

Multimodal therapy the use of multiple oncological strategies for a patient's cancer care. Within Oesophageal Cancer this specifically relates to the use of neoadjuvant therapies as well as formal surgical resection

Neoadjuvant therapy the provision of cancer treatments (typically chemotherapy, radiotherapy, hormone therapy or immunotherapy) prior to formal surgical resection of a tumour to downstage (shrink or improve the size and invasion of) the cancer

Oesophagectomy the surgical removal of part of the oesophagus and typically a portion of the proximal stomach

Performance Status The Eastern Cooperative Oncology Group (ECOG) performance status is a clinical grading scale from 0 to 5 (0 = Fully active, able to carry on all pre-disease performance without restriction, 5 = dead) which has been traditionally used within surgical oncology to evaluate a patient's physical activity levels as a barometer of physiological reserve and fitness for therapeutic interventions

Polypharmacy The simultaneous use of multiple medications

Tertiary referral unit A specialist clinical centre or hospital which has particular expertise in managing a specific clinical condition

Unimodal therapy within Oesophageal cancer this typically relates to forgoing neoadjuvant therapy and proceeding directly to surgery

References

[1] S. Lander, E. Lander, M.K. Gibson, Esophageal cancer: overview, risk factors, and reasons for the rise, Curr. Gastroenterol. Rep. 25 (2023) 275–279, https://doi.org/10.1007/s11894-023-00899-0.

[2] Cancer Research UK, Cancer research UK oesophageal cancer survival. https://www.cancerresearchuk.org/about-cancer/oesophageal-cancer/survival, 2023. (Accessed 27 October 2023).

[3] B. Publishing Asia, M.R. Stephens, W.G. Lewis, A.E. Brewster, I. Lord, G.R.J. C. Blackshaw, I. Hodzovic, G. V Thomas, S.A. Roberts, T.D. L Crosby, C. Gent, M. C. Allison, K. Shute, Multidisciplinary team management is associated with improved outcomes after surgery for esophageal cancer. https://academic.oup.com/dote/article/19/3/164/2194890, 2006.

[4] R.K. Freeman, J.M. Van Woerkom, A. Vyverberg, A.J. Ascioti, The effect of a multidisciplinary thoracic malignancy conference on the treatment of patients with esophageal cancer, Ann. Thorac. Surg. 92 (2011) 1239–1243, https://doi.org/10.1016/j.athoracsur.2011.05.057.

[5] P. Van Hagen, M.C.W. Spaander, A. Van Der Gaast, C.M. Van Rij, H.W. Tilanus, J.J. B. Van Lanschot, B.P.L. Wijnhoven, Impact of a multidisciplinary tumour board meeting for upper-GI malignancies on clinical decision making: a prospective cohort study, Int. J. Clin. Oncol. 18 (2013) 214–219, https://doi.org/10.1007/s10147-011-0362-8.

[6] R. Gray, B. Gordon, M. Meredith, CRUK: meeting patients' Needs - improving the effectiveness of multidisciplinary team meetings in cancer services. www.cancerresearchuk.org, 2017.

[7] B.W. Lamb, K.F. Brown, K. Nagpal, C. Vincent, J.S.A. Green, N. Sevdalis, Quality of care management decisions by multidisciplinary cancer teams: a systematic review, Ann. Surg Oncol. 18 (2011) 2116–2125, https://doi.org/10.1245/s10434-011-1675-6.

[8] T. Soukup, B.W. Lamb, A. Morbi, N.J. Shah, A. Bali, V. Asher, T. Gandamihardja, P. Giordano, A. Darzi, N. Sevdalis, J.S.A. Green, Cancer multidisciplinary team meetings: impact of logistical challenges on communication and decision-making, BJS Open 6 (2022), https://doi.org/10.1093/bjsopen/zrac093.

[9] T. Soukup, A. Morbi, B.W. Lamb, T.A.K. Gandamihardja, K. Hogben, K. Noyes, T. A. Skolarus, A. Darzi, N. Sevdalis, J.S.A. Green, A measure of case complexity for streamlining workflow in multidisciplinary tumor boards: mixed methods development and early validation of the MeDiC tool, Cancer Med. 9 (2020) 5143–5154, https://doi.org/10.1002/cam4.3026.

