ENDORISK

Analysis and extension of a risk assessment tool for endometrial cancer.

Ally Sprik

1108867 - MSc Molecular Life Sciences
Gynaecologische Oncologie - RadboudUMC
Laboratory of Systems and Synthetic Biology - SSB70324

Supervisors
WUR: Dr. Cristina Furlan, PhD
RadboudUMC: Dr. Marike Lombaers, MD PhD-candidate
and Dr. Hanny Pijnenborg, MD PhD

February 26, 2024

Abstract

Background

Endometrial carcinoma is the seventh most common cancer in women worldwide and is generally well-treated. Unfortunately, within the current guidelines, preoperative risk stratification of lymph node metastasis (LNM) both misses patients with LNM and causes unnecessary lymphadenectomies, leading to worse prognoses and a lower quality of life, respectively. The ENDORISK Bayesian network can offer a solution, using preoperative markers to generate a risk assessment for diverse cases. This research aims to extend the model with more data; three new evidence variables: imaging myometrial invasion (MI), and the TCGA markers Polymerase Epsilon Exonuclease (POLE) mutation and Mismatch Repair deficient (MMRd), to enhance its viability in the clinic.

Methods and Findings

In this study, the model was iteratively extended and analysed to detail the continued evolution of the model. To this end, three cohorts were used: The original multi-centre training dataset, which was extended from 763 to 952 patients; a dataset from Brno, the Czech Republic, consisting of 581 patients; and a dataset from Tübingen, Germany, consisting of 247 patients. The Brno and Tübingen datasets were used for validation, and finally, the Brno dataset was also incorporated into the model through cross-validation.

First, the model underwent some changes and the training dataset was extended to 952 patients, which resulted in no change in Receiver Operating Area Under the Curve (ROC AUC), staying at 0.82 and 0.86 for Brno and Tübingen respectively when evaluating LNM. When evaluating five-year survival, the ROC AUC remained the same as well, at 0.84 and 0.66 for Brno and Tübingen respectively. Second, the new variables were added to the model iteratively. After adding the TCGA variables, the ROC AUC changed only when evaluating five-year survival in Tübingen to 0.68. After adding the imaging MI node, the ROC AUCs all stayed the same except for the evaluation of LNM using the Brno validation set, increasing the ROC AUC to 0.84. Adding all three variables in the same network gave the same results. The model underwent a minimal viable set assessment, to analyse the performance under non-ideal conditions. Through this assessment, several new minimal evidence sets were discovered, increasing the versatility of the model. Finally, the Brno set was used to cross-validate the model, resulting in a mean ROC AUC of 0.85 and 0.88 for LNM and five-year survival, respectively.

Conclusions

This study revealed that while the additional variables did not significantly increase the predictive power of the model under optimal conditions, they did increase the versatility of the ENDORISK model. This, in turn, enhances its viability for clinical implementation, through the potential of deployment in a variety of clinical settings and restrictions.

Contents

1 Introduction

Over the past decade, personalised medicine has received a growing interest from both researchers and clinicians [1], reflecting a shift towards more individualised treatment plans for patients. This approach has the potential to enhance the average treatment effectiveness and quality of life for patients due to the by supporting more accurate diagnoses of complex cases, thereby preventing unnecessary procedures and reducing the number of missed cases. In 2020, Reijnen et al. published a paper detailing a Bayesian network that integrates preoperative clinical and molecular biomarkers with the ability to estimate the risk of lymph node metastasis in patients with endometrial carcinoma [2]. This study aims to continue the development of this model by adding clinically relevant variables, thereby enhancing the clinical viability of the ENDORISK model.

1.1 Endometrial carcinoma

Endometrial cancer (EC) is the seventh most common cancer in women worldwide, with 417,000 new cases in 2020 [3]. Additionally, the incidence of this disease is steadily increasing throughout the world [4, 5]. The most common risk factors for endometrial cancer are increased oestrogen exposure, obesity, and age [6, 7].

In general, endometrial cancer has a relatively good prognosis, as long as no distant metastasis is present [8]. Inherently, timely identification of tumour progression towards the lymph nodes is clinically relevant. Currently, lymph nodes are removed along with the uterus only in 'high-risk' patients, which are classified by tumour grade from a cervical smear and selective imaging [9]. In the EC population, this translates to about 80% of tumours that are diagnosed as low-risk and 20% as high-risk, as well as an incidence of lymph node metastasis (LNM) of around 10% overall [10]. Figure 1 presents an overview of the population distribution of cases.

Unfortunately, the risk estimation is suboptimal, especially for lower-risk groups. A comparison of clinical guidelines has shown that the latest modified European Society for Medical Oncology (ESMO) guideline, which is still widely used today, only had a sensitivity of 0.18 (95% CI, 0.109-0.273) and a specificity of 0.51 (95% CI, 0.494-0.527) in low-risk groups [11]. Clearly, this approach can be improved, and the aforementioned Bayesian networks could offer a solution through more comprehensive risk assessment [12].

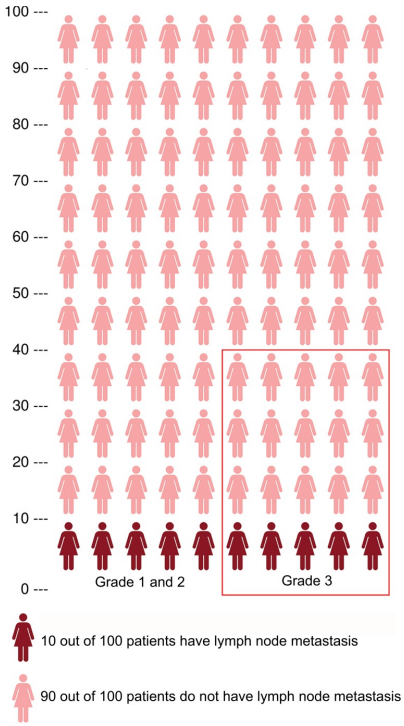


Figure 1: *A simplified view of the division of lymph node metastasis in the population. Reprinted from Reijnen et al. 2019 [10].*

1.2 Bayesian networks

Bayesian networks, also known as belief networks, are probabilistic models where dependencies in the distribution are modelled by a directed acyclic graph. These graphs can be interpreted to represent causal relationships between connected nodes, leading to a clear diagram that is easier to explain and analyse than contemporary machine learning solutions such as random forests or deep learning [13, 14]. Bayesian networks have been applied to a wide array of complex healthcare challenges, most commonly cardiac diseases, cancer, and psychological disorders [12]. One of the reasons they are popular is their ability to use incomplete evidence to arrive at a risk assessment with well-represented uncertainty [15]. This provides the model with a certain clarity when reaching a risk assessment, which is lacking in many other commonly used machine learning techniques [16].

Bayesian networks are built on the concept of conditional probability. This is simply given as P(A|B), representing the concept of the probability of A given the value of B. Since Bayesian networks are directed acyclic graphs, the probability of a single variable can always be abstracted to P(X|parents(X_i)), indicating that the variable probability is a result of the probabilities of its parents. Thereafter, the joint probability of the network can be defined through Equation 1, allowing a distribution to be defined for a specific graph.

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\text{parents}(X_i)) \tag{1}$$

This simple definition of probabilities can easily be applied to common occurrences and logic, e.g., your risk of lung cancer is influenced by potential smoking. This concept can be extended to include factors such as age and Body Mass Index (BMI) categories to encapsulate more causation factors, resulting in the previously mentioned joint probability [17]. Equation 2, provides an example is given of the joint probability in the described lung cancer scenario, where S represents smoking, A represents age groups, B represents BMI categories, and L represents the risk of lung cancer. For ease of demonstration, the causation variables are assumed to be independent of each other. Equation 3 then gives the marginal probability for lung cancer by summing the combinations of probabilities, calculating the overall risk of lung cancer by considering all possible combinations of smoking status, age groups, and BMI categories. Figure 2 shows a representation of the generated Bayesian Network.

$$P(S, A, B, L) = P(S) \cdot P(A) \cdot P(B) \cdot P(L|S, A, B) \tag{2}$$

$$P(L) = \sum_{S,A,B} P(S, A, B, L) \tag{3}$$

Furthermore, Bayesian networks utilise Bayes' theorem to update the prior distributions based on new evidence, as seen in Equation 4 [18]. The equation states that the conditional probability of A given B, also known as the posterior, is dependent on the initial probability of A and B, also known as the prior, and the conditional probability of B given A.

$$P(A|B) = P(A)\frac{P(B|A)}{P(B)} \tag{4}$$

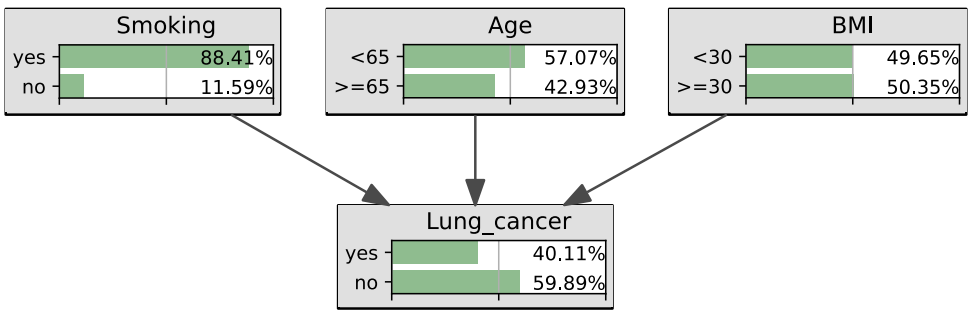


Figure 2: *Example of a Bayesian network.*

A Bayesian network can be developed through two main approaches, which are sometimes combined: structure learning and expert definition [14,19]. Structure learning involves defining the nodes and iteratively adjusting their dependencies to identify the most fitting causal relationships suggested by the data. This method can uncover interesting correlations and potential causations, offering insights into complex datasets. However, it may not always capture the strong causation necessary in the medical field, where justification and interpretability are crucial for clinical decision-making.

1.2.1 Expert Networks

Expert networks leverage domain knowledge of experts to define the relationships within the networks [19]. For example, a doctor can straightforwardly specify that a positive X-ray is caused by lung cancer. This approach eliminates the computational burden of exploring and analysing a large number of possible structures. Another advantage of expert networks is their ability to be trained on incomplete data by incorporating published and expert knowledge [20], an important characteristic in clinical applications, since it is a common challenge in medical research [21]. However, relying solely on expert knowledge is not without its drawbacks. Experts can have biases or may not be up-to-date with the latest research [22], underscoring the need for a diverse panel of experts to mitigate these issues.

Ideally, a fusion of these two methods is applied [23]. In this integrated approach, domain knowledge sets constraints on how the model should be built, such as clarifying that age is not a direct cause of BMI, even though they might be correlated. Combining structure learning with expert insights reduces the computational effort needed to learn the structure while preserving the potential to discover novel relationships between variables. These newly identified relationships can then be evaluated and either integrated into the existing framework or discarded based on further research.

1.3 ENDORISK Model

The original ENDORISK model, published in 2020, is an example of a mixed approach model [2]. The model was initialised through manual definition, where domain knowledge was gained from current literature studies and expert meetings. A clinical tumour progression structure was defined, after which selected preoperative markers were added to the model to estimate the risk of the various disease progression states. Then, using this network as a base, networks were generated by bootstrapping the available data and applying hill-climbing and Tabu-search algorithms. These networks were analysed, and arcs with a strength above 0.7, meaning the arc was present in at least 70% of the networks, were marked for evaluation by the same experts. In Figure 3 an overview of the model published by Reijnen et al. is presented. In Table 1, the nodes of the model are presented and explained.

