Computers in Biology and Medicine 200 (2026) 111373

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Health care professionals' perceptions of machine learning based clinical decision support systems for oesophageal cancer management

Catherine Webb [a,1,*], Navamayooran Thavanesan [b,1], Mohammad Naiseh [c], Rachel Dewar-Haggart [d], Tim Underwood [b], Ganesh Vigneswaran [b]

[a] School of Primary Care, Population Sciences and Medical Education, University of Southampton, UK
[b] Innovation for Translation Research Group (ITRG), School of Cancer Sciences, Faculty of Medicine, University of Southampton, UK
[c] AFG College with the University of Aberdeen, Doha, Qatar
[d] Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

ARTICLE INFO

Keywords:
"Artificial intelligence"
"Machine learning"
"Clinical decision support system"
"Oesophageal cancer"
"Multidisciplinary team"

ABSTRACT

Oesophageal cancer (OC) causes significant morbidity and mortality. Multiple treatment regimens are available, and multidisciplinary team (MDT) decisions over which to offer are complex, multi-faceted and subject to logistical constraints and human factors. A machine learning (ML) model-based clinical decision support system (CDSS) for OC has been developed, trained on historical treatment decisions. However, clinician trust in such systems is not yet established.

This study surveyed clinicians in OC MDTs in the UK and Ireland to investigate which clinical and socio-demographic factors influence conscious decision-making in OC, comparing their relative subjective importance to those derived from the ML model (reflecting previous real-world practice). It also sought to explore clinicians' views on the potential use of artificial intelligence-based CDSSs in OC.

There was agreement between clinicians and the model in many of the most influential factors in decision making, although age and gender had greater influence on the model than their conscious importance to clinicians would support. Clinicians identified a wide range of additional clinical and holistic factors outside the current model which factor into their decision-making, including further investigations, symptoms, nutrition and social factors.

The prospect of utilising an ML CDSS in future received generally positive feedback, although opinions varied widely. However, barriers to implementation were identified, including concerns around perceived clinician superiority over ML CDSSs, patient individuality, transparency and safeguarding, the need for evidence, and additional input requirements. As ML CDSSs are increasingly offered in practice, clinicians' reservations must be addressed and their need for transparency and evidence met.

1. Introduction

Oesophageal cancer (OC) remains a key public health issue [1]. Treatment plans are determined by Upper Gastrointestinal (UGI) multidisciplinary teams (MDTs) which assimilate multi-domain expertise across a broad range of roles [2]. MDT treatment selection is directed by disease burden, patient demographics, functional status and co-morbidities [2,3] and is unsurprisingly a critical determinant of patient outcomes. However, a growing, ageing population has increased MDT caseload volume and complexity [4]. Case discussions may only last a few minutes per patient, suffer from incomplete information, and encounter challenging interpersonal dynamics [5,6].

The need to streamline and reform MDT processes is well recognised [4,5]. One potential solution uses computerised Clinical Decision Support Systems (CDSSs), now emerging across many aspects of cancer-care including screening, diagnosis and treatment planning [7]. CDSSs range from simple tools summarising clinical guidelines to complex systems integrating multiple data sources for patient-specific recommendations

* Corresponding author. South Academic Block, University Hospitals Southampton, Tremona Road, Southampton, SO16 6YD, UK.
E-mail address: Catherine.webb@soton.ac.uk (C. Webb).
@TimTheSurgeon (T. Underwood), @ganesh_vignes (G. Vigneswaran)

[1] These authors are joint First Author for this work.

https://doi.org/10.1016/j.compbiomed.2025.111373
Received 21 February 2025; Received in revised form 18 November 2025; Accepted 3 December 2025
Available online 10 December 2025

[7,8]. The ongoing Medical Artificial Intelligence (MAI) boom within healthcare is one such vehicle, representing a global market worth $5 billion USD in 2022 and projected to rise to $70 billion USD by 2032 [9]. Real-world MAI implementation remains in relative infancy [10,11]. Carefully implemented MAI may improve safety, efficiency, cost-effectiveness and unwarranted variation as well as geographical and sociodemographic inequalities [11–14].

However, without careful analysis of their inputs and processes, there is a risk of ineffective AI CDSSs perpetuating or worsening inequity [15]. Many AI solutions are also typically 'black box' (the machine's underlying logic is unclear to the human user) [16], which can be problematic within healthcare settings in establishing clinician and patient trust [16]. Consequently, they introduce novel questions and challenges ethically, legally, and in their acceptability to clinicians and patients [17].

Machine learning (ML) is a branch of AI increasingly utilised in CDSSs across many specialties [18,19,19,20]. ML uses computational power to identify patterns within large, complex datasets and make predictions. An explainable ML CDSS for OC patients has recently been under development at University Hospital Southampton (UHS), based on data from 893 OC patients discussed in MDTs between 2010 and 2022 [3,21] For the model to be integrated into clinical practice, it is crucial that the factors on which it bases decisions are consistent with standard of care practice and sound human clinical judgement. However, the relative importance clinicians attribute to many of the factors involved in these decisions is currently unknown. Furthermore, inequalities in treatment allocation have been noted by age, gender and ethnicity [22–25]. Identifying which factors clinicians value as explicitly important when compared to ML models trained on historical ground-truth decisions offers insight into future Human-AI interactions where inconsistencies between explicit priorities and observed practice can then be interrogated. Factors influential in the model which are not explicitly important to clinicians, may speak to potential implicit clinician bias, which is capable of significant impact on clinical practice [26].

Within this study, we present the results of a nationwide survey of OC MDT clinicians who were asked to describe how they rationalise the importance of key clinical variables when making treatment plans for OC patients. Their responses were compared against a random forests-based CDSS trained on historical MDT decisions at UHS, a high-volume tertiary referral unit [21], to identify areas of concordance and discordance between the Human and the AI. They were also surveyed on their current perceptions, concerns and views about the use of ML-based CDSSs to identify potential barriers towards implementation of such technologies within the OC space.

2. Methods

An anonymous survey with multiple choice and free text questions was hosted on the Qualtrics™ platform [27] from October–December 2023 (reproduced in supplementary material). Invitations were emailed to clinicians via professional membership organisations (The Association of Upper GI Surgeons of Great Britain and Ireland (AUGIS), the UK and Ireland Oesophagogastric Cancer Group (UKIOG) and the British Society of Gastroenterology (BSG)), and shared on social media by the researchers. Respondents were excluded if they did not self-report as regularly contributing to the UGI MDT in the UK or Ireland.

2.1. Questionnaire

Participants were asked to identify factors important in UGI MDT decision-making, selecting from 18 listed factors (Table 1) plus free-text 'other', and to rank their 10 most important factors. The list included sociodemographic factors plus variables used in the ML model, excluding 2 that were not applicable (timeframe of model training data and local geography). The ML model considers the 19 Charlson Comorbidity Index conditions separately [28], but these were combined in all but 1 question of the survey to reduce questionnaire burden.

Table 1
Factors listed as options in the survey.

Factor	In ML model?	Rationale for inclusion
Age	Y	Risk factor for frailty, comorbidities and complications
Gender	Y	Inequalities evident in literature [23]
Performance status	Y	Fitness to withstand treatment – grading system of maximum activity tolerated
ASA grade	Y	Fitness to withstand treatment – anaesthetist's grading system
Tumour stage	Y	Size and local spread of tumour
Nodal stage	Y	Spread to lymph nodes
Metastasis stage	Y	Spread to distant organs
Tumour location	Y	Influences feasibility of different treatments
Biopsy/tumour histology	Y	Type of cancer cells
Tumour differentiation	Y	Type of cancer cells – degree of abnormality
Comorbidities	Y	Fitness to withstand treatment and risk of complications
Smoking status	N	Risk factor for comorbidities/ complications
Alcohol usage	N	Risk factor for comorbidities/ complications
Patient geographical location	N	Access to various treatment modalities
Presence of local radiotherapy centre	N	Access to radiotherapy
Patient preference	N	Individual choice
Ethnicity	N	Inequalities evident in literature [24]
Socioeconomic status	N	Inequalities evident in literature [25]
Other (free text)	N	Establish additional relevant factors

To assess attitudes to the future use of ML CDSSs in OC, respondents were asked on a Likert scale how likely they would be to use one if available, and to describe any barriers.