[10] J. Wihl, V. Falini, S. Borg, O. Stahl, T. Jiborn, B. Ohlsson, M. Nilbert, Implementation of the measure of case discussion complexity to guide selection of prostate cancer patients for multidisciplinary team meetings, Cancer Med. 12 (2023) 15149–15158, https://doi.org/10.1002/cam4.6189.

[11] B.W. Lamb, N. Sevdalis, S. Arora, A. Pinto, C. Vincent, J.S.A. Green, Teamwork and team decision-making at multidisciplinary cancer conferences: barriers, facilitators, and opportunities for improvement, World J. Surg. 35 (2011) 1970–1976, https://doi.org/10.1007/s00268-011-1152-1.

[12] B. Lamb, J.S.A. Green, C. Vincent, N. Sevdalis, Decision making in surgical oncology, Surg Oncol 20 (2011) 163–168, https://doi.org/10.1016/j.suronc.2010.07.007.

[13] K.C.W.J. Ebben, M.S. Sieswerda, E.J.T. Luiten, J.B. Heijns, C.C. Van Der Pol, M. Bessems, A.H. Honkoop, Mathijs, P. Hendriks, J. Verloop, X.A.A.M. Verbeek, Impact on Quality of Documentation and Workload of the Introduction of a National Information Standard for Tumor Board Reporting (2020), 10.

[14] S.E. Al-Batran, J.A. Ajani, Impact of chemotherapy on quality of life in patients with metastatic esophagogastric cancer, Cancer 116 (2010) 2511–2518, https://doi.org/10.1002/cncr.25064.

[15] R.C. Walker, T.J. Underwood, Oesophageal cancer, Surgery 35 (2017) 627–634, https://doi.org/10.1016/j.mpsur.2017.09.010.

[16] National Institute for Health and Care Excellence, Oesophago-gastric cancer: assessment and management in adults NICE guideline. www.nice.org.uk/guidance/ng83, 2018.

[17] H.S. Lin, J.N. Watts, N.M. Peel, R.E. Hubbard, Frailty and post-operative outcomes in older surgical patients: a systematic review, BMC Geriatr. 16 (2016), https://doi.org/10.1186/s12877-016-0329-8.

[18] Cancer Research UK, Oesophageal cancer incidence statistics. Oesophageal Cancer Incidence Statistics, 2021. (Accessed 19 April 2023).

[19] Z. Cheng, A. Johar, E. Gottlieb-Vedi, M. Nilsson, J. Lagergren, P. Lagergren, Impact of co-morbidity on reoperation or death within 90 days of surgery for oesophageal cancer, BJS Open 5 (2021), https://doi.org/10.1093/bjsopen/zraa035.

[20] J. Racz, L. Dubois, A. Katchky, W. Wall, Elective and emergency abdominal surgery in patients 90 years of age or older, Can. J. Surg. 55 (2012) 322–328, https://doi.org/10.1503/cjs.007611.

[21] E.A. McGillicuddy, K.M. Schuster, K.A. Davis, W.E. Longo, Factors predicting morbidity and mortality in emergency colorectal procedures in elderly patients, Arch. Surg. 144 (2009) 1157–1162, https://doi.org/10.1001/archsurg.2009.203.

[22] D.S. Belinda, C. Fausto, Emergency surgery for colorectal cancer in patients aged over 90 Years: review of the recent literature, Journal of Tumour 4 (2016).

[23] B. Guthrie, K. Barnett, S.W. Mercer, M. Norbury, G. Watt, S. Wyke, Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study, Lancet 380 (2012) 37–43, https://doi.org/10.1016/S0140.

[24] L. Depypere, M. Thomas, J. Moons, W. Coosemans, T. Lerut, H. Prenen, K. Haustermans, H. Van Veer, P. Nafteux, Analysis of patients scheduled for neoadjuvant therapy followed by surgery for esophageal cancer, who never made it to esophagectomy, World J. Surg. Oncol. 17 (2019), https://doi.org/10.1186/s12957-019-1630-8.

[25] N. Ahamat, Access all ages: assessing the impact of age on access to surgical treatment, Bull. Roy. Coll. Surg. Engl. 94 (2012) 300, https://doi.org/10.1308/147363512x13448516926748.

