Mapping the Advanced-Stage Epithelial Ovarian Cancer Landscape Goes Beyond Words: Two Large Language Models, Eight Tasks, One Journey

Michela Quaranta , Alexandros Laios * , Charlie Rogers , Anastasia Mavromatidou , Amudha Thangavelu , Georgios Theophilou , David Nugent , Diederick DeJong , Evangelos Kalampokis

Posted Date: 14 February 2025

doi: 10.20944/preprints202502.1096.v1

Keywords: epithelial ovarian cancer; natural language processing; transfer learning; operative notes; RoBERTa; GatorTron

 doi:10.20944/preprints202502.1096.v1

Article

Mapping the Advanced-Stage Epithelial Ovarian Cancer Landscape Goes Beyond Words: Two Large Language Models, Eight Tasks, One Journey

Michela Quaranta [1,†], Alexandros Laios [1,*,†], Charlie Rogers [1], Anastasia Mavromatidou [2], Amudha Thangavelu [1], Georgios Theophilou [1], David Nugent [1], Diederick DeJong [1] and Evangelos Kalampokis [2]

[1] Department of Gynaecologic Oncology, ESGO Centre of Excellence for Ovarian Cancer Surgery, St James's University Hospital, Leeds, United Kingdom
[2] Information Systems Lab, Department of Business Administration, University of Macedonia, Thessaloniki, Greece
* Correspondence: a.laios@nhs.net
† These authors contributed equally to this work.

Abstract: The advancement of natural language processing (NLP) technologies has transformed various sectors. However, their application in the healthcare domain, particularly for analysing clinical notes, remains underdeveloped. We investigated the use of deep neural networks, specifically transformer-based models, to predict intraoperative and postoperative outcomes related to advanced-stage epithelial ovarian cancer cytoreduction (aEOC) using unstructured surgical notes. We evaluated the performance of RoBERTa, a general-purpose language model and GatorTron, a domain-specific model, across eight binary classification tasks using the same dataset. The dataset consisted of 560 surgical records from patients with aEOC who underwent cytoreductive surgery at a tertiary UK reference centre. Predictive outcomes were converted into binary features to facilitate classification tasks. To enhance the contextual information available to the models, textual data from "operative findings" and "operative notes" were concatenated. Our findings highlight the tangible benefits of employing domain-specific language models for clinical text analysis. GatorTron generally outperformed RoBERTa across most predictive tasks, underscoring the advantages of domain-specific pretraining for understanding medical terminology and context. Both models struggled to predict certain outcomes, particularly those involving postoperative events like major complications and length of hospital stay, despite adjustments in hyperparameters and training strategies. This limitation suggests that operative text alone may not sufficiently capture the complexities of postoperative recovery. These findings have valuable implications for developing medical AI systems to improve the delivery of modern aEOC healthcare.

Keywords: epithelial ovarian cancer; natural language processing; transfer learning; operative notes; RoBERTa; GatorTron

1. Introduction

Advanced stage epithelial ovarian cancer (aEOC) remains clinically challenging as a leading cause of gynecological cancer mortality [1]. Patients often present with nonspecific symptoms, and the disease is frequently diagnosed at a late stage. Optimal surgical cytoreduction combined with platinum-based chemotherapy is the cornerstone of therapy, with the extent of tumour debulking being a strong predictor of overall survival [2]. However, major cytoreductive procedures carry considerable morbidity, making risk stratification and perioperative management paramount for patient counselling and improved outcomes [3].

The Covid-19 pandemic was a catalyst for a rapid digital transformation in healthcare, reflected in the widespread adoption of electronic health records (EHRs) [4]. Despite the wealth of clinical information in EHRs, up to 80% of these data stored in unstructured text, such as clinical notes, operative

 doi:10.20944/preprints202502.1096.v1

findings, and discharge summaries, are not used [5]. Although traditional statistical approaches often overlook these unstructured data, leveraging natural language processing (NLP) could offer additional insights from free-text clinical documents [6].

Numerous language models have been applied in diverse healthcare contexts, but their performance varies depending on factors such as domain-specific training data and complexity of clinical tasks. GatorTron, a domain-specific language model designed for clinical text, has shown promise in several healthcare-related tasks [7]. RoBERTa, built on the Bidirectional Encoder Representations from Transformers (BERT) architecture, has also demonstrated strong performance in general text classification, including biomedical applications [8]. We previously demonstrated that RoBERTa-based classifiers, when employed to process unstructured operative notes outperformed traditional risk factor-based models in the prediction of surgical outcomes [9].