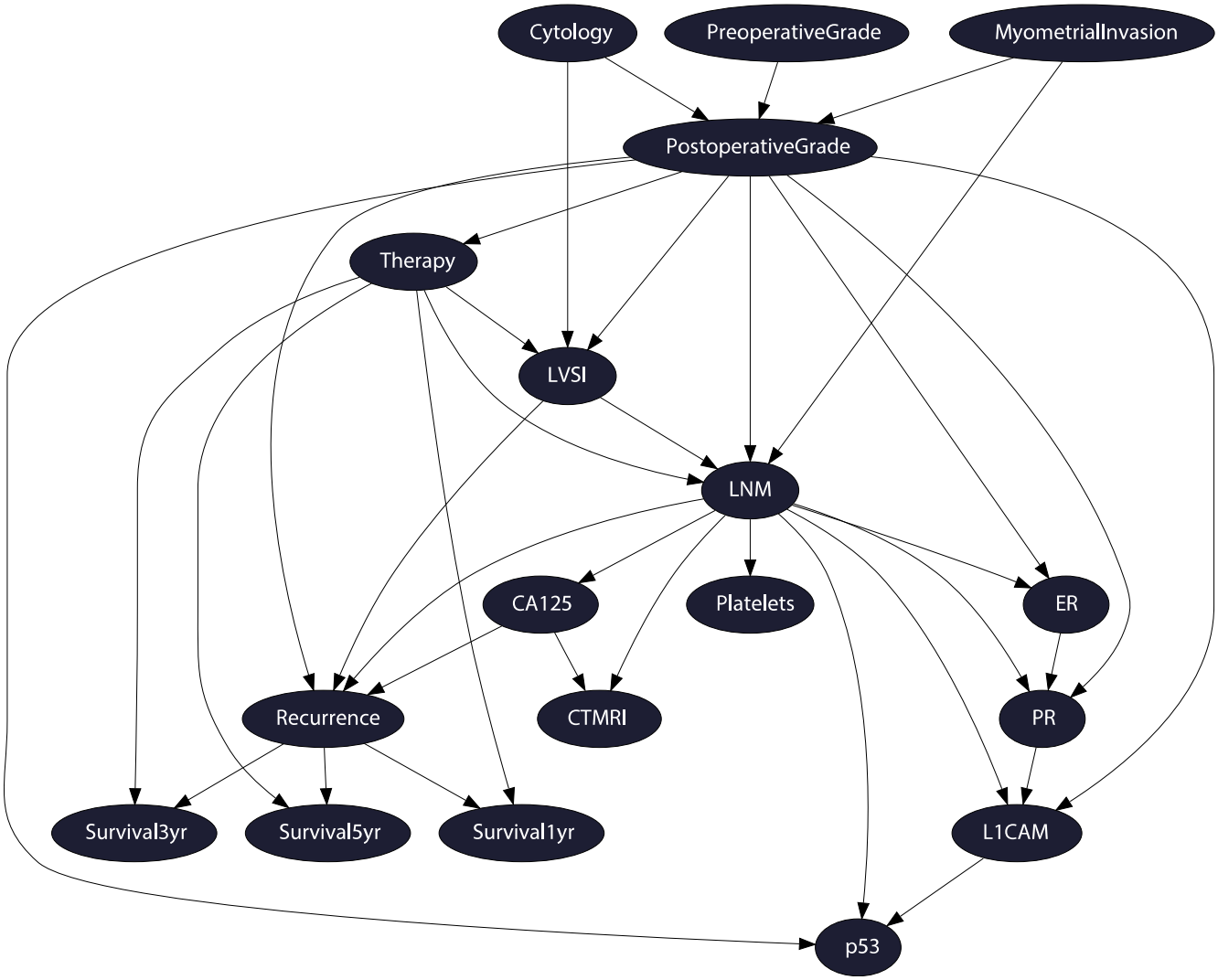


Figure 3: *An overview of the model published by Reijnen et al. 2020*[2].

After the development of the model, it was validated on four different external cohorts from populations in Noord-Brabant in The Netherlands, Bergen in Norway, Tübingen in Germany, and Brno in the Czech Republic[2, 24, 25]. These papers validated the model by generating the risk of LNM and the chance of survival at 5 years using preoperative variables, achieving similar metrics. Evaluating LNM resulted in an area under the receiver operator curve (AUC ROC), a measure of model performance, of 0.82, 0.84, and 0.85 for lymph node metastasis in Norway, the Czech Republic, and Germany respectively[26]. Evaluating 5-year survival resulted in AUC ROCs of 0.70, 0.82, 0.82, and 0.86 for Germany, Norway, The Netherlands, and the Czech Republic respectively. However, when separating patients based on high-grade EC, the Brno team discovered that the model predicted the five-year survival more accurately in low-risk than in high-risk patients. This seems to be supported by the remark of Reijnen et al. that there was a lack of rare high-risk histological subtypes in the original training set.

Table 1: *The nodes present in the model published by Reijnen et al. 2020*[2].

Node	Description	Levels
Cytology	Cervical smear morphology	No, yes (atypical cells present)
Oestrogen receptor	Oestrogen receptor expression in the cervical smear	No ($<$10% of cells), yes ($\geq$10% of cells)
Progesterone receptor	Progesterone receptor expression in the cervical smear	No ($<$10% of cells), yes ($\geq$10% of cells)
L1CAM	L1 Cell-Adhesion Molecule expression in the cervical smear	No ($<$10% of cells), yes ($\geq$10% of cells)
p53	p53 expression in the cervical smear	Wildtype, Mutant
Thrombo-cytosis	Production of platelets	$<$ 400, $\geq$ 400 [$\cdot 10^9$ / l]
CA-125 serum level	Presence of cancer antigen 125	$<$ 35, $\geq$ 35 [IU/ml]
Nodes on CT or MRI	Enlarged nodes on imaging	No, yes ($\geq$ 10 mm short axis diameter)
Lymph vascular Invasion	Tumour growth towards lymph nodes	No, yes

Node	Description	Levels
Preoperative grade	A measure of how differentiated the tumour is, tissue from cervical smear	1, 2, 3
Postoperative grade	A measure of how differentiated the tumour is, tissue from operation	1, 2, 3
Myometrial Invasion	Tumour invasion of myometrial layer	No, yes ($\geq$ 50%)
Lymph node metastasis	Cancer found in lymph nodes	No, yes
Adjuvant therapy	Therapy given to patient	No, chemo-therapy, radiotherapy, chemoradio-therapy
Recurrence	Recurrence of disease	No, local, regional, distant
1-year DSS	Disease specific survival at 1 year	No, yes
3-year DSS	Disease specific survival at 3 years	No, yes
5-year DSS	Disease specific survival at 5 years	No, yes

While the model includes important indicators, additional identified factors could enhance its capabilities. First, in 2013, Levine et al. identified four molecular subgroups that correspond well to disease severity[27]. These groups were incorporated into guidelines in 2021 as Polymerase Epsilon Exonuclease (POLE) mutation, Mismatch Repair deficient (MMRd), No Specific Molecular Profile (NSMP), and abnormal p53 expression, which was already present in the model[28]. Secondly, a preoperative diagnosis of myometrial invasion can be incorporated. Currently, the node for myometrial invasion is based on postoperative assessment of the uterus, which is a significant indicator of LNM and Lymphvascular Invasion (LVSI)[29]. Myometrial Invasion can also be diagnosed preoperatively through imaging, with both magnetic resonance imaging (MRI) and ultrasound (US) performing well at 80% sensitivity, making them good candidates for an imaging node[30, 31].

1.4 Research Aim

The conducted research aims to address several key objectives to answer the central question:

'Can the model's generalisability and preoperative
predictive ability be improved for clinical implementation?'

To formulate an answer to this question, several sub-questions were defined:

1. What correlations and insights can the data reveal concerning the TCGA and imaging MI variables?
2. What effect does the splitting of the therapy node have on the network performance?
3. What effect does the addition of the imaging myometrial invasion and TCGA molecular group nodes have on network performance?
4. What minimal viable evidence sets are available after the changes and updates to the network?
5. What effect does an extension of the training data have on the network performance?

It is anticipated that this research not only has the ability to refine the predictive accuracy of the ENDORISK model but also has the ability to enhance its applicability in clinical settings, contributing to its successful deployment in the field.

2 Methods

The methods employed during the research encompass various applications of data exploration and network permutations. This includes a comprehensive analysis of available patient cohorts to gain more knowledge of the data; an evaluation of different network structures, including the integration of the myometrial invasion imaging and TCGA molecular subgroup nodes; the exploration of extending the training set with newly acquired data; and defining the current clinical risk stratification.

2.1 Languages and Packages

The research was conducted using two programming languages: Python (3.9.7) and R (4.3.2), utilising a variety of packages.

2.1.1 Python Packages

- `itertools (Python 3.9.7 <built-in>)`:
  Data manipulation toolkit
- `numpy (1.26.3)`:
  Numerical Python interface
- `pandas (1.5.3)`:
  Data manipulation and analysis
- `matplotlib (3.8.0)`:
  Plotting library
- `seaborn (0.12.2)`:
  Data visualisation library, enhances matplotlib
- `pyAgrum (1.10.0)`:
  Probabilistic graphical models implemented in Python, used for Bayesian network result generation
- `scikit-learn (1.3.0)`:
  Machine learning library, used for imputation and tools
- `Glasbey (0.2.0)`:
  Colour generation library, used for colour generation in the network visualizations
- `dcurves (1.0.6.2)`:
  Visualization library for decision curves, used for model analysis
- `jupyter (1.0.0)`:
  Implements Jupyter notebooks
- `jupyterlab (4.8.0)`:
  Extends Jupyter notebooks

2.1.2 R Packages

- `dplyr (1.1.3)`:
  Data manipulation toolset
- `ggplot2 (3.4.3)`:
  Data visualisation
- `caret (6.0-94)`:
  Tool package for data manipulation
- `bnlearn (4.8.3)`:
  Bayesian network learning
- `Rgraphviz (2.44.0)`:
  Graph visualisation
- `gRain (1.3.14)`:
  Graphical Independence Networks, used for Bayesian network learning
- `tidyverse (2.0.0)`:
  General data science toolset

2.2 Cohorts

The original training cohort used by Reijnen et al. in 2020 consisted of 763 patients; this cohort was extended to 952 by loosening the definition and including patients for whom the molecular subgroups were also known. Furthermore, the datasets from Tübingen in Germany and Brno in the Czech Republic from the external validations are available. They were slightly extended and redefined through re-examination of samples, after which the Brno dataset consists of 581 patients and the Tübingen dataset of 247 patients. In Appendix A, a table is presented with an overview of the characteristics of the datasets.

2.3 Network training

Through every change and each dataset variation, the network was retrained by using R and the `bnlearn` package and going through the following steps:

1. Define the directed acyclic graph for the network structure.
2. Load the data.
3. Select the nodes from the dataset.
4. Fit the network.
5. Impute the data with the generated network.
6. Fit the network on the imputed data.

2.4 Network Validation

Every variation of the network was validated using the Brno and Tübingen datasets. Except for the exploration of minimal viable sets, this was always performed by making use of all available variables in the preoperative evidence set: ER, PR, p53, L1CAM, CA-125 level, Platelet count, Cytology, imaging MI, MMRd, *POLE*, Preoperative Grade. If a case did not possess the data for a variable, it was left empty. If the network did not possess the node, the value was disregarded for that validation.

2.5 Minimal Set Generation

To evaluate various sets, all combinations of preoperative evidence variables were generated through the Python package `itertools`. These were all evaluated on the Brno validation set to identify well-performing sets. These sets were then abstracted into categories of evidence, such as the category: CA125, three of the immunohistochemical markers (ER, PR, p53, and L1CAM), and preoperative grade. The categories were then individually evaluated, generating several ROC curves. From these results, the mean ROC and its deviation were calculated. Finally, the mean ROC curve, its standard deviation, and the individual ROC curves were plotted to visualise the set performance.

2.6 Cross-validation

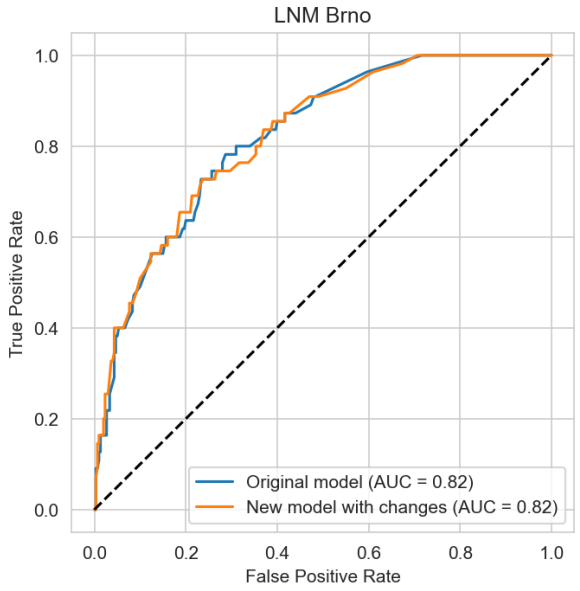
For the cross-validation and addition of the Brno validation set, a stratified five-fold cross-validation was performed using the `caret` package. Each fold comprised 71 cases for validation, while the remaining 510 cases were integrated into the extended training set. This augmented dataset was then utilised to train the network with all additions. Subsequently, the trained networks were validated using the validation folds, resulting in the generation of five ROC curves. These individual ROC curves were averaged to produce a mean ROC curve.