2.2. Development and validation of the ML MDT model

The MDT ML model was developed using a random forests algorithm in Python ("Ranger" Library, sklearn v1.2.2, max depth = 6 (based on k = 5 cross-validation)) using variables consistently available to the MDT prior to a final treatment decision (Table 1). The model was trained on a cohort of 893 oesophageal cancer patients over a 12-year period managed at a tertiary referral centre. Previous publications described the model's development, established utility, and confirmed performance using area under the curve (AUC) from the multi-class Receiver Operator Characteristic (ROC) using "one" vs "others" approach (Fig. 1) [3,21]. Permutation-based feature importance was derived for each included variable. A slightly augmented version of the same model (953 cases) has been externally validated and discussed in detail in Thavanesan et al. [29].

2.3. Analysis

Data were managed in IBM SPSS Statistics for Macintosh v28.0.1.1 and Microsoft Excel v16.80. 'Top ten' responses were converted to numeric scores (top = ten points, second = nine etc).

Qualitative responses were analysed independently by 2 reviewers (CW and MN) through thematic analysis [30]. To ensure consistency and objectivity, an inter-rater reliability test was conducted to assess the level of agreement between different coders. The resulting score of 0.73 indicated good reliability, aligning with the commonly accepted threshold of 0.7 or above. For respondent validation, themes were discussed with clinicians in semi-structured interviews conducted as part of the wider research programme.

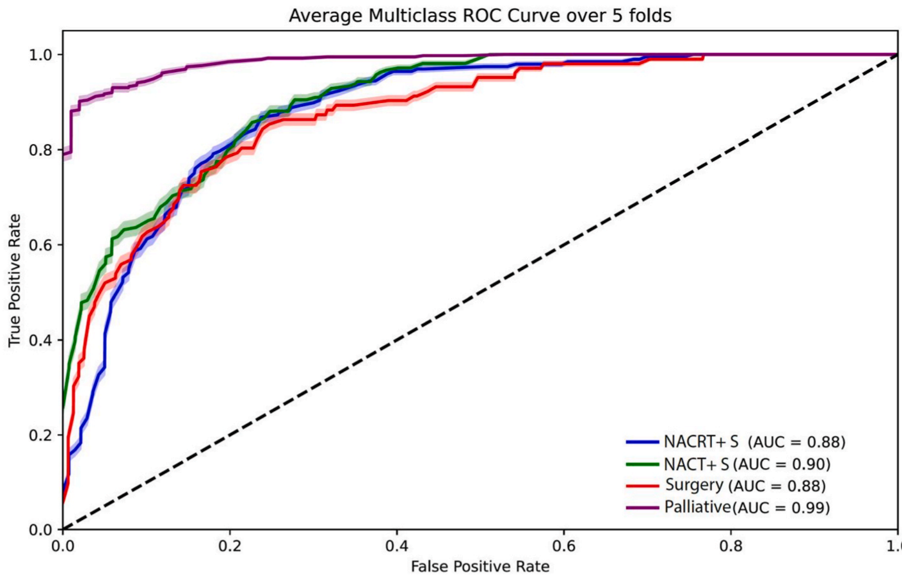

Fig. 1. Multiclass ROC curve for random forests treatment classifier representing a "one vs others" class-prediction performance. K = 5 Cross-validation was conducted using an 80:20 split. Mean ROC is presented ±1x Standard Error of the Mean. Reproduced from Ref. [21]. Computers in Biology and Medicine 180: 108978, p4.

Ethical approval

This work was approved by the University of Southampton Ethics Committee (ERGO: 70375) and the Health Research Authority Research Ethics Committee (IRAS: 319540) as part of the Machine Learning in Oesophageal Cancer (M-LOC) study.

3. Results

3.1. Sample size and demographics

A total of 87 participants accessed the survey, of which 67 consented. Three respondents were excluded as they do not routinely attend the UGI MDT, three as they work outside the UK, and one left the survey immediately after consenting. Table 2 shows the location and professional roles of the remaining 60 respondents. Response numbers vary by question as none were compulsory (except consent) and there was attrition throughout (Fig. 2).

3.2. Factor importance and discussion frequency

Most factors deemed important by respondents were also identified as routinely discussed (Fig. 3). A disparity was seen for patient preference, which was more often deemed important (88 %) than discussed (69 %).

3.3. Relative factor importance in overall treatment decisions

Fig. 4 shows the variable importance plot for the ML model. Table 3 compares the relative importance of each factor from the survey ranking to that of the ML model when determining overall treatment decisions.

Metastasis stage and performance status were top ranking for both the respondents and ML model. However, while age ranked third in the model (with high relative importance at 16 %) it ranked 9th in priority within the survey. Tumour stage and histology both ranked more highly for respondents than the model.

Gender was not ranked in the top ten by any survey respondent but has a 1.13 % variable importance in the ML model, higher than any individual comorbidity. Patient preference and ASA grade featured in the top ten survey rankings but are not featured within the current ML

Table 2
Work location, professional group and seniority of eligible respondents. Valid percentage excludes missing responses.

UK Geographical Area	Number of Respondents (n = 58)	Valid Percentage (%)
East Midlands (England)	3	5.2
West Midlands (England)	3	5.2
East of England	7	12.1
Kent, Surrey and Sussex	1	1.7
London	7	12.1
North East England	2	3.4
Yorkshire and The Humber	3	5.2
North West England	4	6.9
South West England	2	3.4
Thames Valley	3	5.2
Wessex	5	8.6
Scotland	7	12.1
Wales	4	6.9
Northern Ireland	5	8.6
Republic of Ireland	2	3.4
Missing	*2*	
Professional Role	**Number of Respondents (n = 59)**	**Valid Percentage (%)**
Surgeon	21	35.6
Gastroenterologist	7	11.9
Medical Oncologist	5	8.5
Clinical Oncologist	17	28.8
Radiologist	1	1.7
Specialist Nurse	6	10.2
Pathologist[a]	2	3.4
Missing	*1*	
Grade	**Number of Respondents (n = 54)**	**Valid Percentage (%)**
Consultant	43	79.6
Registrar	3	5.6
Fellow	2	3.7
Specialist Nurse	6	11.1
Valid Total	54	100.0
Missing	*6*	

[a] The pathologists completed the demographics section but exited the survey without providing further responses.