Yet, few direct comparisons between these two models exist, particularly within gynaecological oncology and specifically in the context of advanced EOC cytoreductive surgery [10]. This gap could be bridged by combining the NLP work with domain-specific knowledge and deep insight into clinical workflows. To address this, we conducted a head-to-head comparison of GatorTron and RoBERTa for eight binary classification tasks associated with perioperative outcomes and postoperative morbidity in aEOC. By integrating text from operative notes with additional operative findings, we hypothesised that our selected NLP-based neural network models could identify signals predictive of key risk factors, including the likelihood of complete cytoreduction, length of hospital stay, estimated blood loss, operative time, need for intensive treatment unit admission, and postoperative complications. Ultimately, our goal is to demonstrate the potential of modern NLP techniques to improve perioperative risk management and standardize the reporting of key surgical outcomes after advanced cytoreduction of EOC.

2. Materials and Methods

2.1. Study Design and Population

We reviewed the hospital EHRs to identify women with aEOC (FIGO stage III or IV) who underwent cytoreductive surgery at a tertiary UK referral centre between January 2014 and December 2019. The EHR dataset consisted of 560 surgical notes detailing surgical findings. An aEOC clinical database, developed internally, was integrated with the EHR system to provide access to both discrete and engineered data [11]. Institutional research ethics board approval was granted by the LTHT (MO20/133, 163/18.06.20), and informed written consent was obtained from all participants. The study was registered with the UMIN Clinical Trials Registry (UMIN000049480). Exclusion criteria included missing operative notes, lack of histopathological aEOC confirmation, and incomplete postoperative follow-up data.

2.2. Data Extraction and Processing

A customised pipeline was developed to extract and pre-process data from the institutional EHR system. Structured fields—such as patient demographics, date of surgery, date of discharge, and standard laboratory values—were collected using query-based methods. The dataset originally contained 113 fields; however, this was reduced to 10 fields to focus on the unstructured text and most relevant data. These included two text fields (used as model inputs) and eight predictive outcomes (used as output variables) (Table 1).

 doi:10.20944/preprints202502.1096.v1

Table 1. Selected dataset fields.

Text Fields	Patient Characteristics	Original Data Type
Operative Notes (Op Note)	Time procedure (minutes)	Binary
Operative Findings (Op Findings)	Length of stay (days)	Binary
	Time between surgery and end of treatment	Binary
	Intensive Care Admission (after surgery)	Integer
	Estimated Blood Loss (EBL)	Integer
	End of Treatment CA125	Integer
	Complete cytoreduction vs non-complete cytoreduction	Integer
	Major postoperative complications Clavien Dindo (CD) (3–5)	Integer

The unstructured text primarily consisted of:

Operative Notes: detailed narratives documented by the surgeon describing the operative approach, location and extent of disease, specific manoeuvres undertaken at cytoreduction, and immediate surgical events and/or instructions.

Operative Findings: additional text-based reports outlining tumour distribution, and involvement of various anatomical structures.

These narratives were concatenated into a unified corpus for each patient to allow for NLP analysis and improve model performance. All clinical features were transformed into binary to ensure a uniform comparison. For continuous variables (e.g., length of stay, procedure time, blood loss, and time between surgery and treatment), a median split was applied. Binary variables included whether the patient had a complete cytoreduction, major postoperative complications, ITU admission, and CA125 levels posttreatment. To address imbalanced fields, class weights were adjusted during the model training using this formula:

$$\text{Weight} = \frac{n_{\text{samples}}}{(n_{\text{classes}} \cdot \text{np.bincount}(y))}$$

The surgical outcomes included eight binary classification tasks reflecting peri-operative and post-operative parameters:

1. **Complete Cytoreduction**: whether the surgeon recorded no visible residual disease at the end of the procedure.
2. **Length of Stay**: whether the postoperative stay extended beyond 7 days (median), a clinically meaningful threshold to capture extended hospitalisation.
3. **Operative Time**: documented in minutes.
4. **Estimated Blood Loss**: documented in millilitres.
5. **Intensive Care Unit (ICU) Admission**: whether the patient required ICU admission in the immediate postoperative period.
6. **Clavien-Dindo Grade 3–5 postoperative complications**: complications that necessitated surgical, endoscopic, or radiological intervention (grade 3), critical organ support (grade 4), or that resulted in death (grade 5) [12].
7. **Time between surgery and end of treatment**: indicating potential treatment delays.
8. **End-of-Treatment CA125**: whether the patient's serum CA125, measured upon completion of chemotherapy or follow-up, remained elevated or not ($\geq$35 U/mL).

Structured and unstructured entries in operative, discharge summaries, and follow-up notes were thoroughly reviewed by two clinicians (AL, DDJ) to establish ground truth labels. Identifiers and

 doi:10.20944/preprints202502.1096.v1

protected health information were removed to maintain patient confidentiality, in line with institutional guidelines.