[26] N. Thavanesan, I. Bodala, Z. Walters, S. Ramchurn, T.J. Underwood, G. Vigneswaran, Machine learning to predict curative multidisciplinary team treatment decisions in oesophageal cancer, Eur. J. Surg. Oncol. (2023) 106986, https://doi.org/10.1016/j.ejso.2023.106986.

[27] J.M. McLoughlin, J.M. Lewis, K.L. Meredith, The impact of age on morbidity and mortality following esophagectomy for esophageal cancer, Cancer Control 20 (2013) 144–150, https://doi.org/10.1177/107327481302000208.

[28] K.E. Henson, A. Fry, G. Lyratzopoulos, M. Peake, K.J. Roberts, S. McPhail, Sociodemographic variation in the use of chemotherapy and radiotherapy in patients with stage IV lung, oesophageal, stomach and pancreatic cancer: evidence from population-based data in England during 2013-2014, Br. J. Cancer 118 (2018) 1382–1390, https://doi.org/10.1038/s41416-018-0028-7.

[29] I.C. Okereke, J. Westra, D. Tyler, S. Klimberg, D. Jupiter, R. Venkatesan, K. Brooks, Y.-F. Kuo, Disparities in esophageal cancer care based on race: a National Cancer Database analysis, Dis. Esophagus 35 (2022), https://doi.org/10.1093/dote/doab083.

[30] M.C. Kalff, W.P.M. Dijksterhuis, A.D. Wagner, S. Oertelt-Prigione, R.H. A. Verhoeven, V.E.P.P. Lemmens, H.W.M. van Laarhoven, S.S. Gisbertz, M.I. van Berge Henegouwen, Sex differences in treatment allocation and survival of potentially curable gastroesophageal cancer: a population-based study, Eur. J. Cancer 187 (2023) 114–123, https://doi.org/10.1016/j.ejca.2023.04.002.

[31] D.P. Gopal, U. Chetty, P. O'Donnell, C. Gajria, J. Blackadder-Weinstein, Implicit bias in healthcare: clinical practice, research and decision making, Future Healthc J 8 (2021) 40–48, https://doi.org/10.7861/fhj.2020-0233.

[32] R.C. Walker, T.J. Underwood, Oesophageal cancer, Surgery 35 (2017) 627–634, https://doi.org/10.1016/j.mpsur.2017.09.010.

[33] S.L. Favareto, C.F. Sousa, P.J. Pinto, H. Ramos, M.J. Chen, D.G. Castro, M.L. Silva, G. Gondim, A.C.A. Pellizzon, R.C. Fogaroli, Clinical prognostic factors for patients with esophageal cancer treated with definitive chemoradiotherapy, Cureus (2021), https://doi.org/10.7759/cureus.18894.

[34] B. Lamb, J.S.A. Green, C. Vincent, N. Sevdalis, Decision making in surgical oncology, Surg Oncol 20 (2011) 163–168, https://doi.org/10.1016/j.suronc.2010.07.007.

[35] N. Thavanesan, G. Vigneswaran, I. Bodala, T.J. Underwood, The oesophageal cancer multidisciplinary team: can machine learning assist decision-making? J. Gastrointest. Surg. (2023) https://doi.org/10.1007/s11605-022-05575-8.

[36] N. Farzaneh, S. Ansari, E. Lee, K.R. Ward, M.W. Sjoding, Collaborative strategies for deploying artificial intelligence to complement physician diagnoses of acute respiratory distress syndrome, NPJ Digit Med 6 (2023), https://doi.org/10.1038/s41746-023-00797-9.

[37] M. Naiseh, C. Webb, T. Underwood, G. Ramchurn, Z. Walters, N. Thavanesan, G. Vigneswaran, XAI for group-AI interaction: towards collaborative and inclusive explanations. https://orcid.org/0000-0002-4927-5086, 2024.

[38] A. Di Ieva, AI-augmented multidisciplinary teams: hype or hope?. https://doi.org/10.1016/S01406736(19)326261, 2019.

[39] E. Vu, N. Steinmann, C. Schröder, R. Förster, D.M. Aebersold, S. Eychmüller, N. Cihoric, C. Hertler, P. Windisch, D.R. Zwahlen, Applications of machine learning in palliative care: a systematic review, Cancers 15 (2023), https://doi.org/10.3390/cancers15051596.