3 Results

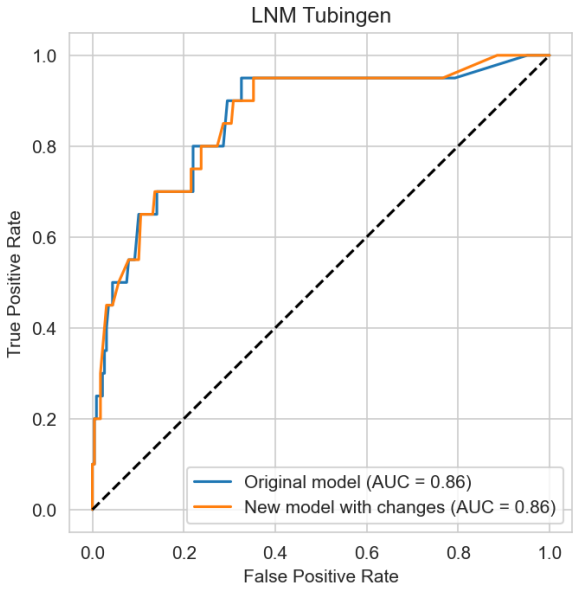
This section outlines the iterative changes made to the network and analyses to systematically present the model's continuous development. The focus is primarily on validating lymph node metastasis (LNM) prediction, with five-year survival validation presented in less detail. Validation graphs for five-year survival are available in Appendix B.

3.1 Basic Network Changes

First, the original network was adjusted by reversing the direction of causality between myometrial invasion and postoperative grade and splitting the therapy node into separate chemotherapy and radiotherapy nodes. The ROC AUC of the original model from Reijnen et al. for LNM was 0.82 and 0.86 using Brno and Tübingen validation set, respectively; for five-year survival it was 0.84 and 0.67 using Brno and Tübingen, respectively. The changed model achieved the same ROC AUC values, with the only difference being the shape of the ROC curves. Figure 4 displays a comparison of the ROC curves for LNM prediction using the Brno and Tübingen datasets.



(a) *ROC curve of the models predicting LNM using the Brno dataset.*

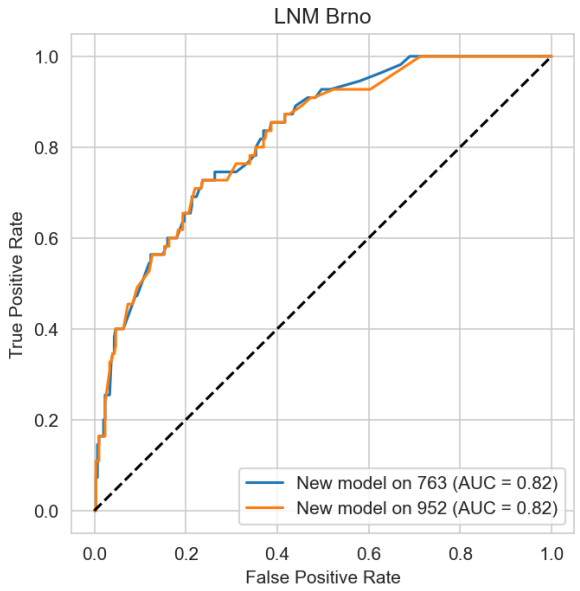


(b) *ROC curve of the models predicting LNM using the Tübingen dataset.*

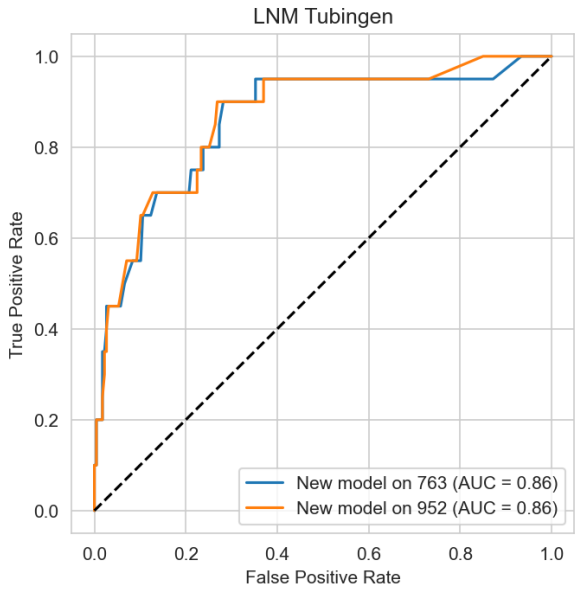
Figure 4: *ROC curves of the original and slightly changed models, trained on the original training set, predicting LNM. The figures show practically no change in ROC AUC and curve after the model changes.*

3.2 Addition of Extra Patients

To facilitate the addition of the TCGA and MRI nodes with more patients, the network was retrained on the newly extended training set. The new model's ROC AUCs remained unchanged when trained on either the original or the extended training set. For LNM, this resulted in ROC AUCs of 0.82 and 0.86 for Brno and Tübingen respectively; for five-year survival, the ROC AUCs were 0.84 and 0.66. Figure 5 shows the ROC curves for LNM comparison.



(a) *ROC curve of the models predicting LNM using the Brno dataset.*



(b) *ROC curve of the models predicting LNM using the Tübingen dataset.*

Figure 5: *ROC curves of the changed model trained on the original 763 patient and the extended 952 patient cohort, predicting LNM. The figures show practically no change in ROC AUC and curve after the model changes.*

3.3 Cohort Analysis of New Variables

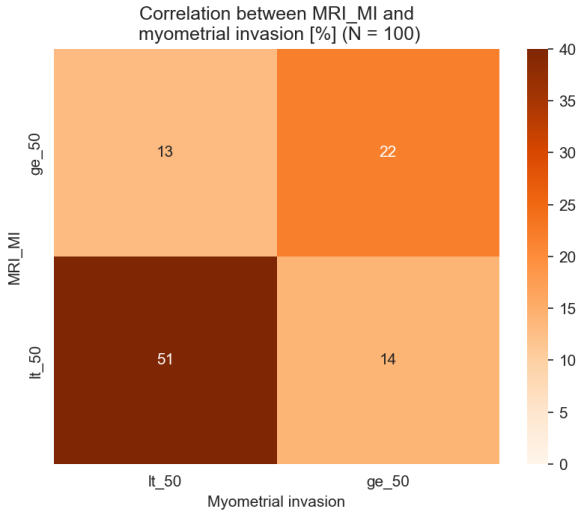
Beyond the dataset overview in Appendix A, additional analyses were conducted to explore the variable distribution within the datasets. These analyses confirmed the presence of known correlations, such as the association between abnormal PR expression and higher tumour grades or the concordance of pre- and postoperative grades. Highlighted below are the analyses of the to-be-added evidence variables: imaging myometrial invasion and the TCGA molecular variables, *POLE* and MSI.

3.3.1 Myometrial Invasion Imaging

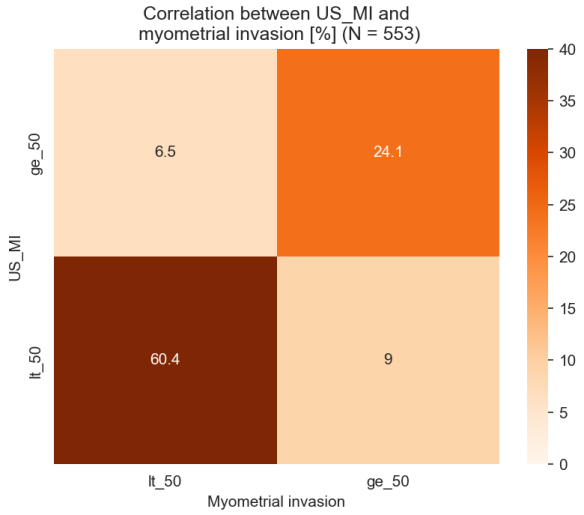
The extended training set and the Czech Republic set include representations of two imaging methods, with the original training set containing MRI MI diagnosis and the Brno set containing US MI diagnosis. The training set, using MRI MI, achieves a True Positive Rate (TPR) of 0.63 and a True Negative Rate (TNR) of 0.78 (n = 100), which falls short of the previously mentioned metrics from the literature. In contrast, the Brno set, utilising US MI, achieves a True Positive Rate (TPR) of 0.79 and a True Negative Rate (TNR) of 0.87 (n = 553), aligning with literature expectations. Cross-tables detailing the correlations are shown in Figures 6a and 6b, respectively.

3.3.2 TCGA molecular groups

The TCGA molecular groups were evaluated in the extended training set and the Brno validation set against LNM to discover their correlations, see Figure 7a and 7b, respectively. The findings are consistent with literature expectations, where abnormal p53 correlates to a higher-risk, MMRd and NSMP are medium-risk, and *POLE* is a lower-risk patient group[27].

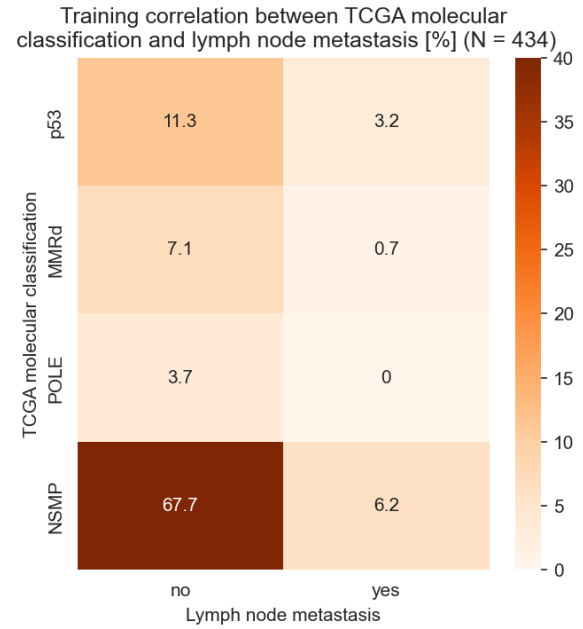


(a) *MRI prediction of myometrial invasion against true myometrial invasion. In this dataset MRI Myometrial Invasion diagnosis has a relatively weak correlation.*

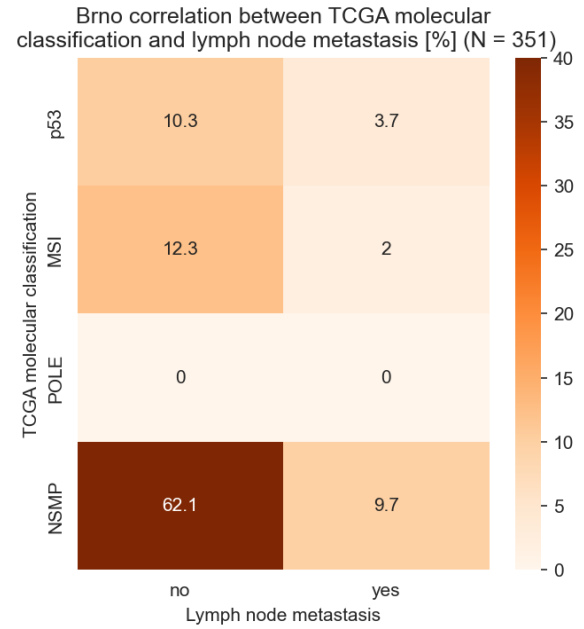


(b) *US prediction of myometrial invasion against true myometrial invasion.*

Figure 6: *Comparison of cross tables for imaging diagnosis of myometrial invasion.*



(a) *Correlation between TCGA molecular groups and LNM in the extended training set.*



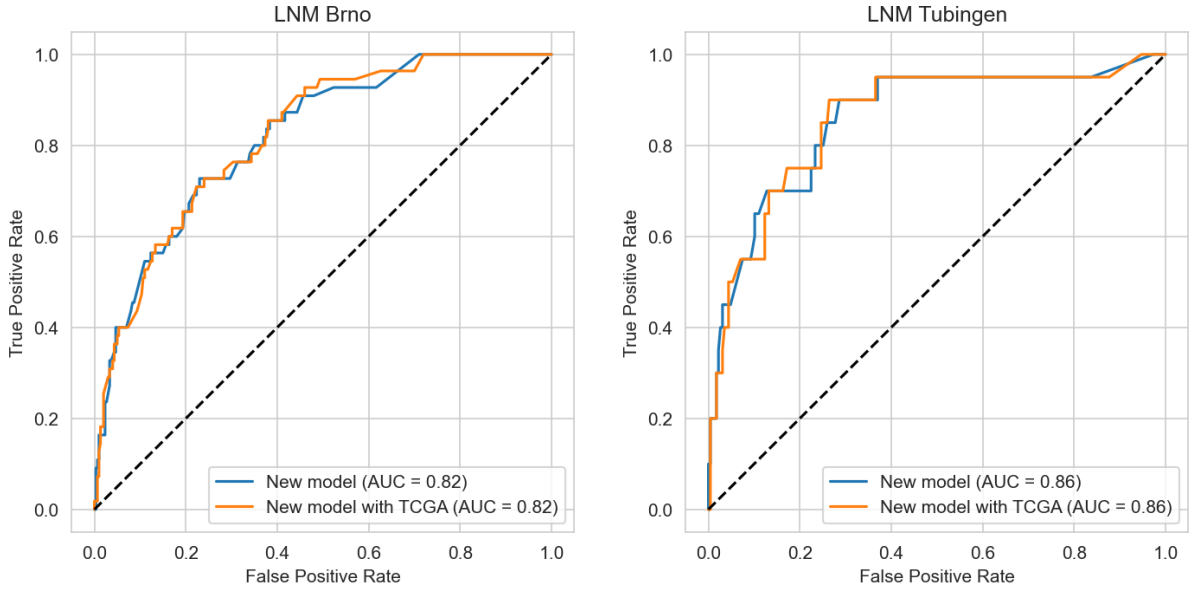
(b) *Correlation between TCGA molecular groups and LNM in the Brno validation set.*

Figure 7: *Comparison of cross-tables for TCGA molecular groups against LNM.*

3.4 Addition of Imaging and TCGA Nodes

3.4.1 Addition of TCGA nodes

Third, the TCGA variables, MMRd and *POLE*, were incorporated into the model, which was then trained on the extended dataset. The ROC AUCs achieved remained consistent at 0.82 and 0.86 for Brno and Tübingen respectively, when evaluating LNM. However, when evaluating five-year survival there was an increase in the ROC AUC for Tübingen, from 0.66 to 0.68; in Brno, the ROC AUC stayed equal in both versions at 0.84. Figure 8 presents a comparison between the ROC curves of the models predicting LNM for Brno and Tübingen.