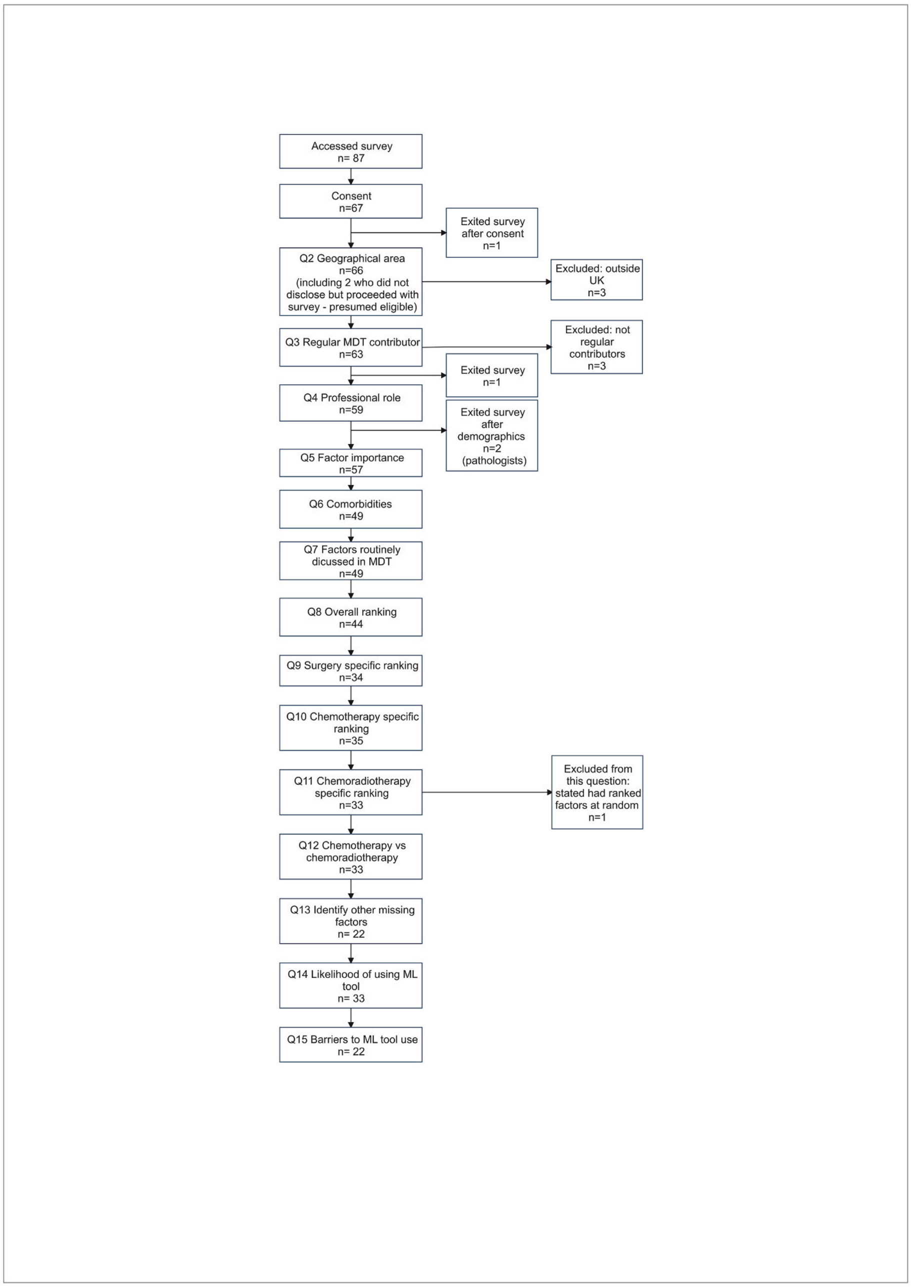

Fig. 2. Flow chart showing number of respondents per question and reasons for exclusions.

model - this data has not been introduced, and in the case of patient preference is frequently unavailable at first MDT discussion.

3.4. Comorbidities

Respondents were asked which co-morbidities they deemed to be important and were free to choose as many or few as they wished (Fig. 5). Congestive heart failure and dementia were considered important by more than 97 % of survey respondents but ranked relatively low within the preliminary ML model at 0.2 % and 0.04 % respectively. Amongst comorbidities, uncomplicated diabetes was considered the least important to respondents, but proved third most important in the model.

3.5. Additional factors recommended by respondents

Respondents were given the opportunity to highlight any additional factors they felt were important to their decision-making, using free-text (Table 4). Responses were organised into 4 categories through thematic analysis – additional investigation findings and assessments, symptoms, medical history, nutrition and social factors.

In addition, some respondents highlighted the impact of differing knowledgeability amongst MDT members, differing policies between units, the expertise of the surgical team in operating after chemoradiotherapy, and local research activity.

As neoadjuvant chemotherapy and chemoradiotherapy are currently both accepted treatments within the UK, respondents were also asked how they ordinarily chose between the two for a given patient. In addition to the general considerations previously outlined, several specific views emerged from across the range of professional roles (n = 33)

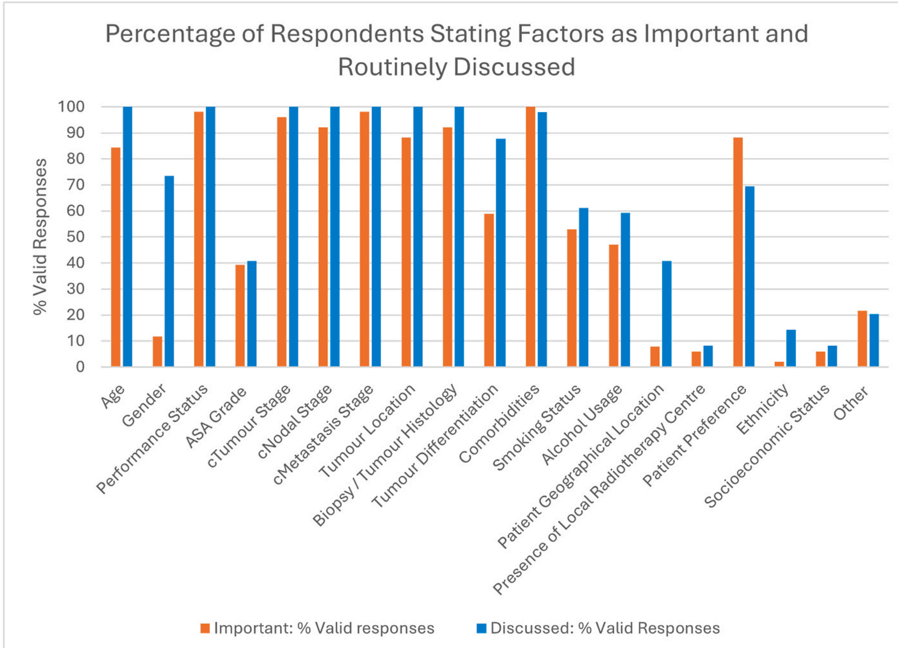

Fig. 3. Clustered bar chart showing the percentage of respondents deeming each factor important to them in OC treatment decisions (n = 57) and the percentage reporting that each factor is routinely discussed in MDTs (n = 49).

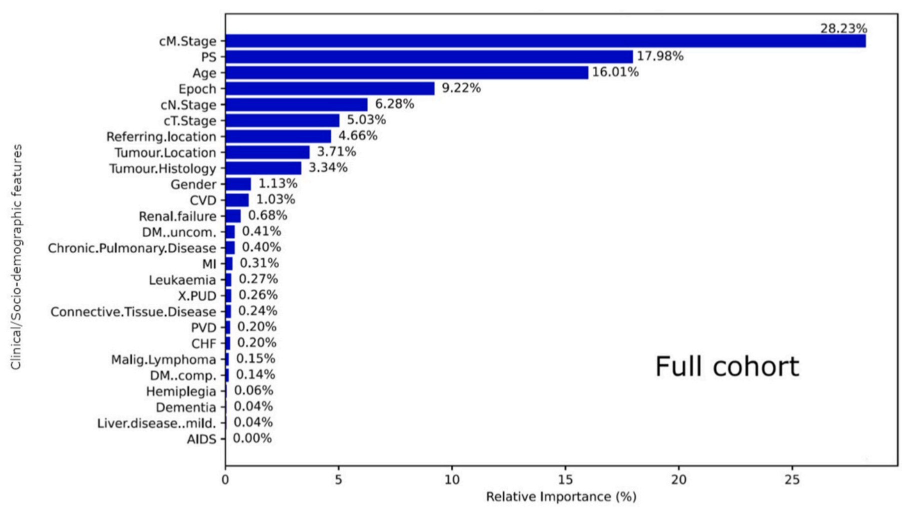

Fig. 4. Relative importance of various clinical and socioeconomic factors in determining patients' treatment pathways for oesophageal cancer, according to a machine learning model trained on data from historical MDT decisions. Adapted from Ref. [21]. Computers in Biology and Medicine 180: 108978, p5. cM.Stage = metastasis stage, PS = performance status, Epoch = date of input data (in relation to the publications of major clinical trials which altered practice in the field), cN. Stage = nodal stage, cT.Stage = tumour stage, CVD = cerebrovascular disease, DM.Uncom = diabetes without complications, MI = myocardial infarction, X.PUD = peptic ulcer disease, PVD = peripheral vascular disease, CHF = chronic heart failure, Malig.Lymphoma = malignant lymphoma, DM.comp = diabetes with complications.