2.3. Large Language Model Architectures

GatorTron is a domain-specific Transformer-based language model developed from clinical notes, PubMed articles, and Wikipedia. It is designed to capture nuances of medical terminology [7].

RoBERTa (Robustly Optimized BERT Pretraining Approach) was chosen for its robust optimization and dynamic masking technique, which enhances generalization and reduces dependency on word order. It is trained on large datasets and has demonstrated superior performance across various NLP benchmarks [13]. RoBERTa has also been applied to detect negations in Dutch clinical texts, showing improved accuracy of clinical information extraction [14]. This suggests that RoBERTa, although not specifically designed for healthcare, can be fine-tuned for medical tasks, whilst specialised models like GatorTron retain their enhanced accuracy and efficiency in medical settings.

RoBERTa was trained and fine-tuned using the Simple Transformers library. Due to its computational cost, GatorTron was trained for 40 epochs using the default Simple Transformers learning rate ($4.00 \cdot 10^{-5}$). The same loss function (cross-entropy loss) was used for consistency. Both models were trained on an 80-20 train-test split. Class weights for imbalanced variables were incorporated during training to mitigate bias. Key hyperparameters are presented in Table 2.

Table 2. Hyperparameter settings for RoBERTa and GatorTron fine-tuning.

	RoBERTa	**GatorTron**
Epochs	40, 60	40
Learning Rate	$4.00 \cdot 10^{-5}$ (*default*), $1.00 \cdot 10^{-7}$, $1.00 \cdot 10^{-5}$	$4.00 \cdot 10^{-5}$ (*default*)
Loss Function	Cross-entropy loss	Cross-entropy loss

2.4. Performance Metrics and Statistical Analysis

We assessed each model's discriminative ability using the area under the receiver operating characteristic curve (AUROC). In addition, we computed accuracy, precision, recall, F1-score, Matthew's correlation coefficient (MCC), and the area under the precision-recall curve (AUPRC). We evaluated differences in performance using DeLong's test for AUROC comparisons where appropriate. Statistical significance was set at $p < 0.05$. Each metric was calculated for every binary classification task, allowing for a thorough comparison of the models' strengths and weaknesses.

3. Results

The descriptive statistics of the continuous variables are summarized in Table 3. The integer to binary transformation was performed by using the median value as the threshold. This concluded in relatively balanced distribution between most variables (Figure 1).

The concatenated field that was created from the merge of operative findings and operative notes is referred to as "Operative Text". The descriptive statistic of the word counts for the three text fields are shown on Table 4.

 doi:10.20944/preprints202502.1096.v1

Table 3. Descriptive statistics for key surgical and clinical parameters.

	Time procedure (minutes)	Length of stay (days)	Time between surgery and end of treatment	Estimated Blood Loss (EBL)	End of Treatment CA125
Count	560	560	540	560	559
Mean	170.39	8.32	91.25	524.50	61.57
Median	150	7	73	425	13
Std	77.55	8.65	85.14	387.78	347.40
Min	30	2	0	50	2
25%	115	6	58	300	8
75%	205	9	119.5	600	23
Max	600	174	1325	4500	5646

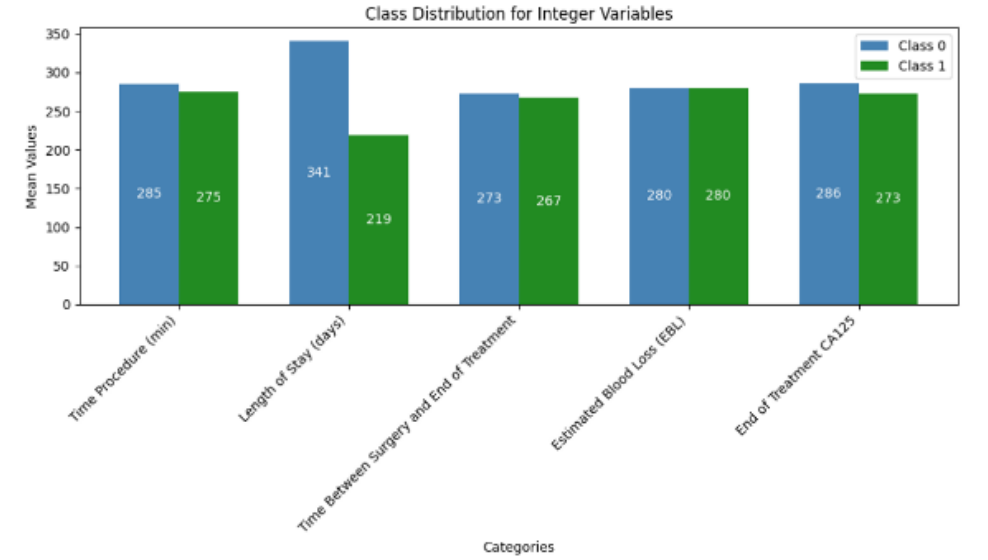

Figure 1. Class distribution of Integer features.