(a) *ROC curve of the models predicting LNM using the Brno dataset.*

(b) *ROC curve of the models predicting LNM using the Tübingen dataset.*

Figure 8: *ROC curves of the changed model with and without the TCGA nodes trained on the extended 952 patient cohort, predicting LNM. The figures show no change in ROC AUC and curve after the model changes.*

3.4.2 Addition of Imaging

Subsequently, the imaging myometrial invasion was added to the model and trained on the extended dataset. This addition was only validated on the Brno set due to the absence of imaging myometrial invasion data in Tübingen. The inclusion led to a slight increase in the ROC AUC for LNM from 0.82 to 0.84; the ROC AUC for five-year survival remained unchanged at 0.84. The ROC curves for LNM are illustrated in Figure 9.

3.4.3 Addition of Both Imaging and TCGA nodes

The ROC AUC for LNM in Brno increased from 0.82 to 0.84, akin to the addition of MRI alone, while the ROC AUC for Tübingen stayed at 0.86. For five-year survival, the ROC AUC remained at 0.84 for Brno, but increased for Tübingen from 0.66 to 0.69. The ROC curves for LNM are available in Figure 10.

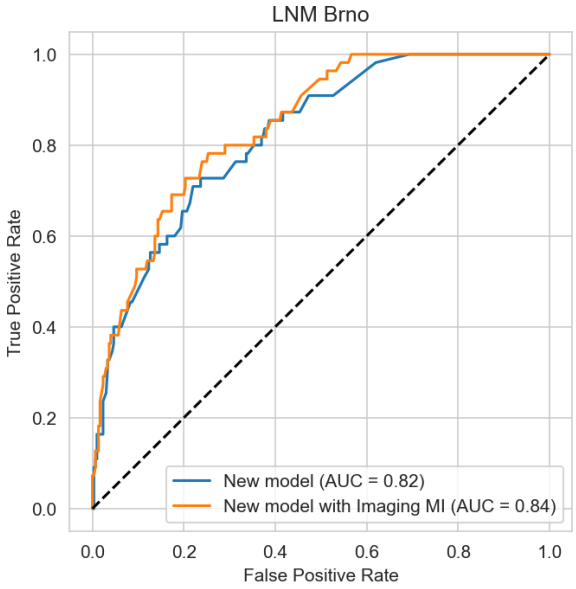
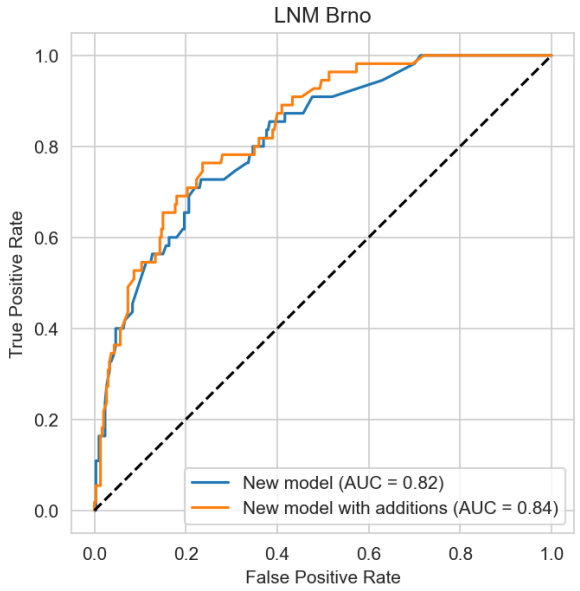
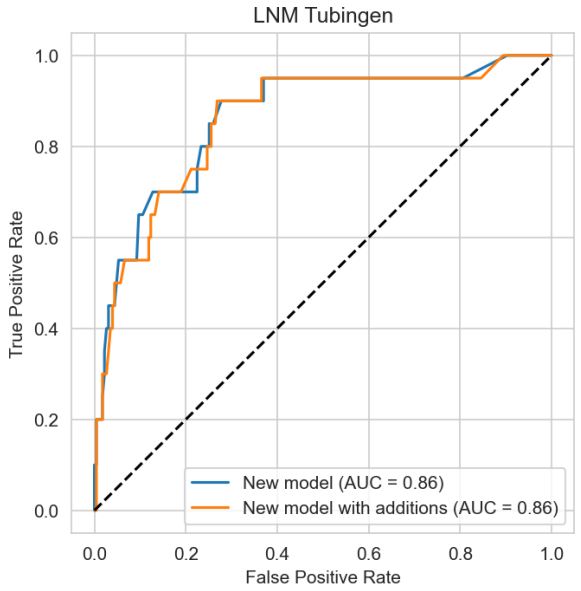


Figure 9: *ROC curves of the model trained on 952 comparing with and without myometrial invasion imaging, using only the Brno dataset. The figures show a slight improvement with the use of imaging to predict LNM.*



(a) *ROC curve of the models predicting LNM using the Brno dataset.*



(b) *ROC curve of the models predicting LNM using the Tübingen dataset.*

Figure 10: *ROC curves of the model trained on 952 comparing with and without both imaging and TCGA nodes, using only the Brno dataset. The figures show a slight improvement with the use of imaging to predict LNM, in accordance with only adding imaging.*

3.4.4 New Network Structure

Figure 11 presents an overview of the new model structure incorporating all the additions.

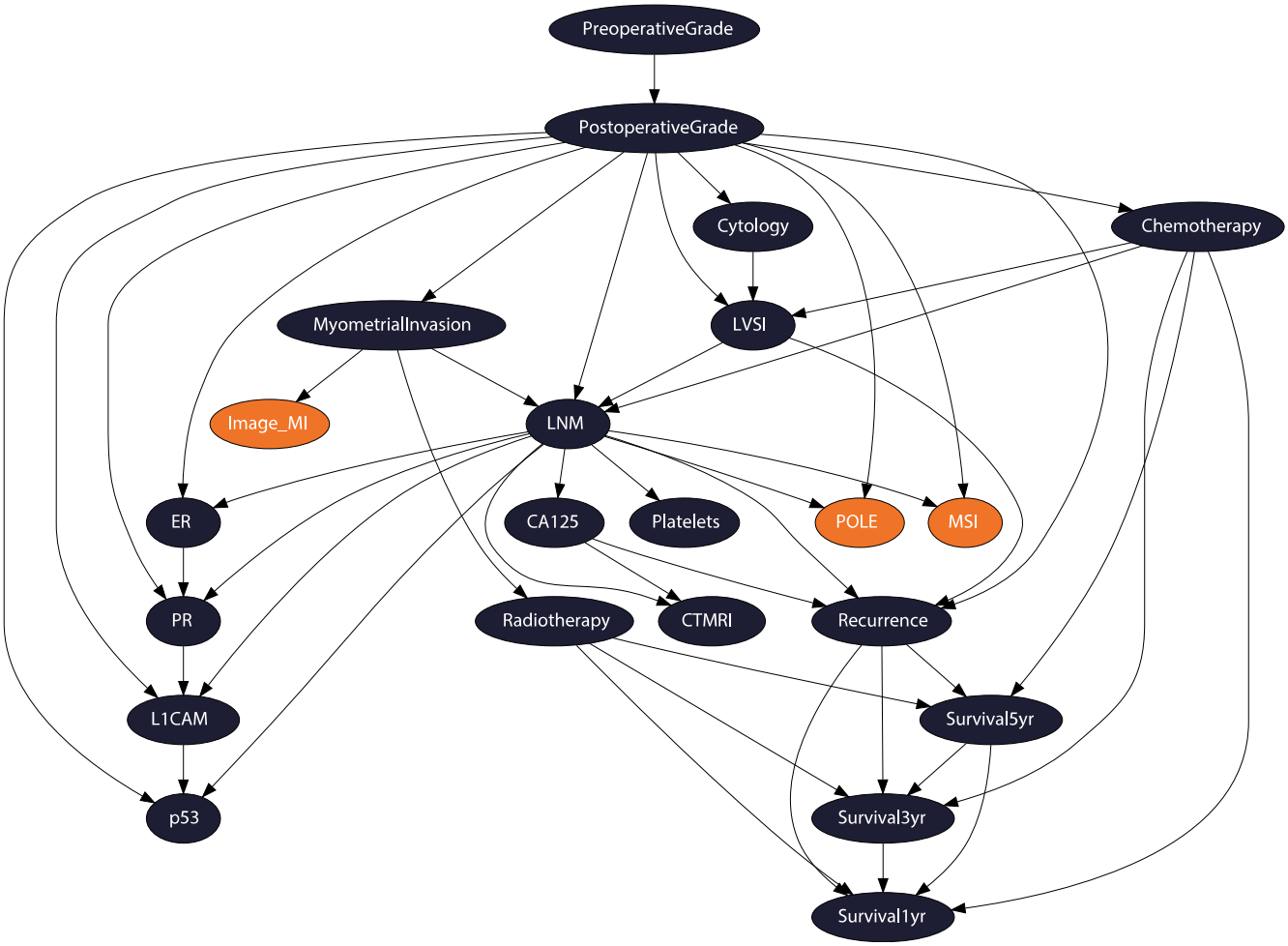
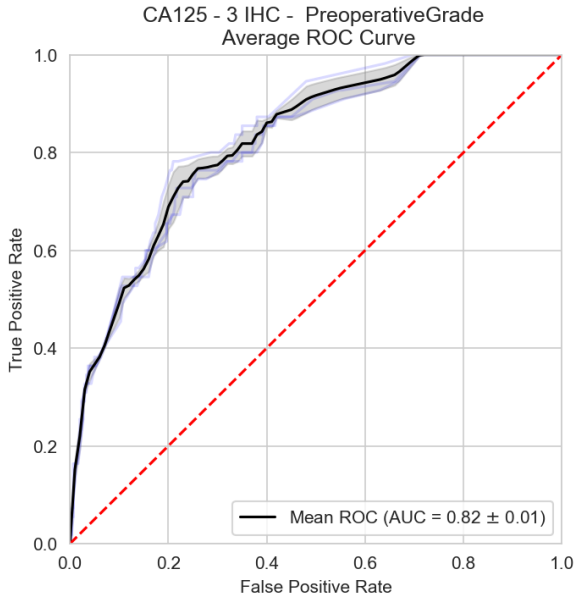


Figure 11: *An overview of the model with all the new additions.*

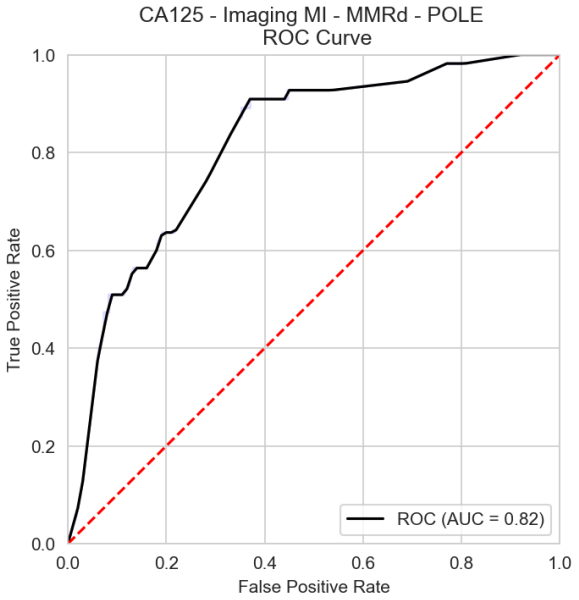
3.5 Minimal Sets

According to Reijnen et al. 2020, for the model to achieve optimal performance, a minimal set was identified that necessarily includes CA125 or Platelets; a minimum of the three immunohistochemical (IHC) markers among ER, PR, p53, L1CAM; and the preoperative grade. With the introduction of the new variables, combinations of all the preoperative variables were created to explore potential minimal evidence sets. These were evaluated on the Brno validation dataset to determine the viability of all possible minimal evidence sets. Given the extensive range of sets, the data is presented as mean ROC curves of generalised groups.