(Table 5).

3.6. Clinician perceptions of a potential ML decision-support system for OC

When scoring their likelihood of using an ML CDSS from 0 (definitely not) to 100 (definitely yes), responses ranged from 0 to 100 (n = 33) with a median of 75.0 (IQR 60.0–82.5).

When asked about perceived barriers to ML CDSS use, five themes were identified (n = 22) relating to: Clinician Superiority, Patient Individuality, Transparency and Safeguarding, a Need for Evidence, and Input Requirements (Table 6).

A conceptual framework created from a literature search of clinician perceptions of AI in cancer (described in supplementary material), is combined with the findings of this study (Fig. 6).

4. Discussion

4.1. Summary

This study identified clinical and sociodemographic factors important in OC treatment decision-making, comparing their relative importance as identified by human experts against an ML model trained on historical MDT decisions from a leading tertiary referral unit. The study

Table 3
Comparison of overall ranking of factors from the survey with rankings from the ML model.

Rank	Survey	Comparison ML model
1	cMetastasis stage	cMetastasis stage
2	Performance status	Performance status
3	cTumour stage	Age
4	Biosy / tumour histology	Epoch *(not in survey)*
5	Comorbidities	cNodal stage
6	cNodal stage	cTumour stage
7	Patient preference *(not in ML model)*	Referring location *(not in survey)*
8	Tumour location	Comorbidities *(summed)*
9	Age	Tumour location
10	ASA grade *(not in ML model)*	Tumour histology

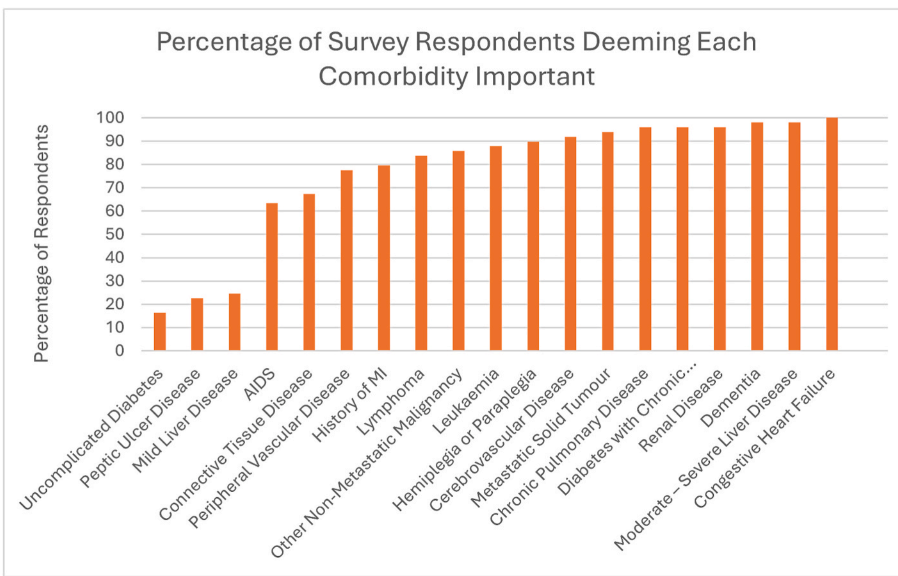

Fig. 5. Survey data - percentage of respondents who considered each comorbidity to be important when able to select as many as they wished (n = 49).

also explored clinicians' perceptions of using ML-driven CDSSs within OC.

Specific factors of importance to the human experts included additional investigation findings and assessments (e.g. molecular markers and frailty), symptoms and medical history (e.g. swallowing difficulties and medication use), nutrition (e.g. body mass index (BMI)) and social factors (e.g. support and coping). Discordance between human and machine were apparent, however, in other areas such as demographic factors, where age and gender were valued much higher within the model than they were consciously to clinicians. Additionally certain comorbidities ranking relatively low in importance in the ML model were weighted highly by participants (congestive heart failure and dementia), and vice versa (uncomplicated diabetes).

Clinicians' perceptions of ML CDSSs were favourable overall. However, several barriers to implementation were identified, grouped into five themes: perceptions of clinician superiority over ML CDSSs (e.g. human intuition), patient individuality (e.g. the need for person-centred care), transparency and safeguarding (e.g. explainability), the need for evidence (e.g. of effectiveness), and input requirements (e.g. molecular data).

4.2. Interpretation

This study has revealed that while ML can model OC decision-making with high performance, there remains some discordance between what humans consciously value within this process versus what the machine identifies when analysing historic decisions. Despite this, the overall sentiment towards the role of ML within MDT processes appears to be positive provided certain barriers can be overcome, such as the sense of clinician superiority, transparency, and granularity of the input variables.

4.3. Implications

Comorbidities such as congestive heart failure and dementia undoubtedly increase perioperative risk in surgical candidates [31,32].

Table 4
Additional factors identified through the survey as important in making OC treatment decisions. Underlined factors were identified by some respondents as also regularly discussed at their MDT. (n = 22).

Category	Factors
Additional investigation findings and assessments	Molecular markers (e.g. PD-L1, CPS, HER-2, MMR, MSI), DYPD genetic testing (determines how well certain chemotherapy agents are metabolised), tumour length, total length of disease (relevant to radiotherapy field), nodal distribution, stomach involvement, lung scan appearances, exercise testing, lung function test results, subjective impression of fitness ('end of bed' assessment), Anaesthetist's opinion, frailty
Symptoms and medical history	Dysphagia (swallowing difficulty), stridor (narrowed windpipe), vomiting (gastric outlet obstruction), disease-related quality of life, medication use, hearing impairment (risk of hearing loss as side-effect)
Nutrition	Weight/weight loss, body mass index, nutritional indicators
Social factors	Social, family and community support, employment, level of understanding and ability to report symptoms/side effects, likely compliance with treatment, psychological wellbeing and coping, religious beliefs (especially for Jehovah's Witnesses), profession and hobbies (risk of neuropathy (nerve damage) as side effect.

Table 5
Respondent views specific to the decision between chemotherapy and chemoradiotherapy. FLOT chemotherapy = 5-Flurouracil (5-FU), Leucovorin, Oxaloplatin, Docetaxel, CROSS Chemoradiotherapy regimen = Carboplatin and Paclitaxel with concurrent radiotherapy).