[40] J.M. Choo, H.S. Ryu, J.S. Kim, J.Y. Cheong, S.J. Baek, J.M. Kwak, J. Kim, Conversational artificial intelligence (chatGPT^{TM}) in the management of complex colorectal cancer patients: early experience, ANZ J. Surg. 94 (2024) 356–361, https://doi.org/10.1111/ans.18749.

[41] Y. Tjhin, B. Kewlani, H.K.S.I. Singh, N. Pawa, Artificial intelligence in colorectal multidisciplinary team meetings. What are the medicolegal implications? Colorectal Dis. (2024) https://doi.org/10.1111/codi.17091.

[42] K. Lee, S.H. Lee, Artificial intelligence-driven oncology clinical decision support system for multidisciplinary teams, Sensors 20 (2020) 1–12, https://doi.org/10.3390/s20174693.

[43] R. Elshawi, M.H. Al-Mallah, S. Sakr, On the interpretability of machine learning-based model for predicting hypertension, BMC Med. Inf. Decis. Making 19 (2019), https://doi.org/10.1186/s12911-019-0874-0.

[44] J. Zhang, Z. ming Zhang, Ethics and governance of trustworthy medical artificial intelligence, BMC Med. Inf. Decis. Making 23 (2023), https://doi.org/10.1186/s12911-023-02103-9.

[45] U. Upadhyay, A. Gradisek, U. Iqbal, E. Dhar, Y.C. Li, S. Syed-Abdul, Call for the responsible artificial intelligence in the healthcare, BMJ Health Care Inform 30 (2023), https://doi.org/10.1136/bmjhci-2023-100920.

[46] P.R. Boshier, A. Swaray, B. Vadhwana, A. O'sullivan, D.E. Low, G.B. Hanna, C. J. Peters, Systematic review and validation of clinical models predicting survival after oesophagectomy for adenocarcinoma, Br. J. Surg. 109 (2022) 418–425, https://doi.org/10.1093/bjs/znac044.

[47] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[48] N. Thavanesan, I. Bodala, Z. Walters, S. Ramchurn, T.J. Underwood, G. Vigneswaran, Machine learning to predict curative multidisciplinary team treatment decisions in oesophageal cancer, Eur. J. Surg. Oncol. (2023) 106986, https://doi.org/10.1016/j.ejso.2023.106986.

[49] S.A. Rahman, R.C. Walker, N. Maynard, N. Trudgill, T. Crosby, D.A. Cromwell, T. J. Underwood, NOGCA project team AUGIS, The AUGIS survival predictor: prediction of long-term and conditional survival after esophagectomy using random survival forests, Ann. Surg. 277 (2023) 267–274, https://doi.org/10.1097/SLA.0000000000004794.

[50] Y. Zhang, Z. Zhang, L. Wei, S. Wei, Construction and validation of nomograms combined with novel machine learning algorithms to predict early death of patients with metastatic colorectal cancer, Front. Public Health 10 (2022) 1008137, https://doi.org/10.3389/fpubh.2022.1008137.

[51] Y. Ren, Y. Zhang, J. Zhan, J. Sun, J. Luo, W. Liao, X. Cheng, Machine learning for prediction of delirium in patients with extensive burns after surgery, CNS Neurosci. Ther. 29 (2023) 2986–2997, https://doi.org/10.1111/cns.14237.

[52] J.H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, 2001.

[53] A. Goldstein, A. Kapelner, J. Bleich, E. Pitkin, Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation, J. Comput. Graph Stat. 24 (2015) 44–65, https://doi.org/10.1080/10618600.2014.907095.

[54] J. Park, W.H. Lee, K.T. Kim, C.Y. Park, S. Lee, T.Y. Heo, Interpretation of ensemble learning to predict water quality using explainable artificial intelligence, Sci. Total Environ. 832 (2022), https://doi.org/10.1016/j.scitotenv.2022.155070.

[55] Y. Feng, X. Wang, J. Zhang, A heterogeneous ensemble learning method for neuroblastoma survival prediction, IEEE J Biomed Health Inform 26 (2022) 1472–1483, https://doi.org/10.1109/JBHI.2021.3073056.