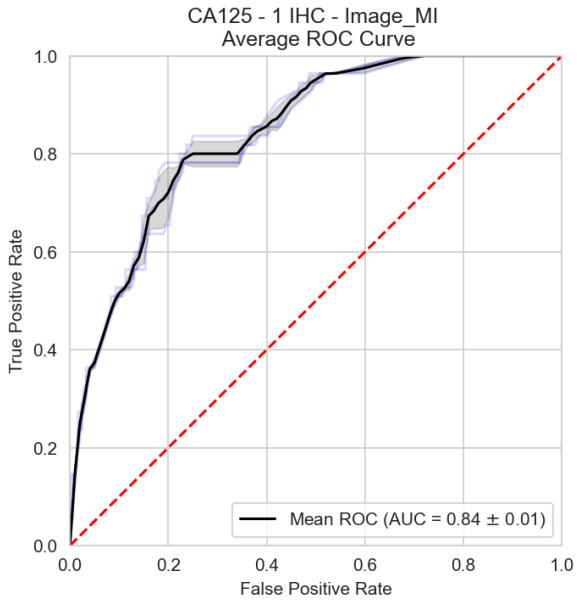
This assessment highlighted CA125 as the strongest LNM predictor, with the mean ROC AUC of sets including it as evidence at 0.83 [0.68 - 0.86] with regards to LNM. Figure 12 showcases a selection of these sets, achieving similar ROC AUCs of 0.82 - 0.84. Furthermore, the analysis identified imaging MI as a significant predictor, with the mean ROC AUC of sets including it at 0.8 [0.7 - 0.86]. A comparison between a set including CA125 with 3 IHCs and one including imaging MI with 3 IHCs is depicted in Figure 13.



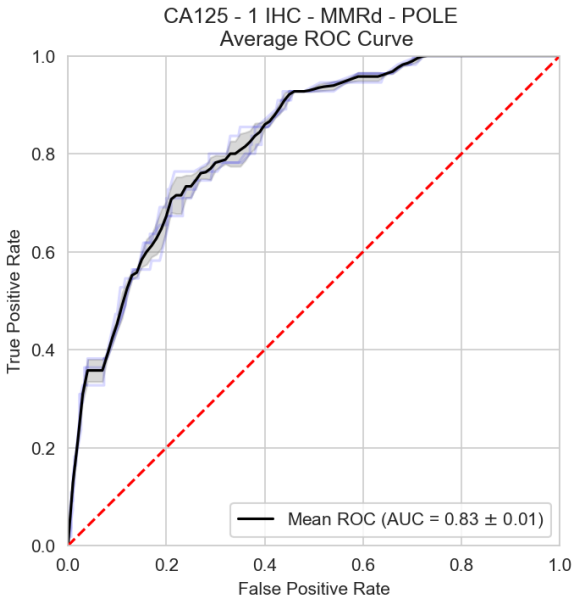
(a) *Mean ROC curve with a standard deviation of the original minimal set of CA125 with 3 immunohistochemical markers and preoperative grade.*



(b) *ROC curve of the minimal set of CA125 with the new additions of imaging MI, MMRd, and POLE.*

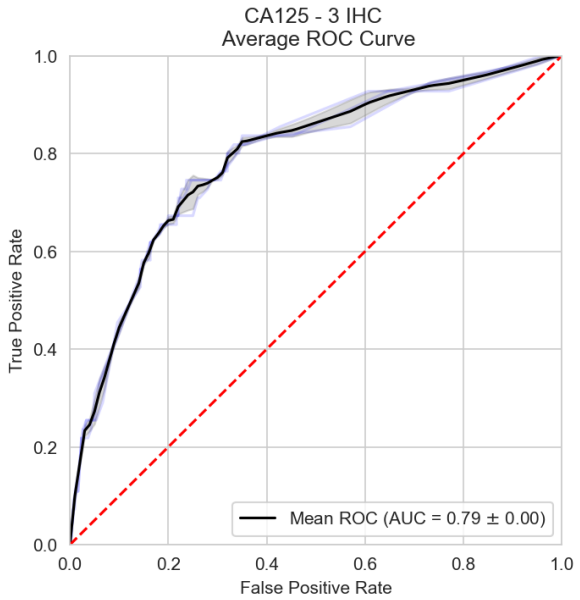


(c) *Mean ROC curve with a standard deviation of the minimal set of CA125 with 1 immunohistochemical marker and imaging myometrial invasion.*

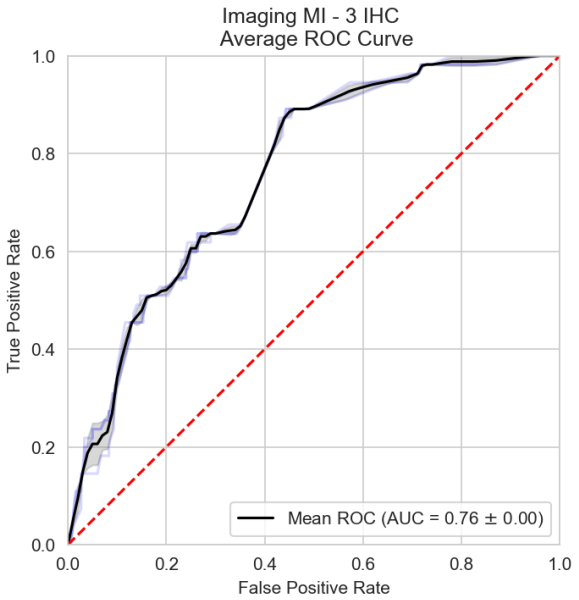


(d) *Mean ROC curve with a standard deviation of the minimal set of CA125 with 1 immunohistochemical marker with MMRd and POLE.*

Figure 12: *Graphs of a selection of well-performing minimal sets evaluated through a mean ROC curve with their standard deviations.*



(a) *Mean ROC with a standard deviation of CA125 with 3 immunohistochemical markers.*

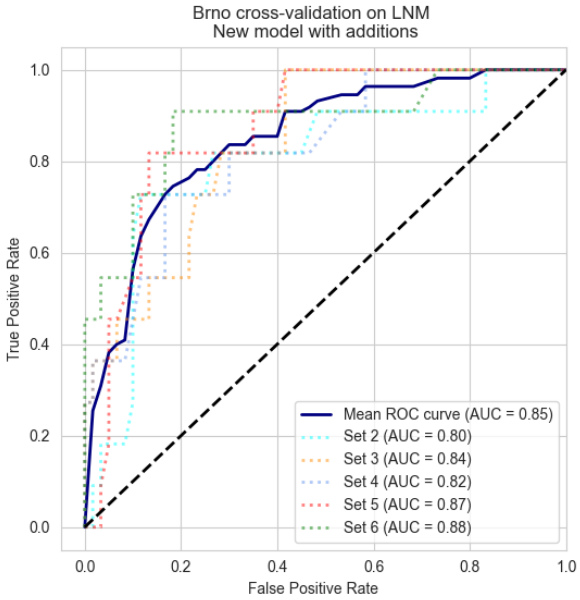


(b) *Mean ROC with a standard deviation of imaging myometrial invasion with 3 immunohistochemical markers.*

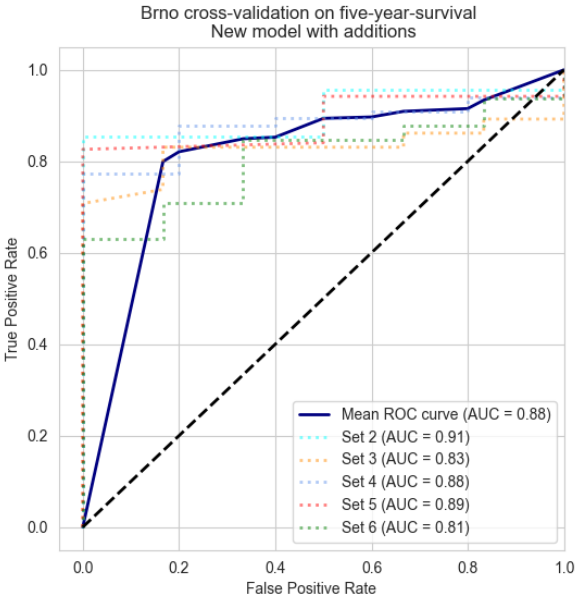
Figure 13: *Comparison between a minimal set of CA125 or imaging myometrial invasion with 3 immunohistochemical markers.*

3.6 Extension with External Dataset

Finally, the training data was augmented and analysed through cross-validation, yielding a mean ROC AUC of 0.85 for LNM and 0.88 for five-year survival. Figure 14 displays the mean ROC curve alongside the ROC curves from the cross-validation subsets.



(a) *Mean ROC of Brno cross-validation on LNM.*



(b) *Mean ROC of Brno cross-validation on five-year survival.*

Figure 14: *Mean ROC of the cross-validation and addition of the Brno dataset with the individual sets plotted opaque.*

4 Discussion

In this study, a comprehensive analysis of the ENDORISK model and the available cohorts was conducted, the main focus of which lay on iterative changes to the model. These adjustments aimed to gain a clear understanding of each step and enhance the predictive power of the model for clinical use. The assessments centred on the risk of lymph node metastasis (LNM) and the chance of five-year survival as the outcomes of the model.

First, the network was slightly adjusted by splitting the therapy node into separate chemo- and radiotherapy nodes, reversing the direction of the relationship between MI and postoperative grade, and the training set was extended from 763 to 952 patients. These changes were implemented to account for the fact that the radiotherapy was dependent on the hospital the data was obtained from [32]; to resolve a causal error, correctly asserting now that myometrial invasion is dependent on the grade; and to accommodate the addition of the imaging MI and TCGA nodes, respectively.

As hypothesised, these changes alone did not impact the predictive capabilities of the network. This is in line with published data, although the TCGA groups do have different risks of LNM and survival [33], with the most significant impact observed in patients with TP53 mutations. These can be identified by p53 immunohistochemical staining, which was already incorporated into the ENDORISK model. This explains the limited impact of adding TCGA classifications to the network. The TCGA groups are primarily relevant for recommending adjuvant therapy treatments and for understanding patterns of recurrence and survival. To ultimately develop ENDORISK as a clinical decision support tool, it is important to integrate these aspects into the current model [34].

Second, the imaging MI and TCGA nodes, MSI and *POLE*, were added to the network. The addition of the TCGA nodes resulted in a significant change in the ROC AUC value only when evaluating five-year survival, increasing from 0.66 to 0.68. The addition of the imaging MI node could only be assessed on the Brno dataset since the Tübingen dataset did not contain any imaging MI data. The cross-validation of Brno added to the original training set increased the ROC AUC from 0.84 to a mean ROC AUC of 0.85 in the case of LNM, and from 0.84 to a mean ROC AUC of 0.88 in the case of five-year survival. This demonstrates the successful and effective capturing of the Brno dataset into the training set and the network.

A notable inconsistency in the validation process was observed in the gap between the ROC AUCs for Brno and Tübingen datasets, when considering five-year survival. As detailed in Appendix A, the Tübingen dataset has a significantly larger population of patients who did not survive beyond five years, with a ratio of 86:14, compared to the Brno dataset ratio of 93:7. This discrepancy underscores the findings of Vinklerová et al. 2022, suggesting that the model tends to be overly optimistic about survival rates, thereby performing better on datasets with a higher survival rate [25].

Due to the marginal impact of integrating new nodes into the model, it was hypothesised that the additional variables were reinforcing the outcomes of the already existing nodes. Therefore, a minimal set assessment was performed to determine the impact of all combinations of preoperative evidence. It was found that CA125 was the strongest preoperative predictor of disease severity in our network, which is consistent with literature [35, 36, 37]. Furthermore, several minimal sets were identified on which the model performed well, validating the hypothesis that the newly added variables bolster the diagnostic capabilities of existing nodes. The discovery of the new minimal viable sets significantly increases the clinical versatility of the model, especially considering the diversity in guidelines and practical constraints such as the availability and cost of equipment.

However, the model still exhibits certain limitations. The MRI data supporting the imaging MI node remains inconsistent with findings reported in the literature, indicating a need for further validation with larger, well-characterised cohorts to clarify this relationship [31, 30]. Additionally, while this study did not primarily focus on five-year survival predictions, the model's optimistic survival forecasts suggest a requirement for additional refinement. Finally, as of this study the validations presented have only been performed by European datasets, an in-depth, successful validation using populations across the world will lend more certainty and support wider implementation of the model.