Consideration	Details From Some Respondents
Histology – Squamous Cell Carcinoma or Adenocarcinoma?	Many reported using chemoradiotherapy for all squamous cell carcinomas (if size and position are appropriate), and defaulting to chemotherapy for most or all adenocarcinomas
Tumour Location	Chemotherapy for Squamous Cell Carcinoma if Siewert grade 2/3 (a grading system describing tumour position) *(Respondent 35, Clinical Oncologist)*
Tumour length/total disease length	Many raised the importance of whether a tumour is small enough to be encompassed in a radiotherapy field, <8–10 cm.
Patient fitness for surgery	Some reported avoiding chemoradiotherapy if the patient is or may be fit for surgery
Is the circumferential resection margin threatened? I.e. is there a risk of incomplete removal with surgery?	May favour chemoradiotherapy for adenocarcinomas *(Respondent 48, Clinical Oncologist)*
Dysphagia (swallowing difficulties)	Chemotherapy may be more likely to improve this symptom *(Respondent 4, Clinical Oncologist)*
Dihdryopyrimidine dehydrogenase (DPD) testing result (a genetic blood test determining ability to metabolise certain chemotherapy agents)	May favour chemoradiotherapy as cannot use 5FU chemotherapy *(Respondent 4, Clinical Oncologist)*
Patient frailty	Frail patients may find chemoradiotherapy more tolerable than FLOT chemotherapy *(Respondent 11, Surgeon)*
Molecular test results	If favour post-operative immunotherapy, may select CROSS chemoradiotherapy regimen *(Respondent 11, Surgeon)*
Impact of potential side effects	Pre-existing hearing impairment, hobbies or profession *(Respondent 2, Medical Oncologist)*

Table 6
Thematic analysis of current barriers to adopting ML CDSSs in OC as highlighted by respondents.

Theme	Comments
Clinician Superiority	Some emphasised the importance of clinicians' intuition, 'gut feeling' and wealth of clinical experience, which they felt an algorithm could not emulate. Some considered clinicians better able to handle uncertainty when investigation results are inconclusive. *"Sometimes the true nodal stage cannot be accurately defined - are the nodes related to cancer or other lung pathology for instance. This uncertainty cannot be entered into an algorithm."* (Respondent 3, Surgeon) Several respondents emphasised that a CDSS should be limited to decision support and not over-ride clinical judgement.
Patient Individuality	Several respondents voiced concerns that ML CDSS use may hinder the individualised, holistic care that clinicians aim to provide, and questioned the application of population-level data to individual decisions. Some emphasised the multitude of factors involved, along with concerns that the model decision may not be valid if the patient's history, tumour or circumstances fell outside of the norm. *"I would be wary of using such a tool as I believe each patient is individual and unless you meet them and use a holistic approach it is hard to decide on the right treatment for them. People often differ from their 'on paper' selves. I do not believe AI can do this."* (Respondent 34, Specialist Nurse)
Transparency and Safeguarding	Respondents emphasised the importance of the model being open and explainable, so the reasoning behind its decision for each patient could be scrutinised to build trust. Some felt that patients should be informed of CDSS use with the right to decline. A number raised the need for a safeguarding process if clinicians disagree with the CDSS.
Need for evidence	Many respondents stated that they would need evidence of benefit before using an ML CDSS, including validity, cost effectiveness, patient benefit and time and resource savings. *"We discuss 40 patients in 90 min. It would have to be very quick and very useful!"* (Respondent 8, Surgeon) One respondent suggested that a CDSS may only save time if it were able to identify a group of patients who did not need to be discussed in the MDT.
Input requirements	Alongside the additional factors highlighted in Table 6, some respondents felt that a CDSS should include information of available clinical trial options. Some questioned how the CDSS would stay up to date, suggesting for example that models trained on data before newer diagnostic and treatment advancements may no longer be valid. *"Because this is an area of rapid flux, I would be concerned that ML models built using data derived before the routine use of molecular testing and the availability of immunotherapy may not be able to keep pace with changes in clinical management."* (Respondent 11, Surgeon)

Their importance to human experts is expected, and their comparative insignificance to the model may be explained by the low overall incidence of these conditions within the cohort (2.4 % for CHF and 1.1 % for dementia). Caution is therefore needed when considering CDSS recommendations based on infrequent features within the training data.

In addition to the assessment of individual and infrequent features, clinical expert decision-making leverages intuitive rapid analysis of interactions between features, which could not be assessed within the methodology of this clinician survey. However, insight into the interactions between such features is a key strength of tree-based models such as the random forests model on which this CDSS is based, and this is discussed in more depth by Thavanesan et al. [21].

The disparate ranking of age in this study was also a notable finding. UK guidelines for OC are not directed by age, as it is a poor predictor at an individual level in cancer treatment decision-making [33]. However, the possibility that clinicians treat it as a surrogate marker of overall fitness for major therapy has been raised previously [21]. It appears that

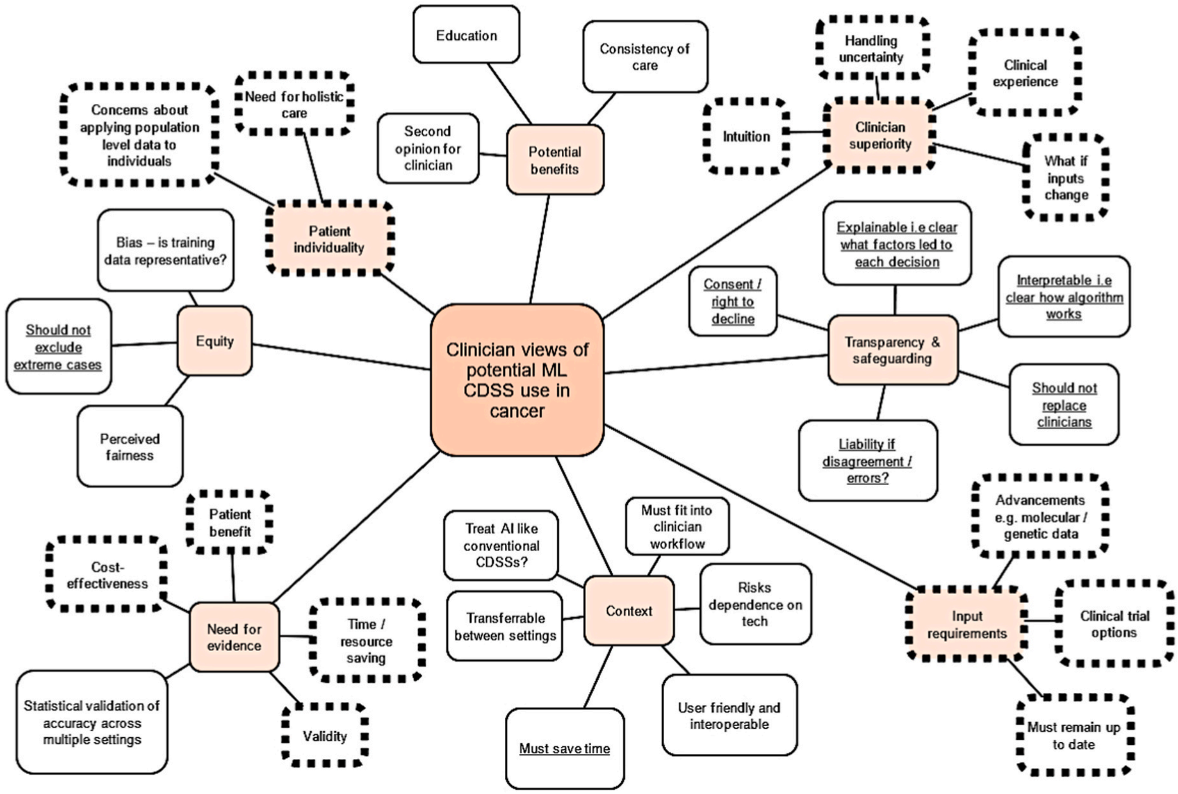

Fig. 6. Conceptual framework of clinician views of potential CDSS use in cancer, updated to include the findings of this study. Dotted borders indicate content added from this work. Underlining indicates content from the literature review which was reinforced by the results of this study.