[56] C. Yu, Y. Li, M. Yin, J. Gao, L. Xi, J. Lin, L. Liu, H. Zhang, A. Wu, C. Xu, X. Liu, Y. Wang, J. Zhu, Automated machine learning in predicting 30-day mortality in patients with non-cholestatic cirrhosis, J. Personalized Med. 12 (2022), https://doi.org/10.3390/jpm12111930.

[57] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you?. Explaining the Predictions of Any Classifier, 2016.

[58] S. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, 2017.

[59] S. Ma, R. Tourani, E. Thuc, D. Le, L. Liu, K. Zhang, E. Kıcıman, P. Cui, A. Hyvärinen, Predictive and causal implications of using Shapley value for model interpretation. http://jmlr.org/papers/v/.html, 2020.

[60] J. Shapiro, J.J.B. van Lanschot, M.C.C.M. Hulshof, P. van Hagen, M.I. van Berge Henegouwen, B.P.L. Wijnhoven, H.W.M. van Laarhoven, G.A.P. Nieuwenhuijzen, G.A.P. Hospers, J.J. Bonenkamp, M.A. Cuesta, R.J.B. Blaisse, O.R.C. Busch, F.J. W. ten Kate, G.J.M. Creemers, C.J.A. Punt, J.T.M. Plukker, H.M.W. Verheul, E.J. S. Bilgen, H. van Dekken, M.J.C. van der Sangen, T. Rozema, K. Biermann, J. C. Beukema, A.H.M. Piet, C.M. van Rij, J.G. Reinders, H.W. Tilanus, E. W. Steyerberg, A. van der Gaast, Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial, Lancet Oncol. 16 (2015) 1090–1098, https://doi.org/10.1016/S1470-2045(15)00040-6.

[61] S.E. Al-Batran, N. Homann, C. Pauligk, T.O. Goetze, J. Meiler, S. Kasper, H. G. Kopp, F. Mayer, G.M. Haag, K. Luley, U. Lindig, W. Schmiegel, M. Pohl, J. Stoehlmacher, G. Folprecht, S. Probst, N. Prasnikar, W. Fischbach, R. Mahlberg, J. Trojan, M. Koenigsmann, U.M. Martens, P. Thuss-Patience, M. Egger, A. Block, V. Heinemann, G. Illerhaus, M. Moehler, M. Schenk, F. Kullmann, D.M. Behringer, M. Heike, D. Pink, C. Teschendorf, C. Löhr, H. Bernhard, G. Schuch, V. Rethwisch, L.F. von Weikersthal, J.T. Hartmann, M. Kneba, S. Daum, K. Schulmann, J. Weniger, S. Belle, T. Gaiser, F.S. Oduncu, M. Güntner, W. Hozaeel, A. Reichart, E. Jäger, T. Kraus, S. Mönig, W.O. Bechstein, M. Schuler, H. Schmalenberg, R. D. Hofheinz, Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial, Lancet 393 (2019) 1948–1957, https://doi.org/10.1016/S0140-6736(18)32557-1.

[62] J. Lagergren, M. Bottai, G. Santoni, Patient age and survival after surgery for esophageal cancer, Ann. Surg Oncol. 28 (2021) 159–166, https://doi.org/10.1245/s10434-020-08653-w.

[63] D. Tougeron, H. Hamidou, M. Scotté, F. Di Fiore, M. Antonietti, B. Paillot, P. Michel, Esophageal cancer in the elderly: an analysis of the factors associated with treatment decisions and outcomes. https://doi.org/10.1186/1471-2407-10-510, 2010.

[64] P. Vining, T.J. Birdas, Management of clinical T2N0 esophageal cancer: a review, J. Thorac. Dis. 11 (2019) S1629–S1632, https://doi.org/10.21037/jtd.2019.07.85.

[65] J.P. Dolan, T. Kaur, B.S. Diggs, R.A. Luna, B.C. Sheppard, P.H. Schipper, B.H. Tieu, G. Bakis, G.M. Vaccaro, J.M. Holland, K.M. Gatter, M.A. Conroy, C.A. Thomas, J. G. Hunter, Significant understaging is seen in clinically staged T2N0 esophageal cancer patients undergoing esophagectomy, Dis. Esophagus 29 (2016) 320–325, https://doi.org/10.1111/dote.12334.