In examining the landscape of predictive modelling for preoperative endometrial carcinoma assessment, the ENDORISK model emerges as a robust tool. Its performance surpasses that of previous models, including those reviewed by Koskas et al. in 2016 and the more recent advancements by Asami et al. in 2022 [38, 39]. Specifically, the ENDORISK model achieves ROC AUC scores that not only surpass the previously benchmarked threshold of 0.75 but also demonstrate an edge over the logistic regression, support vector machine, and random forest models developed by Asami et al., which presented scores ranging from 0.76 to 0.82 across different datasets.

Furthermore, the ENDORISK model distinguishes itself through its wider array of minimal viable evidence sets, enhancing its clinical relevance. Such versatility is pivotal in medical diagnostics, where the ability to tailor predictive tools to specific patient profiles and clinical guidelines can significantly impact treatment outcomes. Overall, the ENDORISK model stands as a strong tool for preoperative endometrial carcinoma risk assessment in a wide variety of situations.

5 Conclusion

In this research, the ENDORISK model was successfully enhanced with the integration of three novel nodes: imaging for myometrial invasion (MI), and The Cancer Genome Atlas (TCGA) nodes for mismatch repair deficiency (MMRd) and Polymerase Epsilon Exonuclease (*POLE*) mutation. The addition of the imaging MI node marginally improved the model's ability to predict the risk of lymph node metastasis (LNM), elevating the Receiver Operating Area Under the Curve (ROC AUC) from 0.82 to 0.84 in the Brno validation set, without affecting five-year survival outcomes. Conversely, the incorporation of TCGA nodes slightly enhanced the model's predictive accuracy for five-year survival, with the ROC AUC in the Tübingen cohort increasing from 0.66 to 0.68.

Importantly, the inclusion of these nodes revealed new minimal configurations for optimal model operation, significantly enhancing the ENDORISK model's versatility and adaptability for clinical use. This development facilitates the model's integration across varied medical environments, accommodating differences in preoperative diagnostic practices and guidelines. The model's flexibility to adapt to diverse regional medical standards and resources highlights its utility as a versatile tool for personalised patient care. Furthermore, validating these enhancements, the extension of the model to include the Brno cohort through cross-validation revealed a final mean ROC AUC of 0.85 for LNM and 0.88 for five-year survival validation, further reinforcing the predictive performance of the model.

The consistent performance of the ENDORISK model across different cohorts and evidence sets underscores the potential for it to refine clinical decision-making, providing a more customised, precise, and therefore, effective approach to disease management. With its improved predictive accuracy and demonstrated clinical versatility, the updated ENDORISK model is set to contribute to the advance of prognostic practices in oncology, outlining a promising future for personalised medicine.

6 Acknowledgements

I would like to extend my deepest gratitude to Dr. Marike Lombaers, MD PhD-candidate, and Dr. Hanny Pijnenborg, MD PhD, for their daily support and supervision. Furthermore, I would like to extend my gratitude to Prof. Dr. Peter Lucas, Dr. Arjen Hommersom, PhD, and Dr. Casper Reijnen, MD PhD, for their regular assistance and support during this research. I would also like to thank Dr. Cristina Furlan, PhD, for her guidance during several moments in this study. Finally, I would like to thank the research group Gynaecological Oncology for their warm welcome.

7 Future research

The conducted study, while enhancing the ENDORISK model, also highlighted certain limitations, creating opportunities for several research paths to continue the development of the ENDORISK model. First, a pivotal area for future investigation involves focusing on the analysis and improvement of the survival and postoperative sections of the model. Enhancing the accuracy and reliability of these aspects of the model could significantly bolster its overall reliability and clinical applicability. Second, aconducting a comprehensive study with a well-defined dataset to explore the correlation between myometrial invasion and the diagnostic capabilities of magnetic resonance imaging (MRI) and ultrasound methods could substantiate and reinforce the imaging myometrial invasion node. Such research would provide critical insights into the efficacy of these imaging techniques in predicting myometrial invasion, potentially leading to more accurate preoperative assessments. Third, a validation of the ENDORISK model with cohorts from regions outside of Europe will cement its reliability. Utilising well-annotated cohorts from diverse geographical locations would shed light on the model's global applicability and scalability, paving the way for broader implementation. Finally, exploring and developing advanced imputation methods based on published ground truths could enhance the model's performance, especially in scenarios with limited data availability. Effective imputation strategies could improve the model's accuracy and predictive power by providing more comprehensive and reliable datasets for training. Together, these avenues of future research hold the promise of advancing the ENDORISK model, making it a more robust and universally applicable tool for clinical decision-making in the management of endometrial carcinoma. Through targeted improvements and validations, the ENDORISK model can continue to evolve, contributing to the field of personalised medicine and improving outcomes for patients with endometrial carcinoma.

8 Repository information

A copy of the code used to generate and analyse the networks is available at: `https://github.com/allysprik/Internship_2024`. As the data in this study is sensitive, these notebooks have been wiped of all outputs and no data has been included.

References

[1] D. Stefanicka-Wojtas and D. Kurpas. "Personalised MedicineImplementation to the Healthcare System in Europe (Focus Group Discussions)". In: *J Pers Med* 13.3 (Feb. 21, 2023), p. 380. ISSN: 2075-4426. DOI: 10.3390/jpm13030380. pmid: 36983562. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10058568/ (visited on 02/02/2024).

[2] C. Reijnen et al. "Preoperative Risk Stratification in Endometrial Cancer (ENDORISK) by a Bayesian Network Model: A Development and Validation Study". In: *PLoS Med* 17.5 (May 2020), e1003111. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1003111. pmid: 32413043.

[3] H. Sung et al. "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". In: *CA Cancer J Clin* 71.3 (May 2021), pp. 209–249. ISSN: 1542-4863. DOI: 10.3322/caac.21660. pmid: 33538338.

[4] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal. "Cancer Statistics, 2023". In: *CA Cancer J Clin* 73.1 (Jan. 2023), pp. 17–48. ISSN: 1542-4863. DOI: 10.3322/caac.21763. pmid: 36633525.

[5] S. Zhang et al. "Global, Regional, and National Burden of Endometrial Cancer, 19902017: Results From the Global Burden of Disease Study, 2017". In: *Front Oncol* 9 (Dec. 19, 2019), p. 1440. ISSN: 2234-943X. DOI: 10.3389/fonc.2019.01440. pmid: 31921687. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6930915/ (visited on 02/02/2024).

[6] K. H. Lu and R. R. Broaddus. "Endometrial Cancer". In: *New England Journal of Medicine* 383.21 (Nov. 19, 2020), pp. 2053–2064. ISSN: 0028-4793. DOI: 10.1056/NEJMra1514010. pmid: 33207095. URL: https://doi.org/10.1056/NEJMra1514010 (visited on 02/02/2024).

[7] C. Arroyo-Johnson and K. D. Mincey. "Obesity Epidemiology Worldwide". In: *Gastroenterol Clin North Am* 45.4 (Dec. 2016), pp. 571–579. ISSN: 1558-1942. DOI: 10.1016/j.gtc.2016.07.012. pmid: 27837773.

[8] *Cancer of the Endometrium - Cancer Stat Facts*. SEER. URL: https://seer.cancer.gov/statfacts/html/corp.html (visited on 02/02/2024).

[9] A. Oaknin et al. "Endometrial Cancer: ESMO Clinical Practice Guideline for Diagnosis, Treatment and Follow-Up". In: *Annals of Oncology* 33.9 (Sept. 1, 2022), pp. 860–877. ISSN: 0923-7534, 1569-8041. DOI: 10.1016/j.annonc.2022.05.009. pmid: 35690222. URL: https://www.annalsofoncology.org/article/S0923-7534(22)01207-8/fulltext (visited on 02/02/2024).

[10] C. Reijnen et al. "Diagnostic Accuracy of Clinical Biomarkers for Preoperative Prediction of Lymph Node Metastasis in Endometrial Carcinoma: A Systematic Review and Meta-Analysis". In: *Oncologist* 24.9 (Sept. 2019), e880–e890. ISSN: 1549-490X. DOI: 10.1634/theoncologist.2019-0117. pmid: 31186375.

[11] S. Bendifallah et al. "Just How Accurate Are the Major Risk Stratification Systems for Early-Stage Endometrial Cancer?" In: *Br J Cancer* 112.5 (5 Mar. 2015), pp. 793–801. ISSN: 1532-1827. DOI: 10.1038/bjc.2015.35. URL: https://www.nature.com/articles/bjc201535 (visited on 02/02/2024).

[12] S. McLachlan, K. Dube, G. A. Hitman, N. E. Fenton, and E. Kyrimi. "Bayesian Networks in Healthcare: Distribution by Medical Condition". In: *Artificial Intelligence in Medicine* 107 (July 1, 2020), p. 101912. ISSN: 0933-3657. DOI: 10.1016/j.artmed.2020.101912. URL: https://www.sciencedirect.com/science/article/pii/S0933365720300774 (visited on 02/07/2024).

[13] E. Kyrimi et al. "A Comprehensive Scoping Review of Bayesian Networks in Healthcare: Past, Present and Future". In: *Artificial Intelligence in Medicine* 117 (July 1, 2021), p. 102108. ISSN: 0933-3657. DOI: 10.1016/j.artmed.2021.102108. URL: https://www.sciencedirect.com/science/article/pii/S0933365721001019 (visited on 02/05/2024).

[14] N. K. Kitson, A. C. Constantinou, Z. Guo, Y. Liu, and K. Chobtham. "A Survey of Bayesian Network Structure Learning". In: *Artif Intell Rev* 56.8 (Aug. 1, 2023), pp. 8721–8814. ISSN: 1573-7462. DOI: 10.1007/s10462-022-10351-w. URL: https://doi.org/10.1007/s10462-022-10351-w (visited on 02/05/2024).

[15] P. J. F. Lucas, L. C. van der Gaag, and A. Abu-Hanna. "Bayesian Networks in Biomedicine and Health-Care". In: *Artificial Intelligence in Medicine*. Bayesian Networks in Biomedicine and Health-Care 30.3 (Mar. 1, 2004), pp. 201–214. ISSN: 0933-3657. DOI: 10.1016/j.artmed.2003.11.001. URL: https://www.sciencedirect.com/science/article/pii/S0933365703001313 (visited on 02/07/2024).

[16] M. Mourby, K. Ó Cathaoir, and C. B. Collin. "Transparency of Machine-Learning in Healthcare: The GDPR & European Health Law". In: *Computer Law & Security Review* 43 (Nov. 1, 2021), p. 105611. ISSN: 0267-3649. DOI: 10.1016/j.clsr.2021.105611. URL: https://www.sciencedirect.com/science/article/pii/S0267364921000844 (visited on 02/07/2024).

[17] R. L. Schilling. *Measure, Integral, Probability & Processes: A Concise Introduction to Probability and Random Processes. Probab(Ilistical)Ly the Theoretical Minimum*. Dresden: Technische Universität Dresden, 2021. 438 pp. ISBN: 9798599104889.

[18] A. Stuart and J. K. Ord. *Distribution Theory*. Ed. by M. G. Kendall. sixth edition. Kendall's Advanced Theory of Statistics Volume 1. Chichester: Wiley & Sons, 2004. 676 pp. ISBN: 978-0-470-66530-5.

[19] P. Philipp, J. Beyerer, Y. Fischer, and J. Beyerer. "Expert-Based Probabilistic Modeling of Workflows in Context of Surgical Interventions". In: *2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA). Mar. 2017, pp. 1–7. DOI: 10.1109/COGSIMA.2017.7929589. URL: https://ieeexplore.ieee.org/document/7929589 (visited on 02/05/2024).

[20] W. Liao and Q. Ji. "Learning Bayesian Network Parameters under Incomplete Data with Domain Knowledge". In: *Pattern Recognition* 42.11 (Nov. 1, 2009), pp. 3046–3056. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2009.04.006. URL: https://www.sciencedirect.com/science/article/pii/S0031320309001472 (visited on 02/05/2024).

[21] K. Dickersin and I. Chalmers. "Recognizing, Investigating and Dealing with Incomplete and Biased Reporting of Clinical Research: From Francis Bacon to the WHO". In: *J R Soc Med* 104.12 (Dec. 2011), pp. 532–538. ISSN: 0141-0768. DOI: 10.1258/jrsm.2011.11k042. pmid: 22179297. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3241511/ (visited on 02/05/2024).