age's disproportionate influence in practice may be sub-conscious, and independent of performance status in planning cancer treatment - a finding echoed elsewhere in the literature [25]. Gender, ethnicity, socioeconomic status and geographical location were only deemed important by a minority of clinicians. However previous studies have found inequitable treatments for women, non-caucasians and those of lower socioeconomic backgrounds [22–25]. ML CDSSs learning from historical decisions risk perpetuating inequalities, and the European Union Artificial Intelligence Act seeks to mitigate against this [34,35]. A key benefit of the use of ML within cancer care decision-making is not just the modelling and automation of treatment planning, but also in the potential of explainable AI (XAI) to interrogate the rationale behind team-based decisions. This offers the capability to audit decisions over time and elicit potential sources of bias. Within this cohort, our model weighed age as a significant factor, and we have shown previously that age influences OC treatment options [3,21]. There is some nuance to this phenomenon however, as we must balance the drive to minimise bias from ML models with the recognition that not all biases are indeed problematic and may even represent domain expertise borne form experiential learning. In these situations, it is recognised that some biases may be useful provided they do not propagate inequity [36]. In our use-case we know that advanced age carries additional clinical risk in aggressive medical interventions and so must be considered when treatment planning. Once such biases are recognised this may then allow future training datasets to be modified to up-scale representation in under-represented groups, or be fed-back to clinicians directly to modify human decision-making prior to training subsequent model generations.

Desirable factors outside of the current model identified by respondents included exercise testing, BMI, presenting symptoms, genetic testing and novel molecular markers. Additionally, future models may consider using multimodal machine learning techniques to incorporate a range of other input types [37–39]. Patient preference was identified as a key factor but could not be included within the current ML model as it is trained on MDT discussions made early in the patient pathway. An estimated 11 % of treatment decisions deviate from clinicians' recommendations due to patient preference [20], and some have argued that preferences should be more formally investigated and recorded for MDTs [40]. Shared decision-making between patient and clinician may benefit from using future CDSSs during consultations to support this process [20].

Participant perceptions of ML CDSSs in OC conveyed previously identified preferences for explainability, interpretability [10,41–44], a safeguarding process in the event of clinician disagreement with the CDSS [45,46], the need for speed [20], and the importance of reliability for extreme clinical cases [45]. Building trust between clinicians and AI tools is paramount if adoption of such tools is to be achieved and requires ethical, transparent innovation [47]. Predictive performance alone is insufficient to gain clinician trust; a model must be able to explain its reasoning [48]. In addition to quality control, data governance and bias mitigation, explainability (the ability to extract insight and understanding of machine logic when provided a given prediction or AI output) has become key for Medical AI in recent years [49,50]. To avoid biasing respondents in their survey responses the feature importance values were not provided within the survey. However, in the context of a validated ML model being used in clinical scenarios, the authors advocate for the insights from explainability tools to be made readily available for clinicians using medical CDSSs. This allows clinicians the opportunity to evaluate how well such rationales align with their own human decision making to develop trust in recommendations or to exercise clinician agency in disagreeing with the output. Within our modelling we utilised a form of global explainability technique (permutation-based feature importance) to provide insight. Global techniques provide a static over-arching overview of how the model uses and weighs inputs and this does not change from instance to instance. However, at the point of use, local-explainability tools such as Local Interpretable Model Agnostic Explanations (LIME [51] and Shappley

Values (SHAP) [52] are currently of great interest in providing insight at the individual patient level [50]. In combining these techniques and presenting them to clinicians as built-in aspects of user interfaces maximises transparency. This may then be combined with features such as counterfactual explanations (explanations derived from changing maximally one or two model inputs for a given patient and comparing how the resultant output varies from the original scenario), clinician override (the ability to override an AI recommendation by the human operator and record those specific cases for future training) and feedback systems for clinicians to highlight areas of discrepancy to incorporate into subsequent model iterations. Finally, traceability of model development also remains a critical aspect in building and maintaining trust and accountability. It requires the recording of successive model life cycles, their input data, pre-processing procedures and outputs [53].

Areas of hesitancy for respondents surrounded factoring in patient individuality, perceived clinician superiority, and novel clinical predictors. In a survey of doctors and medical students in Korea, only 44 % believed that AI was diagnostically superior to doctors [54]. Research has also highlighted the limitations of AI CDSSs in handling uncertainty [55], a sentiment strongly echoed by participants within this study. With regards to patient individuality, it remains controversial whether medical AI technologies inhibit or enhance person-centred care. Some fear MAI may fail to incorporate patients' values and preferences, while proponents argue for its potential to release clinician time to build patient trust and aid counselling for shared decision-making [56]. This demonstrates that in the short term, such tools will play a primarily decision-support role rather than an automated decision-maker role.

The results of this study have highlighted that while MAI represents a growth market globally, offering potential gains across several clinical performance indicators, successful implementation requires buy-in by end-users through overcoming the barriers highlighted here. It has also demonstrated discordance between human perception and machine learned variables for OC decision making, suggesting that what the human agents within MDTs perceive they value may not necessarily match subsequent decision outcomes. Human oversight remains key to mitigating MAI bias [15], and research such as this study, is necessary to achieve long-term clinician buy-in, as well as safe and equitable implementation.

4.4. Study limitations and strengths

The study was undoubtedly subject to limitations, the main one being a low response rate (approximately 6 %–8 % based on society mailing list sizes) - a common challenge with clinician surveys [57], as well as response-attrition over the course of the survey. Only the most motivated respondents continued onto the latter stages of the questionnaire exploring perceptions of ML CDSSs, potentially exacerbating response bias as questions exploring perceptions of ML CDSSs were towards the end of the survey. The overall sentiment towards ML found in this study may thus be in part influenced by self-selection of respondents holding stronger views on the subject; however, a full range of positive and negative perceptions were reported by respondents.

Exploring nuanced decision-making within an online survey naturally risks over-simplifying a complex process where factors may not carry equal weight across their range and may even act synergistically with others. Nevertheless this study leveraged a mixed methodology approach to bridge knowledge between ML modelling of MDT decision-making with quantitative and qualitative data on respondent's priorities [58] allowing direct comparison between real world practice as viewed through an AI paradigm versus clinicians' own conscious decision-making. The anonymous, online nature of the survey mitigates some of the risk of social acceptability response bias, particularly compared to co-design approaches in which interviewers may be affiliated to the CDSS. This survey also achieved strong geographical coverage across the UK, spanning all relevant professional groups within the MDT and representing a full range of opinions on using ML CDSSs.

The insights derived from the study will be key in guiding future CDSS design and ML modelling approaches if designers wish to maximise the adoption of these CDSSs into widespread clinical practice. The need for a "co-design" ethos within medical AI which factors in stakeholder needs and concerns is increasingly important if the barriers identified within this study are to be overcome.

5. Conclusion

This survey revealed insights into clinicians' priorities in OC decision-making and highlighted the range and complexity of factors influencing treatment recommendations. Sociodemographic factors appear to have greater impact in practice than would be anticipated by their conscious importance to clinicians. A wide range of potential barriers to the future use of an ML CDSS in OC were identified, many of which aligned with the findings of work on other cancers and throughout medicine. These must be addressed if the potential gains of clinical AI are to be realised.

CRediT authorship contribution statement

Catherine Webb: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Navamayooran Thavanesan:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohammad Naiseh:** Writing – review & editing, Visualization, Methodology, Formal analysis, Data curation. **Rachel Dewar-Haggart:** Writing – review & editing, Visualization, Methodology, Formal analysis. **Tim Underwood:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Ganesh Vigneswaran:** Writing – review & editing, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Ethical approval

This work was approved by the University of Southampton Ethics Committee (ERGO: 70375) and the Health Research Authority Research Ethics Committee (IRAS: 319540) as part of the Machine Learning in Oesophageal Cancer (M-LOC) study.