Table 4. Descriptive statistics for operative text, operative findings, and operative notes.

	Operative Text	Operative Findings	Operative Note
Count	560	560	560
Mean	165.12	43.12	122.98
Median	139.5	41	99
Std	90.84	16.44	80.44
Min	0	0	0
25%	101	31.75	64
75%	214	54.25	165
Max	643	86	565

The word count distribution showed significant variation between Operative Findings and Operative Notes fields with the Operative Findings typically having shorter word counts (30-35 words). In contrast, Operative Notes had a wider spread, peaking around 100–150 words and extending up to 565 words in length (Figure 2a). To gain insights into commonly used terms, word clouds were generated (Figure 2b).

 doi:10.20944/preprints202502.1096.v1

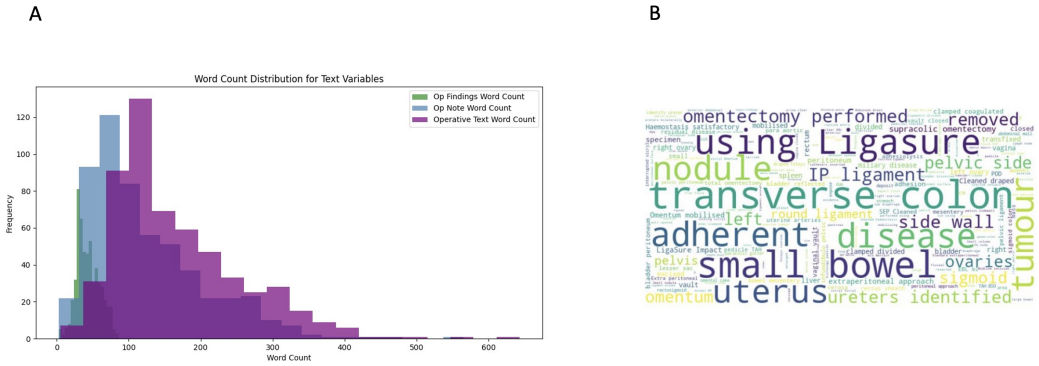

Figure 2. A) Word count distribution of text variables. The variation reflects the different purposes of the fields B) Word cloud visualization of the concatenated texts extracted from operative notes and operative findings. More frequent terms appear larger.

During model convergence, GatorTron displayed a steady increase in validation AUROC with fewer epochs, whilst RoBERTa required more tuning cycles. For RoBERTa, using only the text field of "operative findings" performance accuracies for most outcomes were slightly above 50%, which is a little better than random chance. Similar observations were made for GatorTron (data not shown). Therefore, the "Operative Text" was used as data input. All variables except ICU admission and major post-operative complications were predicted by a 40-epoch training with the default learning rate of 4.00×10^{-5}. The two imbalanced fields underwent weight readjustment to improve model performance. Tables 5 and 6 summarise the performance metrics for RoBERTa and GatorTron across the eight binary tasks, respectively.

Table 5. RoBERTa evaluation metrics – operative text with 40 epochs training.

Target Field	Accuracy	Recall	Precision	F1-score	AUPRC	AUROC	MCC
CCO vs nonCCO	0.802	0.756	0.756	0.756	0.83	0.87	0.589
EBL	0.595	0.661	0.609	0.634	0.61	0.60	0.182
End of Treatment CA125	0.505	1.000	0.505	0.671	0.40	0.33	0.000
Length of stay	0.568	1.000	0.568	0.724	0.51	0.42	0.000
Time between surgery and end of treatment	0.505	0.480	0.471	0.475	0.53	0.52	0.006
Time procedure	0.703	0.857	0.618	0.718	0.79	0.83	0.446
HDU/ITU admission	0.811	0.462	0.632	0.533	0.64	0.86	0.426
Major postop complications	0.910	0.125	0.250	0.167	0.23	0.66	0.133

Table 6. GatorTron evaluation metrics -operative text with 40 epochs training.