[66] S.R. Markar, C. Gronnier, A. Pasquer, A. Duhamel, H. Beal, J. Théreaux, J. Gagnière, G. Lebreton, C. Brigand, B. Meunier, D. Collet, C. Mariette, Role of neoadjuvant treatment in clinical T2N0M0 oesophageal cancer: results from a retrospective multi-center European study, Eur. J. Cancer 56 (2016) 59–68, https://doi.org/10.1016/j.ejca.2015.11.024.

[67] P.S.N. van Rossum, H.W.M. van Laarhoven, CROSS versus modified MAGIC or FLOT in oesophageal and gastro-oesophageal junction adenocarcinoma, Lancet Gastroenterol Hepatol (2023), https://doi.org/10.1016/S2468-1253(23)00278-9.

[68] J. V Reynolds, S.R. Preston, B. O'Neill, M.A. Lowery, L. Baeksgaard, T. Crosby, M. Cunningham, S. Cuffe, G.O. Griffiths, I. Parker, S.L. Risumlund, R. Roy, S. Falk, G.B. Hanna, F.R. Bartlett, A. Alvarez-Iglesias, M.P. Achiam, M. Nilsson, G. Piessen, N. Ravi, D. O'Toole, C. Johnston, R.S. McDermott, R.C. Turkington, S. Wahed, S. Sothi, H. Ford, M.S. Wadley, D. Power, S. Mukherjee, C. Morgan, S.L. Parsons, N. Bhuva, S. Campbell, L. Grogan, G. Leonard, A.R. Bateman, C. Mitchell, S. O'Reilly, E. Mulroe, O. McLoughlin, A. Shevlin, A.M. Shannon, J. Marron, M. Nolan, G. Burch, M. Costello, D. Griffiths, K. Cozens, E. Foley, C.L. Donohoe, C. O'Farrell, J. Moore, J. O'Sullivan, Trimodality therapy versus perioperative chemotherapy in the management of locally advanced adenocarcinoma of the oesophagus and oesophagogastric junction (Neo-AEGIS): an open-label, randomised, phase 3 trial, Lancet Gastroenterol Hepatol (2023), https://doi.org/10.1016/s2468-1253(23)00243-1.

[69] X. Zhao, Y. Ren, Y. Hu, N. Cui, X. Wang, Y. Cui, Neoadjuvant chemotherapy versus neoadjuvant chemoradiotherapy for cancer of the esophagus or the gastroesophageal junction: a meta-analysis based on clinical trials, PLoS One 13 (2018), https://doi.org/10.1371/journal.pone.0202185.

[70] J.A. Abrams, D.L. Buono, J. Strauss, R.B. McBride, D.L. Hershman, A.I. Neugut, Esophagectomy compared with chemoradiation for early stage esophageal cancer in the elderly, Cancer 115 (2009) 4924–4933, https://doi.org/10.1002/cncr.24536.

[71] K. Sugimura, H. Miyata, K. Tanaka, T. Makino, A. Takeno, O. Shiraishi, M. Motoori, M. Yamasaki, Y. Kimura, M. Hirao, K. Fujitani, T. Yasuda, M. Mori, H. Eguchi, M. Yano, Y. Doki, Multicenter randomized phase 2 trial comparing chemoradiotherapy and docetaxel plus 5-fluorouracil and cisplatin chemotherapy as initial induction therapy for subsequent conversion surgery in patients with clinical T4b esophageal cancer: short-term results, Ann. Surg. 274 (2021) E465–E472, https://doi.org/10.1097/SLA.0000000000004564.

[72] N.E. Donlon, B. Moran, A. Kamilli, M. Davern, A. Sheppard, S. King, C.L. Donohoe, M. Lowery, M. Cunningham, N. Ravi, C. Mueller, J. Cools-Lartigue, L. Ferri, J. V Reynolds, CROSS versus FLOT regimens in esophageal and esophagogastric junction adenocarcinoma: a propensity-matched comparison, Ann. Surg. 276 (2022) 792–798, https://doi.org/10.1097/SLA.0000000000005617.