[22] D. P. Gopal, U. Chetty, P. O'Donnell, C. Gajria, and J. Blackadder-Weinstein. "Implicit Bias in Healthcare: Clinical Practice, Research and Decision Making". In: *Future Healthc J* 8.1 (Mar. 2021), pp. 40–48. ISSN: 2514-6645. DOI: 10.7861/fhj.2020-0233. pmid: 33791459. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8004354/ (visited on 02/05/2024).

[23] H. Amirkhani, M. Rahmati, P. J. F. Lucas, and A. Hommersom. "Exploiting Experts Knowledge for Structure Learning of Bayesian Networks". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 39.11 (Nov. 1, 2017), pp. 2154–2170. ISSN: 0162-8828, 2160-9292. DOI: 10.1109/TPAMI.2016.2636828. URL: http://ieeexplore.ieee.org/document/7776879/ (visited on 02/05/2024).

[24] M. Grube et al. "Improved Preoperative Risk Stratification in Endometrial Carcinoma Patients: External Validation of the ENDORISK Bayesian Network Model in a Large Population-Based Case Series". In: *J Cancer Res Clin Oncol* 149.7 (2023), pp. 3361–3369. ISSN: 0171-5216. DOI: 10.1007/s00432-022-04218-4. pmid: 35939115. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10314833/ (visited on 02/12/2024).

[25] P. Vinklerová et al. "External Validation Study of Endometrial Cancer Preoperative Risk Stratification Model (ENDORISK)". In: *Frontiers in Oncology* 12 (2022). ISSN: 2234-943X. URL: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2022.939226 (visited on 02/07/2024).

[26] K. Hajian-Tilaki. "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation". In: *Caspian J Intern Med* 4.2 (2013), pp. 627–635. ISSN: 2008-6164. pmid: 24009950. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3755824/ (visited on 02/07/2024).

[27] D. A. Levine. "Integrated Genomic Characterization of Endometrial Carcinoma". In: *Nature* 497.7447 (7447 May 2013), pp. 67–73. ISSN: 1476-4687. DOI: 10.1038/nature12113. URL: https://www.nature.com/articles/nature12113 (visited on 09/02/2023).

[28] M. Alexa, A. Hasenburg, and M. J. Battista. "The TCGA Molecular Classification of Endometrial Cancer and Its Possible Impact on Adjuvant Treatment Decisions". In: *Cancers (Basel)* 13.6 (Mar. 23, 2021), p. 1478. ISSN: 2072-6694. DOI: 10.3390/cancers13061478. pmid: 33806979. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8005218/ (visited on 02/09/2024).

[29] J. Wang et al. "Association of Myometrial Invasion With Lymphovascular Space Invasion, Lymph Node Metastasis, Recurrence, and Overall Survival in Endometrial Cancer: A Meta-Analysis of 79 Studies With 68,870 Patients". In: *Front Oncol* 11 (Oct. 21, 2021), p. 762329. ISSN: 2234-943X. DOI: 10.3389/fonc.2021.762329. pmid: 34746002. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8567142/ (visited on 02/12/2024).

[30] G. Spagnol et al. "Threedimensional Transvaginal Ultrasound vs Magnetic Resonance Imaging for Preoperative Staging of Deep Myometrial and Cervical Invasion in Patients with Endometrial Cancer: Systematic Review and Metaanalysis". In: *Ultrasound Obstet Gynecol* 60.5 (Nov. 2022), pp. 604–611. ISSN: 0960-7692. DOI: 10.1002/uog.24967. pmid: 35656849. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9828663/ (visited on 02/09/2024).

[31] A. Ziogas et al. "The Diagnostic Accuracy of 3D Ultrasound Compared to 2D Ultrasound and MRI in the Assessment of Deep Myometrial Invasion in Endometrial Cancer Patients: A Systematic Review". In: *Taiwanese Journal of Obstetrics and Gynecology* 61.5 (Sept. 1, 2022), pp. 746–754. ISSN: 1028-4559. DOI: 10.1016/j.tjog.2022.06.002. URL: https://www.sciencedirect.com/science/article/pii/S1028455922002030 (visited on 02/09/2024).

[32] A. Zanga et al. *The Impact of Missing Data on Causal Discovery: A Multicentric Clinical Study*. Nov. 3, 2023. DOI: 10.48550/arXiv.2305.10050. arXiv: 2305.10050 [cs, stat]. URL: http://arxiv.org/abs/2305.10050 (visited on 02/24/2024). preprint.

[33] A. Jamieson et al. "Endometrial Carcinoma Molecular Subtype Correlates with the Presence of Lymph Node Metastases". In: *Gynecologic Oncology* 165.2 (May 1, 2022), pp. 376–384. ISSN: 0090-8258. DOI: 10.1016/j.ygyno.2022.01.025. URL: https://www.sciencedirect.com/science/article/pii/S0090825822000646 (visited on 02/24/2024).

[34] F. Siegenthaler et al. "Time to First Recurrence, Pattern of Recurrence, and Survival after Recurrence in Endometrial Cancer According to the Molecular Classification". In: *Gynecologic Oncology* 165.2 (May 1, 2022), pp. 230–238. ISSN: 0090-8258. DOI: 10.1016/j.ygyno.2022.02.024. URL: https://www.sciencedirect.com/science/article/pii/S0090825822001433 (visited on 02/24/2024).

[35] C. Reijnen et al. "Improved Preoperative Risk Stratification with CA-125 in Low-Grade Endometrial Cancer: A Multicenter Prospective Cohort Study". In: *Journal of Gynecologic Oncology* 30.5 (Sept. 2019). DOI: 10.3802/jgo.2019.30.e70. pmid: 31328454. URL: https://www-ncbi-nlm-nih-gov.ezproxy.library.wur.nl/pmc/articles/PMC6658593/ (visited on 02/22/2024).

[36] K. Shawn LyBarger, H. A. Miller, and H. B. Frieboes. "CA125 as a Predictor of Endometrial Cancer Lymphovascular Space Invasion and Lymph Node Metastasis for Risk Stratification in the Preoperative Setting". In: *Sci Rep* 12.1 (1 Nov. 17, 2022), p. 19783. ISSN: 2045-2322. DOI: 10.1038/s41598-022-22026-1. URL: https://www.nature.com/articles/s41598-022-22026-1 (visited on 02/22/2024).

[37] K. U. Nithin, M. G. Sridhar, K. Srilatha, and S. Habebullah. "CA 125 Is a Better Marker to Differentiate Endometrial Cancer and Abnormal Uterine Bleeding". In: *African Health Sciences* 18.4 (Dec. 2018), p. 972. DOI: 10.4314/ahs.v18i4.17. pmid: 30766562. URL: https://www-ncbi-nlm-nih-gov.ezproxy.library.wur.nl/pmc/articles/PMC6354887/ (visited on 02/22/2024).

[38] M. Koskas et al. "Evaluation of Models to Predict Lymph Node Metastasis in Endometrial Cancer: A Multicentre Study". In: *European Journal of Cancer* 61 (July 1, 2016), pp. 52–60. ISSN: 0959-8049. DOI: 10.1016/j.ejca.2016.03.079. URL: https://www.sciencedirect.com/science/article/pii/S095980491632041X (visited on 02/25/2024).

[39] Y. Asami et al. "Predictive Model for the Preoperative Assessment and Prognostic Modeling of Lymph Node Metastasis in Endometrial Cancer". In: *Sci Rep* 12.1 (1 Nov. 8, 2022), p. 19004. ISSN: 2045-2322. DOI: 10.1038/s41598-022-23252-3. URL: https://www.nature.com/articles/s41598-022-23252-3 (visited on 02/25/2024).

Appendices

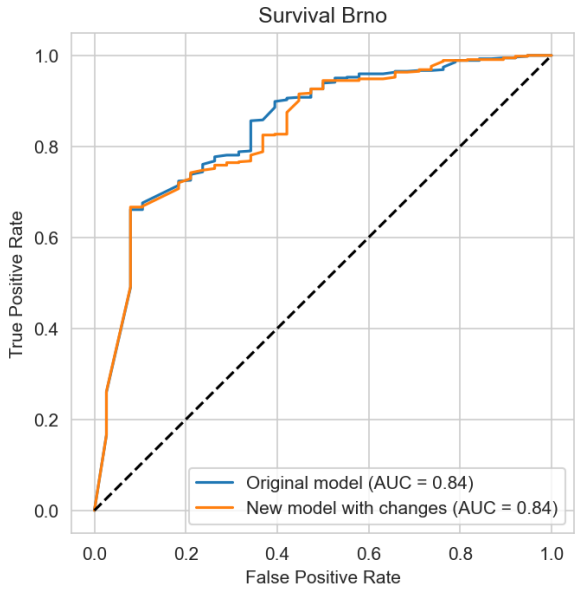
A Dataset table

		Training	Extended Training	Brno	Tübingen
Total N		763	952	581	247
Age (years)		64.9 ± 10.3	64.9 ± 10.2	64.8 ± 9.9	64.0 ± 10.9
Preoperative tumour grade	1	372 (56.8%)	372 (56.8%)	161 (50.3%)	78 (31.6%)
	2	173 (26.4%)	173 (26.4%)	291 (27.8%)	115 (46.6%)
	3	110 (16.8%)	110 (16.8%)	127 (21.9%)	52 (21.9%)
	Unknown	108	297	2	0
ER expression	Negative	76 (10.0%)	76 (10.0%)	43 (7.4%)	27 (11.0%)
	Positive	686 (90.0%)	686 (90.0%)	538 (92.6%)	219 (89.0%)
	Unknown	1	190	0	1
PR expression	Negative	137 (18.1%)	137 (18.1%)	73 (12.6%)	45 (18.2%)
	Positive	620 (81.9%)	620 (81.9%)	508 (87.4%)	202 (81.8%)
	Unknown	6	195	0	0
L1CAM expression	Negative (<10%)	665 (89.4%)	665 (89.4%)	487 (83.8%)	192 (77.7%)
	Positive (≥ 10%)	79 (10.6%)	79 (10.6%)	94 (16.2%)	55 (22.3%)
	Unknown	19	208	0	0
p53 expression	Wild-type	583 (84.9%)	583 (84.9%)	480 (86.0%)	209 (84.6%)
	Mutated	112 (16.1%)	112 (16.1%)	78 (14%)	38 (15.4%)
	Unknown	68	257	23	0
CA-125	< 35 IU/mL	318 (77.9%)	318 (77.9%)	422 (79.8%)	197 (87.2%)
	≥ 35 IU/mL	90 (22.1%)	90 (22.1%)	107 (20.2%)	29 (12.8%)
	Unknown	355	544	52	21
Platelet count	$< 400 \cdot 10^9$1/L	557 (95.7%)	557 (95.7%)	553 (97.4%)	225 (91.5%)
	$\geq 400 \cdot 10^9$1/L	36 (4.3%)	36 (4.3%)	15 (2.6%)	21 (8.5%)
	Unknown	181	370	14	1
MMRd	No	197 (77.3%)	351 (79.2%)	89 (64.0%)	170 (68.8%)
	Yes	58 (22.7%)	92 (20.8%)	50 (36.0%)	77 (31.2%)
	Unknown	508	509	442	0