Target Field	Accuracy	Recall	Precision	F1-score	AUPRC	AUROC	MCC
Time procedure	0.766	0.857	0.689	0.764	0.81	0.85	0.550
Length of stay	0.577	0.667	0.618	0.641	0.61	0.56	0.127
Time between surgery and end of treatment	0.505	0.540	0.474	0.505	0.53	0.54	0.014
HDU/ITU admission	0.838	0.462	0.750	0.571	0.73	0.88	0.500
EBL	0.658	0.661	0.684	0.672	0.72	0.73	0.314
End of Treatment CA125	0.532	0.429	0.545	0.480	0.51	0.52	0.066
CCO vs nonCCO	0.829	0.756	0.810	0.782	0.87	0.86	0.642
Major postop complications CD (3-5)	0.937	0.125	1.000	0.222	0.28	0.72	0.342

GatorTron consistently outperformed RoBERTa, with the most pronounced differences observed in classification tasks requiring greater domain-specific knowledge—such as predicting the need for ICU admission, identifying postoperative CD grade 3–5 complications, and detecting the possibility of incomplete cytoreduction. Across all tasks, GatorTron's AUROC values showed statistically significant improvements compared to RoBERTa ($p < 0.05$).

 doi:10.20944/preprints202502.1096.v1

The greatest improvement was for the ICU admission outcome (AUROC +0.06), which relied heavily on textual indicators such as extensive organ resection or blood transfusion, both of which were better captured by GatorTron's domain-specific embeddings (Figure 3).

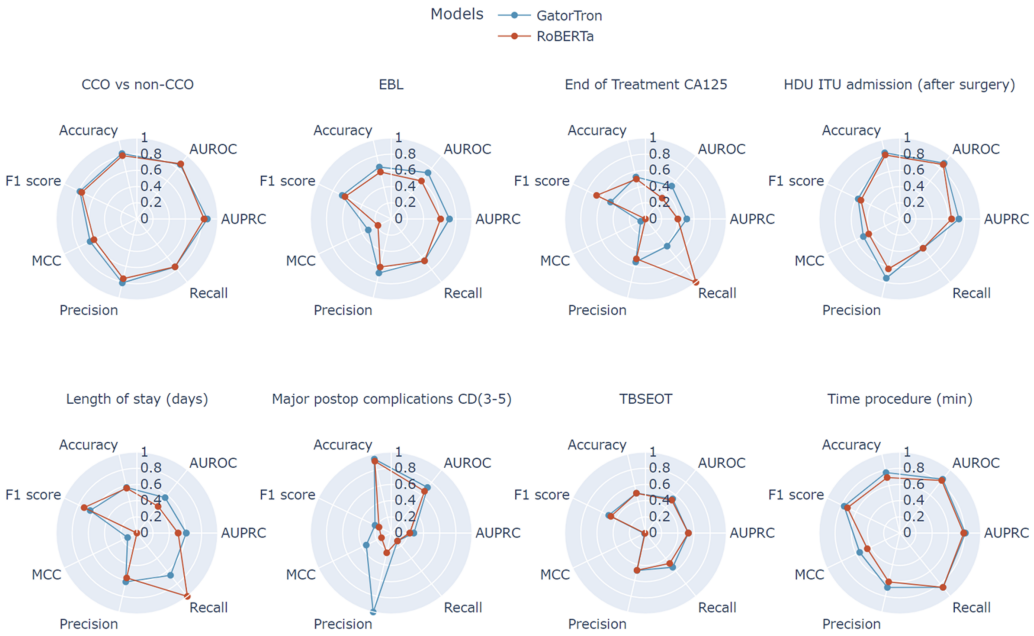

Figure 3. A) Radar plots comparing performance between RoBerta and GatorTron models for all examined clinical tasks using Matthew's correlation coefficient (MCC), recall, precision, F1 score, accuracy, area under precision-recall curve (AURPC), and area under receiver operating characteristic curve (AUROC).

In addition to quantitative performance, the unstructured concatenated text revealed several unexpected correlations when analysed qualitatively. Shorter operative times did not always translate to minimal disease burdens, as some fast-track surgeries involved critically unstable patients or limited attempts at cytoreduction owing to borderline feasibility. Similarly, while one might anticipate that higher EBL would correlate with prolonged hospital stay, certain cases displayed speedy recovery despite substantial intraoperative bleeding. These nuances underline the value of systematic text analysis, which can unveil latent factors influencing EOC surgical outcomes.

4. Discussion

This work showcases the growing intersection of clinical medicine and AI-driven NLP that holds vast promise for improving surgical oncology. Our results demonstrate that domain-specific NLP architectures like GatorTron possess strong discriminatory power for predicting multiple perioperative and postoperative outcomes in advanced EOC surgery. By employing unstructured free-text fields that are typically underutilized or entirely overlooked, GatorTron outperformed the more general-purpose RoBERTa model in nearly every classification task examined. For instance, in predicting procedure time, GatorTron achieved an accuracy of 76.6% and an MCC of 0.550, whereas RoBERTa attained 70.3% accuracy and 0.446 MCC.