[73] G. Zhang, C. Zhang, N. Sun, L. Xue, Z. Yang, L. Fang, Z. Zhang, Y. Luo, S. Gao, Q. Xue, J. Mu, Y. Gao, F. Tan, J. He, Neoadjuvant chemoradiotherapy versus neoadjuvant chemotherapy for the treatment of esophageal squamous cell carcinoma: a propensity score-matched study from the National Cancer Center in China, J. Cancer Res. Clin. Oncol. 148 (2022) 943–954, https://doi.org/10.1007/s00432-021-03659-7.

[74] J. Hoeppner, F. Lordick, T. Brunner, T. Glatz, P. Bronsert, N. Röthling, C. Schmoor, D. Lorenz, C. Ell, U.T. Hopt, J.R. Siewert, ESOPEC: prospective randomized controlled multicenter phase III trial comparing perioperative chemotherapy (FLOT protocol) to neoadjuvant chemoradiation (CROSS protocol) in patients with adenocarcinoma of the esophagus (NCT02509286), BMC Cancer 16 (2016) 503, https://doi.org/10.1186/s12885-016-2564-y.

[75] A.M. Bouvier, G. Launoy, C. Lepage, J. Faivre, Trends in the management and survival of digestive tract cancers among patients aged over 80 years, Aliment. Pharmacol. Ther. 22 (2005) 233–241, https://doi.org/10.1111/j.1365-2036.2005.02559.x.

[76] C. Shea, A.R. Khawaja, K. Sofi, G. Nabi, Association of metabolic equivalent of task (MET) score in length of stay in hospital following radical cystectomy with urinary diversion: a multi-institutional study, Int. Urol. Nephrol. 53 (2021) 1305–1310, https://doi.org/10.1007/s11255-021-02813-x.

[77] M.A. Brzezicki, N.E. Bridger, M.D. Kobetić, M. Ostrowski, W. Grabowski, S.S. Gill, S. Neumann, Artificial intelligence outperforms human students in conducting neurosurgical audits, Clin. Neurol. Neurosurg. 192 (2020), https://doi.org/10.1016/j.clineuro.2020.105732.

[78] H.S. Ólafsdóttir, E. Dalqvist, E. Onjukka, F. Klevebro, M. Nilsson, G. Gagliardi, G. Alexandersson von Döbeln, Postoperative complications after esophagectomy for cancer, neoadjuvant chemoradiotherapy compared to neoadjuvant chemotherapy: a single institutional cohort study, Clin Transl Radiat Oncol 40 (2023) 100610, https://doi.org/10.1016/j.ctro.2023.100610.

[79] J.F. Bosset, M. Gignoux, J.P. Triboulet, E. Tiret, G. Mantion, D. Elias, P. Lozach, J. C. Ollier, J.J. Pavy, M. Mercier, T. Sahmoud, Chemoradiotherapy followed by surgery compared with surgery alone in squamous-cell cancer of the esophagus, N. Engl. J. Med. 337 (1997) 161–167, https://doi.org/10.1056/NEJM199707173370304.

[80] T.N. Walsh, N. Noonan, D. Hollywood, A. Kelly, N. Keeling, T.P. Hennessy, A comparison of multimodal therapy and surgery for esophageal adenocarcinoma, N. Engl. J. Med. 335 (1996) 462–467, https://doi.org/10.1056/NEJM199608153350702.

[81] F. Fiorica, D. Di Bona, F. Schepis, A. Licata, L. Shahied, A. Venturi, A.M. Falchi, A. Craxì, C. Cammà, Preoperative chemoradiotherapy for oesophageal cancer: a systematic review and meta-analysis, Gut 53 (2004) 925–930, https://doi.org/10.1136/gut.2003.025080.

[82] S.K.R. Mukkamalla, A. Recio-Boiles, H.M. Babiker, Esophageal Cancer, StatPearls Publishing, 2023.

[83] J.V. Reynolds, S.R. Preston, B. O'Neill, M.A. Lowery, L. Baeksgaard, T. Crosby, M. Cunningham, S. Cuffe, G.O. Griffiths, R. Roy, S. Falk, G. Hanna, F.R. Bartlett, I. Parker, A. Alvarez-Iglesias, M. Nilsson, G. Piessen, S. Risum, N. Ravi, R. S. McDermott, Neo-AEGIS (neoadjuvant trial in adenocarcinoma of the esophagus and esophago-gastric junction international study): preliminary results of phase III RCT of CROSS versus perioperative chemotherapy (modified MAGIC or FLOT protocol). (NCT01726452), J. Clin. Oncol. 39 (2021) 4004, https://doi.org/10.1200/JCO.2021.39.15_suppl.4004.