Funding support acknowledgement

NT receives a joint studentship from the Institute for Life Sciences (University of Southampton) and University Hospital Southampton. The project receives additional funding from the UKRI Trustworthy Autonomous Systems Hub (TAS Hub) Pump Priming Fund. The funding sources were not involved in study design, data collection, analysis, interpretation of data or writing of this manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.111373.

Appendix 1. Literature search details in development of conceptual framework in Fig. 5, performed on March 6, 2024

"Artificial Intelligence" OR "AI" OR "Machine Learning" OR "ML" OR "neural net*" OR "deep learning" OR "expert system".

AND.

Clinician OR doctor OR physician OR nurse OR "health* professional" OR "health* staff" OR "medic* professional" OR "medic* staff".

AND.

Views OR perceptions OR thoughts OR beliefs OR opinions OR qualitative OR interviews OR focus groups.

AND.

"Decision support" OR "decision aid" or "decision assist*"

(MESH: Decision Support Systems, Clinical/or Decision Support Techniques/)

AND.

Cancer OR tumour OR tumor OR neoplas*

Inclusion: all study designs, all applications of clinical decision support system (e,g. screening, diagnosis, treatment).

Exclusion: studies not specific to cancer (e.g. nephrology in general, not only kidney cancer).

Data availability

Anonymised data included within this study may be shared on request.

References

[1] J. Huang, A. Koulaouzidis, W. Marlicz, V. Lok, C. Chu, C.H. Ngai, L. Zhang, P. Chen, S. Wang, J. Yuan, X.Q. Lao, S.L.A. Tse, W. Xu, Z.J. Zheng, S.H. Xie, M.C.S. Wong, Global burden, risk factors, and trends of esophageal cancer: an analysis of cancer registries from 48 countries, Cancers (Basel, Switzerland) 13 (2021).
[2] NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE, Oesophago-gastric cancer: assessment and management in adults [Online]. Available: https://www.nice.org.uk/guidance/ng83/resources/oesophagogastric-cancer-assessment-and-management-in-adults-pdf-1837693014469, 2018. Accessed 23/June/23.
[3] 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. 49 (2023) 106986.
[4] INDEPENDENT CANCER TASKFORCE, Achieving world-class Cancer Outcomes: a Strategy for England 2015-2020, Independent Cancer Taskforce, London, 2015 [Online], http://bit.ly/1Idwf5W. Accessed 19/04/24.
[5] CANCER RESEARCH UK, Meeting patients' needs: improving the effectiveness of multidisciplinary team meetings in cancer services [Online]. Available: https://www.cancerresearchuk.org/about-us/we-develop-policy/our-policy-on-cancer-services/improving-the-effectiveness-of-mdts-in-cancer-services, 2017. Accessed 01/04/24.
[6] B.W. Lamb, K.F. Brown, K. Nagpal, C. Vincent, J. Green, N. Sevdalis, Quality of care management decisions by multidisciplinary cancer teams: a systematic review, Ann. Surg Oncol. 18 (2011) 2116–2125.
[7] M. Beauchemin, M.T. Murray, L. Sung, D.L. Hershman, C. Weng, R. Schnall, Clinical decision support for therapeutic decision-making in cancer: a systematic review, Int. J. Med. Inf. 130 (2019) 103940.
[8] NHS ENGLAND, Supporting clinical decisions with health information technology: an implementation guide for clinical decision support systems [Online]. Available: https://www.england.nhs.uk/long-read/supporting-clinical-decisions-with-health-information-technology/, 2023. Accessed 23/04/24.
[9] GLOBAL MARKET INSIGHTS, Artificial intelligence in healthcare market - by application (medical imaging & diagnosis, drug discovery, therapy planning, hospital workflow, wearables, virtual assistants), by region & global forecast, 2023 - 2032 [Online]. Available: https://www.gminsights.com/industry-analysis/healthcare-artificial-intelligence-market, 2023. Accessed 02/04/24.
[10] S. Emani, A. Rui, H.A.L. Rocha, R.F. Rizvi, S.F. Juaçaba, G.P. Jackson, D.W. Bates, Physicians' perceptions of and satisfaction with artificial intelligence in cancer treatment: a clinical decision support system experience and implications for low-middle-income countries, JMIR Cancer 8 (2022).
[11] Choudhury, Asan, Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature Review *JMIR Medical Informatics*, 8, 2020 18599.
[12] N.N. Khanna, M.A. Maindarkar, V. Viswanathan, J.F.E. Fernandes, S. Paul, M. Bhagawati, P. Ahluwalia, Z. Ruzsa, A. Sharma, R. Kolluri, I.M. Singh, J.R. Laird, M. Fatemi, A. Alizad, L. Saba, V. Agarwal, A. Sharma, J.S. Teji, M. AL-Maini, V. Rathore, S. Naidu, K. Liblik, A.M. Johri, M. Turk, L. Mohanty, D.W. Sobel, M. Miner, K. Viskovic, G. Tsoulfas, A.D. Protogerou, G.D. Kitas, M.M. Fouda, S. Chaturvedi, M.K. Kalra, J.S. Suri, Economics of artificial intelligence in healthcare: diagnosis vs. treatment, Healthcare (Basel, Switzerland) 10 (2022).
[13] N. Thavanesan, G. Vigneswaran, I. Bodala, T.J. Underwood, The oesophageal cancer multidisciplinary team: can machine learning assist decision-making? J. Gastrointest. Surg. 27 (2023) 807–822.
[14] C. Brown, R. Nazeer, A. Gibbs, P. Le Page, A.R. Mitchell, Breaking bias: the role of artificial intelligence in improving clinical decision-making, Cureus 15 (2023) e36415.
[15] M. Mittermaier, M.M. Raza, J.C. Kvedar, Bias in AI-based models for medical applications: challenges and mitigation strategies, npj Digit. Med. 6 (2023) 113.
[16] E.H. Shortliffe, M.J. Sepúlveda, Clinical decision support in the era of artificial intelligence, J. Am. Med. Assoc. 320 (2018) 2199–2200.
[17] J. Amann, A. Blasimme, E. Vayena, D. Frey, V.I. Madai, Explainability for artificial intelligence in healthcare: a multidisciplinary perspective, BMC Med. Inf. Decis. Making 20 (2020) 310.
[18] S. Sanchez-Martinez, O. Camara, G. Piella, M. Cikes, M.Á. González-Ballester, M. Miron, A. Vellido, E. Gómez, A.G. Fraser, B. Bijnens, Machine learning for clinical decision-making: challenges and opportunities in cardiovascular imaging, Front. Cardiovasc. Med. 8 (2022).
[19] L. Adlung, Y. Cohen, U. Mor, E. Elinav, Machine learning in clinical decision making, Med 2 (2021) 642–665.
[20] A. Ankolekar, B. van der Heijden, A. Dekker, C. Roumen, D. de Ruysscher, B. Reymen, A. Berlanga, C. Oberije, R. Fijten, Clinician perspectives on clinical decision support systems in lung cancer: implications for shared decision-making, Health Expect. 25 (2022) 1342–1351.
[21] N. Thavanesan, A. Farahi, C. Parfitt, Z. Belkhair, T. Azim, I. Bodala, Z. Walters, S. Ramchurn, T.&G.V. Underwood, Insights from explainable AI in oesophageal cancer team decisions, Comput. Biol. Med. 180 (2024).
[22] M. Kalff, W. Dijksterhuis, A. Wagner, S. Oertelt-Prigione, R. Verhoeven, V. Lemmens, H. van Laarhoven, S. Gisbertz, M. 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.
[23] W. Dijksterhuis, M. Kalff, A. Wagner, R. Verhoeven, V. Lemmens, M. van Oijen, S. Gisbertz, M. van Berge Henegouwen, H. van Laarhoven, Gender differences in treatment allocation and survival of advanced gastroesophageal cancer: a population-based study, J. Natl. Cancer Inst. 113 (2021) 1551–1560.
[24] 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).
[25] 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.
[26] D.P. Gopal, U. Chetty, P. O'Donnell, C. Gajria, J. Blackadder-Weinstein, Implicit bias in healthcare: clinical practice, research and decision making, Future Healthcare J. 8 (2021) 40–48.
[27] QUALTRICS, Qualtrics Homepage, 2024 [Online]. [Accessed 23/10/24].
[28] M.E. Charlson, P. Pompei, K.L. Ales, C.R. Mackenzie, A new method of classifying prognostic comorbidity in longitudinal studies: development and validation, J. Chron. Dis. 40 (1987) 373–383.
[29] N. Thavanesan, M. Naiseh, M. Terol, S. Rahman, S. Hill, C. Parfitt, Z. Walters, S. Ramchurn, S. Markar, R. Owen, N. Maynard, T. Azim, Z. Belkatir, E. Perez, M. Mccord, T. Underwood, G. Vigneswaran, Oesophageal cancer multi-disciplinary tool: a co-designed, externally validated, machine learning tool for oesophageal cancer decision making, eClinicalMedicine 89 (2025) 103527.
[30] V. Braun, V. Clarke, Using thematic analysis in psychology, Qual. Res. Psychol. 3 (2006) 77–101.
[31] S. Alcorn, I. Foo, Perioperative management of patients with dementia, BJA Educat. 17 (2017) 94–98.
[32] A. Kotzé, S.J. Howell, Heart failure: pathophysiology, risk assessment, community management and anaesthesia, Cont. Educ. Anaesth. Crit. Care Pain 8 (2008) 161–166.
[33] E. Soto-Perez-de-Celis, D. Li, Y. Yuan, Y.M. Lau, A. Hurria, Functional versus chronological age: geriatric assessments to guide decision making in older patients with cancer, Lancet Oncol. 19 (2018) e305–e316.
[34] H. van Kolfschooten, The AI cycle of health inequity and digital ageism: mitigating biases through the EU regulatory framework on medical devices, J. Law Biosci. 10 (2023) lsad031.
[35] THE EUROPEAN UNION, Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689, 2024. Accessed 20/08/24.
[36] M. Pot, N. Kieusseyan, B. Prainsack, Not all biases are bad: equitable and inequitable biases in machine learning and radiology, Pot M, Kieusseyan N, Prainsack B, Insights Imag. 10 (1) (2021) 13, 12.
[37] G. Zhao, Y. Lin, Y. Lu, Z. Chen, W. Guo, Lightweight bilateral network of Mura detection on micro-OLED displays, Measurement 255 (2025) 117937.
[38] X. Yu, X. Liang, Z. Zhou, B. Zhang, H. Xue, Deep soft threshold feature separation network for infrared handprint identity recognition and time estimation, Infrared Phys. Technol. 138 (2024) 105233.
[39] X. Yu, X. Liang, Z. Zhou, B. Zhang, Multi-task learning for hand heat trace time estimation and identity recognition, Expert Syst. Appl. 255A (2024) 124551.
[40] T. Soukup, N. Sevdalis, J. Green, B.W. Lamb, C. Chapman, T.A. Skolarus, Making tumor boards more patient-centered: let's start with the name, JCO Oncol. Pract. 17 (2021) 591–593.