There are several potential explanations for GatorTron's superior performance. First, its training on large volumes of clinical text exposes it to unique context not seen in generic language corpora. Consequently, GatorTron more readily identifies conceptually relevant phrases such as "optimal cytoreduction," "residual nodules," or "significant ascites." Secondly, the model's domain-specific embedding space allows for improved handling of ambiguous terms (e.g., "debulking" or "interval cytoreduction"), which may be interpreted differently by general-purpose models lacking clinical context. Finally, aEOC operative reports often contain complex, highly technical descriptions of procedures or tumour spread. GatorTron's capacity to incorporate domain-sensitive relationships likely offers fine-grained predictions. That said, to improve model performance, more precise definitions of

 doi:10.20944/preprints202502.1096.v1

surgical subprocedures—such as the various types of peritonectomy—are required [15], or the use of standardized operative templates is warranted [16].

Beyond comparing these two models, our work underscores the importance of structured and standardized reporting of perioperative details. The combined textual data from operative notes and findings proved more effective for predictive modelling than either alone, suggesting that future informatics in aEOC (and in other oncological fields) should facilitate the integration of multiple textual sources. By capturing potential confounders such as tumour location, infiltration depth, and surgical complexity, text-based approaches can offer a holistic representation of the intraoperative scenario than simple numeric fields like duration of surgery or estimated blood loss.

Despite these promising results, both models encountered challenges in predicting certain outcomes, particularly those involving postoperative events like major complications and length of hospital stay. This limitation suggests that capturing the complexities of postoperative recovery solely from operative text may not be sufficient.

4.1. Clinical Implications

By combining both quantitative and qualitative data, our work offers guidance towards the integration of NLP systems that provide clinicians with evidence-based, risk-adjusted patient counselling [17,18]. Integrative qualitative analysis revealed unexpected nuances, which underline the value of systematic text interrogation to unveil latent factors influencing aEOC surgical outcomes. From a practical viewpoint, the capacity to accurately stratifying perioperative risks in aEOC has clear implications for patient counselling, resource allocation, and postoperative care planning. For instance, an NLP system integrated into the electronic health environment could flag high-risk cases for additional monitoring or perhaps guide surgeons in discussing potential complications and recovery trajectories with patients [19]. This approach might drive more personalised surgical plans, optimising the likelihood of complete cytoreduction whilst mitigating undue harm.

Our study highlights prospects for future AI applications in aEOC more broadly. One intriguing application is the development of ovarian cancer-specific chatbots that could field patient questions on perioperative events, expected complications, or postoperative recovery. By drawing on real-world operative data and outcomes, such a chatbot could deliver targeted, evidence-based advice, aiding both clinicians and patients throughout the treatment continuum.

4.2. Strengths and Novel Contributions

The main strength of the study is the head-to-head comparison of state-of-art GatorTron and RoBERTa models for aEOC-related perioperative predictions, contributing to the ongoing discourse around domain-specific versus general-purpose language models in healthcare, based on same input data. We explored not only immediate measures of surgical success but also post-operative morbidity, resource utilisation (ICU admission), and biochemical endpoints (end-of-treatment CA125), thereby painting the bigger picture of the perioperative aEOC journey. By using actual operative notes and findings from a high-volume tertiary referral centre, our results are highly relevant to clinical practice, with potential for direct translational impact. Counterintuitive associations highlighted the complexity of surgical outcomes, and illustrated how NLP can uncover nuanced insights that might not conform to standard assumptions.

4.3. Limitations

Several limitations warrant consideration. Firstly, a single-institution study with a modest sample size may limit the generalisability of findings. Nevertheless, replication in diverse settings is essential to affirm the broader utility of these models. Secondly, the manual labelling performed by two clinicians was resource-intensive and might have introduced subjective bias. Future research could explore semi-automated labelling or active learning to reduce reviewer burden [20]. Thirdly, both models could be vulnerable to domain shifts—unexpected changes in documentation style, or use of different “macro” templates—potentially affecting model accuracy. There was an exclusive reliance on

 doi:10.20944/preprints202502.1096.v1

the operative notes as input data. These narratives might not capture all factors affecting postoperative complications or time between surgery and end of treatment. An end-to-end risk prediction tool should better combine both numeric and free-text information for maximal predictive precision.

Based on our findings, we propose the prospective implementation of GatorTron or similar domain-specific models into real-time clinical decision support tools within EHR platforms to help surgeons identify patients at higher risk of morbidity or incomplete cytoreduction. Although we prioritised predictive performance over interpretability, the adoption of model explainability might assist clinicians in understanding the basis for aEOC-specific predictions [21]. Finally, this type of work can potentially pave the way towards the development of a bespoke, proprietary, domain-specific chatbot capable of synthesising textual EHR data. Such a system could join similar efforts [22,23] and act as a first point of contact for patient questions, bridging the gap between appointments and streamlining communications with clinical teams.