[84] H.Y. Deng, W.P. Wang, Y.C. Wang, W.P. Hu, P.Z. Ni, Y.D. Lin, L.Q. Chen, Neoadjuvant chemoradiotherapy or chemotherapy? A comprehensive systematic review and meta-analysis of the options for neoadjuvant therapy for treating oesophageal cancer, Eur. J. Cardio. Thorac. Surg. 51 (2017) 421–431, https://doi.org/10.1093/ejcts/ezw315.

[85] M. Khosravi, Z. Zare, S.M. Mojtabaeian, R. Izadi, Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews, Health Serv Res Manag Epidemiol 11 (2024), https://doi.org/10.1177/23333928241234863.

[86] R. Haward, Z. Amir, C. Borrill, J. Dawson, J. Scully, M. West, R. Sainsbury, Breast cancer teams: the impact of constitution, new cancer workload, and methods of operation on their effectiveness, Br. J. Cancer 89 (2003) 15–22, https://doi.org/10.1038/sj.bjc.6601073.

[87] M.P. Achiam, M. Nordsmark, M. Ladekarl, A. Olsen, A. Loft, R.S. Garbyal, M. H. Larsen, A.P. Ainsworth, T.S. Kristensen, S. Dikinis, D.W. Kjær, L. Bæksgaard, M. Siemsen, M.B. Nielsen, S. Schlander, S. Kramer, N. Katballe, I. Kruhlikava, E. Tabaksblat, R.V. Fisker, P.B. Mortensen, E. Holtved, J. Eckardt, S. Detlefsen, G. Naujokaite, C.D. Lütken, Clinically decisive (dis)agreement in multidisciplinary team assessment of esophageal squamous cell carcinoma; a prospective, national, multicenter study, Acta Oncol. (Madr.) 60 (2021) 1091–1099, https://doi.org/10.1080/0284186X.2021.1937308.

[88] NHS England, NHS Improvement, Streamlining Multi-Disciplinary Team Meetings Guidance for Cancer Alliances, 2019.

[89] D. Ueda, T. Kakinuma, S. Fujita, K. Kamagata, Y. Fushimi, R. Ito, Y. Matsui, T. Nozaki, T. Nakaura, N. Fujima, F. Tatsugami, M. Yanagawa, K. Hirata, A. Yamada, T. Tsuboyama, M. Kawamura, T. Fujioka, S. Naganawa, Fairness of artificial intelligence in healthcare: review and recommendations, Jpn. J. Radiol. 42 (2024) 3–15, https://doi.org/10.1007/s11604-023-01474-3.

[90] M. Nagendran, P. Festor, M. Komorowski, A.C. Gordon, A.A. Faisal, Quantifying the impact of AI recommendations with explanations on prescription decision making, NPJ Digit Med 6 (2023), https://doi.org/10.1038/s41746-023-00955-z.

[91] M. Naiseh, D. Al-Thani, N. Jiang, R. Ali, How the different explanation classes impact trust calibration: the case of clinical decision support systems, Int. J. Hum. Comput. Stud. 169 (2023), https://doi.org/10.1016/j.ijhcs.2022.102941.

[92] R. Charow, T. Jeyakumar, S. Younus, E. Dolatabadi, M. Salhia, D. Al-Mouaswas, M. Anderson, S. Balakumar, M. Clare, A. Dhalla, C. Gillan, S. Haghzare, E. Jackson, N. Lalani, J. Mattson, W. Peteanu, T. Tripp, J. Waldorf, S. Williams, W. Tavares, D. Wiljer, Artificial intelligence education programs for health care professionals: scoping review, JMIR Med Educ 7 (2021), https://doi.org/10.2196/31043.

[93] M. Jirotka, B. Grimpe, B. Stahl, G. Eden, M. Hartswood, Responsible research and innovation in the digital age, Commun. ACM 60 (2017) 62–68, https://doi.org/10.1145/3064940.