[41] S.G. Anjara, A. Janik, A. Dunford-Stenger, K. Mc Kenzie, A. Collazo-Lorduy, M. Torrente, L. Costabello, M. Provencio, Examining explainable clinical decision support systems with think aloud protocols, PLoS One 18 (2023).

[42] T. Brigden, C. Mitchell, E.R. Hill, A. Hall, Ethical and legal implications of implementing risk algorithms for early detection and screening for oesophageal cancer, now and in the future, PLoS One 18 (2023).

[43] L. Pumplun, F. Peters, J.F. Gawlitza, P. Buxmann, Bringing machine learning systems into clinical practice: a design science approach to explainable machine learning-based clinical decision support systems, J. Assoc. Inf. Syst. Online 24 (2023) 953–979.

[44] C.J. Staes, A.C. Beck, G. Chalkidis, C.H. Scheese, T. Taft, J.W. Guo, M.G. Newman, K. Kawamoto, E.A. Sloss, J.P. Mcpherson, Design of an interface to communicate artificial intelligence-based prognosis for patients with advanced solid tumors: a user-centered approach, J. Am. Med. Inf. Assoc. 31 (2023) 174–187.

[45] L. Koco, C.C.N. Siebers, M. Schlooz, C. Meeuwis, H.S.A. Oldenburg, M. Prokop, R. M. Mann, S. Ailawadhi, The facilitators and barriers of the implementation of a clinical decision support system for breast cancer multidisciplinary team meetings - an interview study, Cancers 16 (2024).

[46] J. Helenason, C. Ekström, M. Falk, P. Papachristou, Exploring the feasibility of an artificial intelligence based clinical decision support system for cutaneous melanoma detection in primary care - a mixed method study, Scand. J. Prim. Health Care 42 (2024) 51–60.

[47] 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. 21 (1) (2023) e100920, 30.

[48] M. Mediouni, V. Marenkov, A.B. Diallo, Towards an interpretable machine learning model for predicting antimicrobial resistance, J. Glob. Antimicrob. Resist. 45 (2025) 47–51, 2025 Aug 25.

[49] J. Zhang, Z.M. Zhang, Ethics and governance of trustworthy medical artificial intelligence, BMC Med. Inf. Decis. Making 13 (1) (2023) 7, 23.

[50] M. Frasca, D. la Torre, G. Pravettoni, I. Cutica, Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review, Discov. Artif. Intell. 4 (2024) 15.

[51] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you?. Explaining the Predictions of Any Classifier ArXiv ID 1602.04938, 2016.

[52] S. Lundberg, S. Lee, A Unified Approach to Interpreting Model Predictions, 2017. ArXiv ID 1705.07874.

[53] K. Lekadir, A.F. Frangi, A.R. Porras, B. Glocker, C. Cintas, C.P. Langlotz, et al., FUTURE-AI: International Consensus Guideline for Trustworthy and Deployable Artificial Intelligence in Healthcare *BMJ*, 2025 e081554, 2025; 388.

[54] S. Oh, J.H. Kim, S.W. Choi, H.J. Lee, J. Hong, S.H. Kwon, Physician confidence in artificial intelligence: an online mobile survey, J. Med. Internet Res. 21 (2019) e12422.

[55] K.J. Pfisterer, R. Lohani, E. Janes, D. Ng, D. Wang, D. Bryant-Lukosius, R. Rendon, A. Berlin, J. Bender, I. Brown, A. Feifer, G. Gotto, S. Saha, J.A. Cafazzo, Q. Pham, An actionable expert-system algorithm to support nurse-led cancer survivorship care: algorithm development study, JMIR Cancer 9 (2023) e44332.

[56] A. Sauerbrei, A. Kerasidou, F. Lucivero, N. Hallowell, The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions, BMC Med. Inf. Decis. Making 23 (2023) 73.

[57] W.E. Shiyab, C. Ferguson, K. Rolls, E. Halcomb, Solutions to address low response rates in online surveys, Eur. J. Cardiovasc. Nurs. 22 (2023) 441–444.

[58] P. Bazeley, Analysing mixed methods data. Mixed Methods Research for Nursing and the Health Sciences, 2009.