5. Conclusions

Our study confirms that domain-specific NLP models, exemplified by GatorTron, can effectively employ unstructured EHR data to predict a broad range of important clinical outcomes at aEOC cytoreductive surgery. In comparison, a general-purpose model such as RoBERTa displayed inferior performance, highlighting the value of clinical context and domain-specific language embeddings. Investing in domain-specific architectures, robust data extraction protocols, and comprehensive outcome metrics can yield actionable intelligence that informs the delivery of high-quality, patient-centric care in EOC and beyond. Our work lays the groundwork for developing advanced AI applications, including chatbots.

Author Contributions: Conceptualization, A.L. and E.K.; Data curation, A.L., E.R., C.D. and M.Q.; Software, A.M., E.K., C.D. and E.R.; Supervision, E.K. and D.D.J.; Writing—original draft, E.R., C.D. and A.L.; Writing and editing, M.Q., S.M., A.T., T.B., D.N. and D.D.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was approved by the Institutional Review Board (23/NE/0229/328779/12.01.24).

Informed Consent Statement: Informed consent was obtained from all the patients involved in the study.

Data Availability Statement: Data presnted in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Doufekas, K.; Olaitan, A. Clinical epidemiology of epithelial ovarian cancer in the UK. *International journal of women's health* **2014**, pp. 537–545.
2. du Bois, A.; Reuss, A.; Pujade-Lauraine, E.; Harter, P.; Ray-Coquard, I.; Pfisterer, J. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: A combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials. *Cancer* **2009**, *115*, 1234–1244. https://doi.org/https://doi.org/10.1002/cncr.24149.
3. Chi, D.S.; Franklin, C.C.; Levine, D.A.; Akselrod, F.; Sabbatini, P.; Jarnagin, W.R.; DeMatteo, R.; Poynor, E.A.; Abu-Rustum, N.R.; Barakat, R.R. Improved optimal cytoreduction rates for stages IIIC and IV epithelial ovarian, fallopian tube, and primary peritoneal cancer: a change in surgical approach. *Gynecologic oncology* **2004**, *94*, 650–654.
4. Dagliati, A.; Malovini, A.; Tibollo, V.; Bellazzi, R. Health informatics and EHR to support clinical research in the COVID-19 pandemic: an overview. *Briefings in Bioinformatics* **2021**, *22*, 812–822. https://doi.org/10.1093/bib/bbaa418.
5. Martin-Sanchez, F.; Verspoor, K. Big data in medicine is driving big changes. *Yearbook of medical informatics* **2014**, *23*, 14–20.