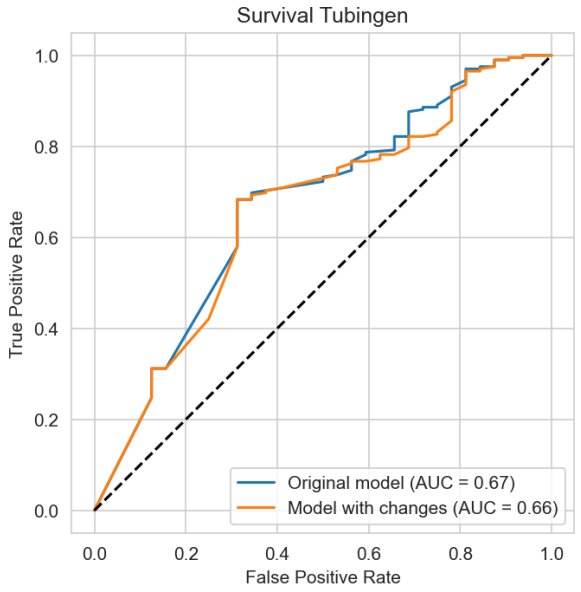
		Training	Extended Training	Brno	Tübingen
POLE mutation	No	233 (91.4%)	408 (92.1%)	137 (98.6%)	218 (88.3%)
	Yes	22 (8.6%)	35 (7.9%)	2 (1.4%)	29 (11.7%)
	Unknown	508	509	442	0
Imaging myometrial invasion	$< 50\%$	65 (63.7%)	65 (63.7%)	384 (69.4%)	0
	$\geq 50\%$	37 (36.3%)	37 (36.3%)	169 (30.6%)	0
	Unknown	661	850	28	247
Cervical cytology	Normal	497 (93.6%)	660 (94.3%)	411 (96.3%)	0
	Abnormal	34 (6.4%)	40 (5.7%)	16 (3.7%)	0
	Unknown	232	252	154	247
Histological subtype	Endometrioid	682 (94.6%)	682 (94.6%)	512 (88.1%)	193 (78.1%)
	Non-endometrioid	39 (5.4%)	39 (5.4%)	69 (11.9%)	54 (21.9%)
	Unknown	42	231	0	0
Myometrial invasion	$< 50\%$	462 (63.0%)	573 (63.5%)	392 (67.5%)	165 (66.8%)
	$\geq 50\%$	271 (37%)	329 (36.5%)	189 (32.5)	82 (33.2%)
	Unknown	30	50	0	0
Postoperative tumour grade	1	317 (41.6%)	377 (40.5%)	148 (51.0%)	84 (46.0%)
	2	289 (37.9%)	357 (38.3%)	296 (25.5%)	110 (35.2%)
	3	157 (20.6%)	198 (21.2%)	137 (23.5%)	45 (18.8%)
	Unknown	0	20	0	8
FIGO stage (surgical)	IA	428 (56.09%)	530 (56.87%)	343 (59.04%)	156 (63.16%)
	IB	196 (25.69%)	230 (24.68%)	94 (16.18%)	52 (21.05%)
	II	51 (6.68%)	64 (6.87%)	68 (11.70%)	13 (5.26%)
	IIIA	20 (2.62%)	23 (2.47%)	18 (3.10%)	23 (9.31%)
	IIIB	4 (0.52%)	6 (0.64%)	6 (1.03%)	1 (0.40%)
	IIIC	43 (5.64%)	54 (5.79%)	45 (7.75%)	1 (0.40%)
	IVA	2 (0.26%)	2 (0.21%)	0	1 (0.40%)
	IVB	19 (2.49%)	23 (2.47%)	7 (1.20%)	0
	Unknown	0	20	0	0

		Training	Extended Training	Brno	Tübingen
LVSI	No	667 (87.4%)	792 (86.7%)	485 (83.9%)	213 (86.2%)
	Yes	96 (12.6%)	121 (13.3%)	93(16.1%)	34 (13.8%)
	Unknown	0	39	3	0
Lymph node metastasis	No	440 (89.3%)	440 (89.3%)	300 (84.5%)	227 (91.9%)
	Yes	53 (10.7%)	53 (10.7%)	55 (15.5%)	20 (8.1%)
	Unknown	270	459	226	0
Five-year-survival	Yes	704 (92.3%)	873 (93.7%)	543 (93.5%)	202 (86.3%)
	No	59 (7.7%)	59 (6.3%)	38 (6.5%)	32 (13.7%)
	Unknown	0	20	0	13
Adjuvant treatment	None	415 (54.5%)	465 (50%)	311 (54.5%)	104 (42.1%)
	Radiotherapy	297 (39.0%)	394 (42.3%)	187 (32.8%)	119 (48.2%)
	Chemotherapy	38 (5.0)	41 (4.4%)	29 (5.1%)	12 (4.9%)
	Chemoradiotherapy	12 (1.5%)	31 (3.33%)	44 (7.7%)	12 (4.9%)
	Unknown	1	21	10	0

B Five-year-survival validations

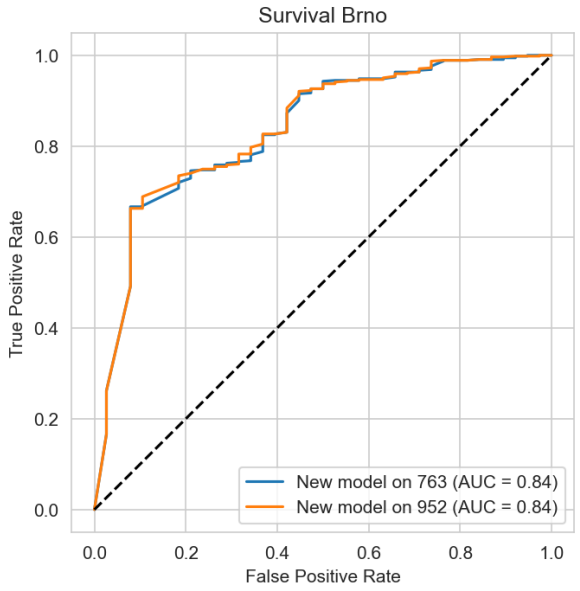


(a) *ROC curve of the models predicting five-year-survival using the Brno dataset.*

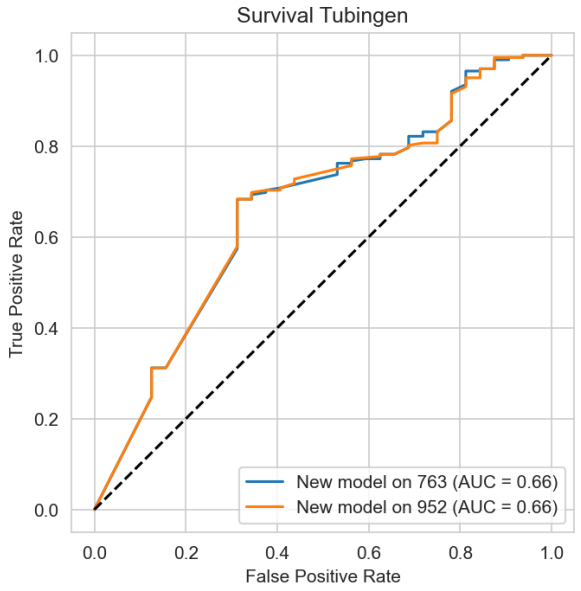


(b) *ROC curve of the models predicting five-year-survival using the Tübingen dataset.*

Figure 15: *ROC curves of the original and slightly changed models, trained on the original training set, predicting five-year-survival. The figures show practically no change in ROC AUC and curve after the model changes.*

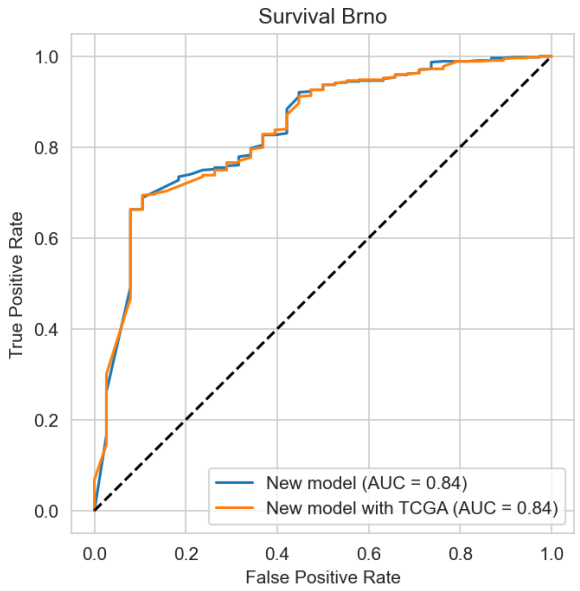


(a) *ROC curve of the models predicting five-year-survival using the Brno dataset.*

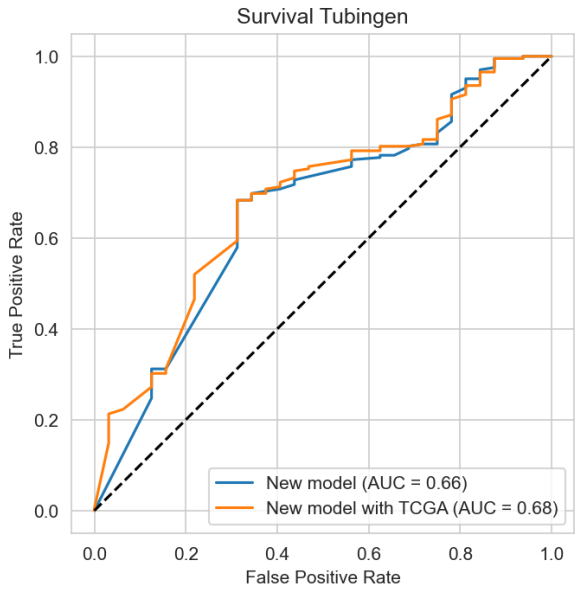


(b) *ROC curve of the models predicting five-year-survival using the Tübingen dataset.*

Figure 16: *ROC curves of the changed model trained on the original 763 patient and the extended 952 patient cohort, predicting five-year-survival. The figures show practically no change in ROC AUC and curve after the model changes.*



(a) *ROC curve of the models predicting five-year-survival using the Brno dataset.*



(b) *ROC curve of the models predicting five-year-survival using the Tübingen dataset.*

Figure 17: *ROC curves of the changed model with and without the TCGA nodes trained on the extended 952 patient cohort, predicting five-year-survival. The figures only show a change in ROC AUC for Tübingen.*

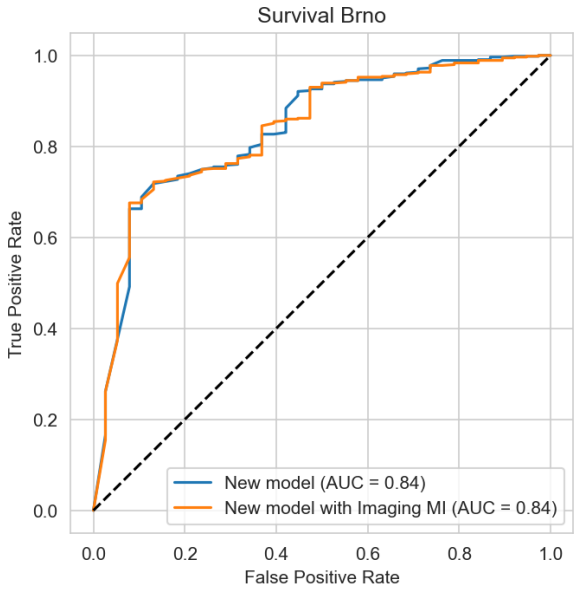
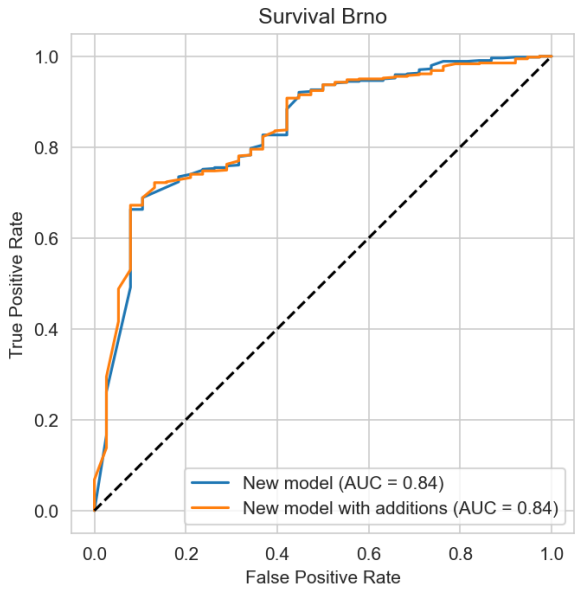
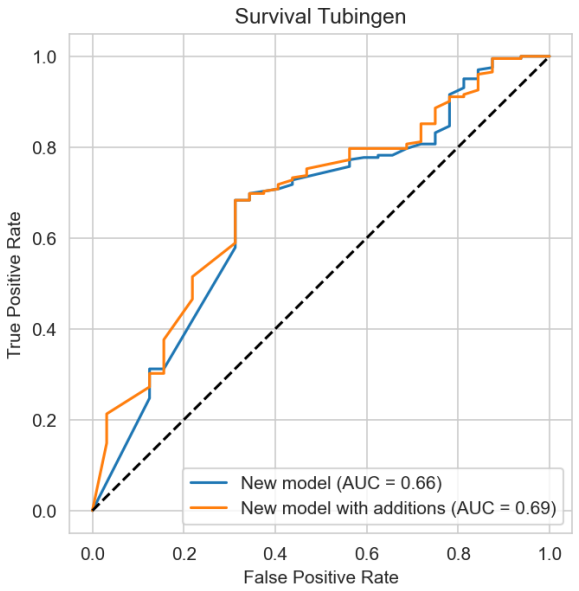


Figure 18: *ROC curves of the model trained on 952 comparing with and without myometrial invasion imaging, using only the Brno dataset. The figure shows no change with the use of imaging to predict five-year-survival.*



(a) *ROC curve of the models predicting five-year-survival using the Brno dataset.*



(b) *ROC curve of the models predicting five-year-survival using the Tübingen dataset.*

Figure 19: *ROC curves of the model trained on 952 comparing with and without both imaging and TCGA nodes, using only the Brno dataset. The figures show a slight improvement with the use of the TCGA variables, MMRd and POLE, to predict five-year-survival for Tübingen, in accordance with only adding MMRd and POLE.*