 doi:10.20944/preprints202502.1096.v1

6. Khurana, D.; Koli, A.; Khatter, K.; Singh, S. Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications* **2023**, *82*, 3713–3744.
7. Yang, X.; Chen, A.; PourNejatian, N.; Shin, H.C.; Smith, K.E.; Parisien, C.; Compas, C.; Martin, C.; Flores, M.G.; Zhang, Y.; et al. GatorTron: A Large Clinical Language Model to Unlock Patient Information from Unstructured Electronic Health Records, 2022, [arXiv:cs.CL/2203.03540].
8. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); Burstein, J.; Doran, C.; Solorio, T., Eds., Minneapolis, Minnesota, 2019; pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423.
9. Laios, A.; Kalampokis, E.; Mamalis, M.E.; Tarabanis, C.; Nugent, D.; Thangavelu, A.; Theophilou, G.; Jong, D.D. RoBERTa-Assisted Outcome Prediction in Ovarian Cancer Cytoreductive Surgery Using Operative Notes. *Cancer Control* **2023**, *30*, 10732748231209892. PMID: 37915208, https://doi.org/10.1177/10732748231209892.
10. Wu, Y.; Huang, J.; Xu, C.; Zheng, H.; Zhang, L.; Wan, J. Research on Named Entity Recognition of Electronic Medical Records Based on RoBERTa and Radical-Level Feature. *Wireless Communications and Mobile Computing* **2021**, *2021*, 2489754. https://doi.org/https://doi.org/10.1155/2021/2489754.
11. Newsham, A.C.; Johnston, C.; Hall, G.; Leahy, M.G.; Smith, A.B.; Vikram, A.; Donnelly, A.M.; Velikova, G.; Selby, P.J.; Fisher, S.E. Development of an advanced database for clinical trials integrated with an electronic patient record system. *Computers in Biology and Medicine* **2011**, *41*, 575–586. https://doi.org/https://doi.org/10.1016/j.compbiomed.2011.04.014.
12. Clavien, P.A.; Barkun, J.; de Oliveira, M.L.; Vauthey, J.N.; Dindo, D.; Schulick, R.D.; de Santibañes, E.; Pekolj, J.; Slankamenac, K.; Bassi, C.; et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Annals of Surgery* **2009**, *250*, 187–196. https://doi.org/10.1097/SLA.0b013e3181b13ca2.
13. Yang, X.; Bian, J.; Hogan, W.R.; Wu, Y. Clinical concept extraction using transformers. *Journal of the American Medical Informatics Association* **2020**, *27*, 1935–1942. https://doi.org/10.1093/jamia/ocaa189.
14. van Es, B.; Reteig, L.C.; Tan, S.C.; Schraagen, M.; Hemker, M.M.; Arends, S.R.S.; Rios, M.A.R.; Haitjema, S. Negation detection in Dutch clinical texts: an evaluation of rule-based and machine learning methods. *BMC Bioinformatics* **2023**, *24*, 10. https://doi.org/10.1186/s12859-022-05130-x.
15. Laios, A.; Kalampokis, E.; Mamalis, M.E.; Thangavelu, A.; Hutson, R.; Broadhead, T.; Nugent, D.; De Jong, D. Exploring the Potential Role of Upper Abdominal Peritonectomy in Advanced Ovarian Cancer Cytoreductive Surgery Using Explainable Artificial Intelligence. *Cancers* **2023**, *15*. https://doi.org/10.3390/cancers15225386.
16. Querleu, D.; Planchamp, F.; Chiva, L.; Fotopoulou, C.; Barton, D.; Cibula, D.; Aletti, G.; Carinelli, S.; Creutzberg, C.; Davidson, B.; et al. European Society of Gynaecological Oncology (ESGO) Guidelines for Ovarian Cancer Surgery. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society* **2017**, *27*, 1534—1542. https://doi.org/10.1097/igc.0000000000001041.
17. Yang, S.; Yang, X.; Lyu, T.; Huang, J.L.; Chen, A.; He, X.; Braithwaite, D.; Mehta, H.J.; Wu, Y.; Guo, Y.; et al. Extracting Pulmonary Nodules and Nodule Characteristics from Radiology Reports of Lung Cancer Screening Patients Using Transformer Models. *Journal of Healthcare Informatics Research* **2024**, *8*, 463–477. https://doi.org/10.1007/s41666-024-00166-5.
18. Klug, K.; Beckh, K.; Antweiler, D.; Chakraborty, N.; Baldini, G.; Laue, K.; Hosch, R.; Nensa, F.; Schuler, M.; Giesselbach, S. From admission to discharge: a systematic review of clinical natural language processing along the patient journey. *BMC Medical Informatics and Decision Making* **2024**, *24*, 238. https://doi.org/10.1186/s12911-024-02641-w.
19. Sheikh, A.; Jha, A.; Cresswell, K.; Greaves, F.; Bates, D.W. Adoption of electronic health records in UK hospitals: lessons from the USA. *The Lancet* **2014**, *384*, 8–9. https://doi.org/10.1016/S0140-6736(14)61099-0.
20. Chen, K.; Xu, W.; Li, X. The Potential of Gemini and GPTs for Structured Report Generation based on Free-Text ^{18}F-FDG PET/CT Breast Cancer Reports. *Academic Radiology* **2024**. doi: 10.1016/j.acra.2024.08.052, https://doi.org/10.1016/j.acra.2024.08.052.
21. Banegas-Luna, A.J.; Peña-García, J.; Iftene, A.; Guadagni, F.; Ferroni, P.; Scarpato, N.; Zanzotto, F.M.; Bueno-Crespo, A.; Pérez-Sánchez, H. Towards the Interpretability of Machine Learning Predictions for Medical Applications Targeting Personalised Therapies: A Cancer Case Survey. *International Journal of Molecular Sciences* **2021**, *22*. https://doi.org/10.3390/ijms22094394.

 doi:10.20944/preprints202502.1096.v1

22. Siglen, E.; Vetti, H.H.; Lunde, A.B.F.; Hatlebrekke, T.A.; Strømsvik, N.; Hamang, A.; Hovland, S.T.; Rettberg, J.W.; Steen, V.M.; Bjorvatn, C. Ask Rosa – The making of a digital genetic conversation tool, a chatbot, about hereditary breast and ovarian cancer. *Patient Education and Counseling* **2022**, *105*, 1488–1494. https://doi.org/https://doi.org/10.1016/j.pec.2021.09.027.
23. Finch, L.; Broach, V.; Feinberg, J.; Al-Niaimi, A.; Abu-Rustum, N.R.; Zhou, Q.; Iasonos, A.; Chi, D.S. ChatGPT compared to national guidelines for management of ovarian cancer: Did ChatGPT get it right? – A Memorial Sloan Kettering Cancer Center Team Ovary study. *Gynecologic Oncology* **2024**, *189*, 75–79. https://doi.org/10.1016/j.ygyno.2024.07.007.