

Lifestyle Cancer Survival Predictors: Influence of Vegetarian Diet on the Relapse of Endometrial Cancer

Georgy KOPANITSA^a, Oleg METSKER^{b,1}, Ekaterina BOLGOVA^a, Sergey KOVALCHUK^a

^aITMO University, Saint-Petersburg, Russia

^bAlmazov National Medical Research Centre, Saint-Petersburg, Russia

Abstract. Endometrial cancer (EC) is the most common gynecological tumor in high-income countries, and its incidence has increased over time. The most critical risk factor for EC is the long-term unopposed exposure to increased estrogens both exogenous and endogenous. Machine learning can be used as a promising tool to resolve longstanding challenges and support identification of the risk factors and their correlations before the clinical trials and make them more focused. In this paper we present the results of the research of the correlation analysis of Endometrial cancer risk factors. The study was performed with EC patients of the Almazov center in Saint-Petersburg, Russia. All women involved in the current study underwent radical surgical intervention due to EC. After initial cancer treatment, they were referred to the Almazov center outpatient specialists for follow-up visits. Many of them were readmitted of the inpatient clinic due to relapse. We extracted a variety of parameters related to lifestyle, dietary habits, socioeconomic, and reproductive features from the inpatient and outpatient databases of Almazov center. The medical records of the women with enough data were included in the study. Prediction of Progression-free survival (PFS) and overall survival (OS) were analyzed respectively. The AUC of ROC was calculated for PFS = 0.93 and for OS = 0.94.

Keywords. Oncogynecology, Machine learning, Prediction, Lifestyle

1. Introduction

Oncogynecology embraces a group of diseases that originate in the female reproductive organs. There are five main sites of gynecologic cancers: ovaries, uterus (corpus and cervix separately), vulva, and vagina. In the year 2019, more than 1.3 million new cancer cases, arising from those five sites occurred worldwide [1].

Although being grouped, each disease represents a distinct clinical entity with different incidence, risk factors, clinical presentation, and prognosis [2].

Endometrial cancer (EC) is the most common gynecological tumor in high-income countries, and its incidence has increased over time [3-4]. In 2018, there were 382,069 cases of EC worldwide [1]. The majority of patients suffering from EC are over 40 years, although rare cases also occur in younger women [5]. Overall, the median age of women diagnosed with EC peaks around 60 years.

¹ Corresponding Author: Oleg Metsker, Almazov National Research Center, Saint-Petersburg, Russia,
E-mail: olegmetsker@gmail.com

EC belongs to so-called hormone-responsive cancers, which implies that hormone exposure, more specifically, estrogen exposure, is involved in EC occurrence and progression. Unbalanced estrogen exposure may originate from drugs (e.g., hormone replacement therapy) or could be a consequence of early menarche, late menopause, and obesity [3]. Other factors that mimic estrogens, such as endocrine disruptors or cadmium, are also associated with increased risk of EC [6].

The most critical risk factor for EC is the long-term unopposed exposure to increased estrogens both exogenous and endogenous [7]. This can be potentially associated with dietary preferences of a woman. All of them directly or indirectly affect the balance of estrogen in the body of a woman. Early diagnosis of the EC and its relapses plays an essential role. The risk of EC relapse depends on patient age and tumor characteristics, such as type, stage, present of metastases and progesterone receptor expression [8].

Despite the known risk factors, several challenges and poor diagnostic and treatment outcomes have been reported in the recent systematic review, where failure to diagnose malignancies early affect treatment provision and prognosis [9]. Therefore, a deeper analysis of the risk factors and their combinations can provide better insights on the diagnostics and prevention of Endometrial cancer. As clinical trials take much time and effort, study the combination of factors can become even more challenging task. Machine learning approach helps to process large amount of information within shorter time, which makes it useful in the search for factors affecting the progression of cancer. Although it is clear that the disposition to cancer is due to unmodified factors, it may affect when the cancer "shows itself".

Machine learning can be used as a promising tool to resolve longstanding challenges and to support identification of the risk factors and their correlations prior to clinical trials and make them more focused.

In this paper, we present the results of the research of the correlation analysis of Endometrial cancer risk factors.

2. Methods

The study was performed with EC patients of the Almazov national research center.

2.1. Participation Criteria

Women, diagnosed with EC and admitted to the Almazov National research center hospital within the period 2011-2020.

2.2. Clinical Data Collection

All women involved in the current study underwent radical surgical intervention due to EC. After initial cancer treatment, they were referred to the outpatient specialists for follow-up visits. If the relapse was suspected, a patient was referred back to the inpatient hospital.

From the Almazov inpatient and outpatient databases we extracted parameters related to lifestyle, dietary habits, socioeconomic, and reproductive features. Corresponding medical records were retrieved from the existing database. All data was analyzed in unidentified fashion. In total, data from 3845 women were processed. We processed only data of women, who had enough information about their diet in the

anamnesis. Of these, 9.8% women had relapse or died due to EC progression. Patients' median age (IQR) at the time of diagnosis was 70.0 (65.3 - 77.0) and BMI = 26.3 (23.7 - 30.1).

2.3. Endpoint Events

The following indicators were calculated – progression-free survival (PFS) and overall survival (OS), respectively. PFS is the time which a patient lives with cancer without it worsening. OS is the length of time from the defined start point to death from any cause.

2.4. Data Analysis

Each experiment ran in the setting of stratified 5-fold cross-validation (i.e., random 80% of patients were used for training and 20% for testing, target class ratios in the folds were preserved), which represents the trade-off between sensitivity and specificity of the model. The AUC was calculated based on an average of 5 curves (one curve per fold in the setting of 5-fold cross-validation). All the measurements were performed separately per dataset and per model parameter value to determine the best parameters for classifiers as well as optimal data preprocessing.

After determining the optimal dataset and model parameters, we performed a more thorough ROC analysis (100 x 5-fold cross-validation) with the given parameters to find the optimal probability threshold for further label assignment, and the final assessment that utilizes the threshold to calculate the remaining prediction quality characteristics.

We used a series of classification models available within scikit-learn as a pool for the selection of the best predictive methods to be applied within the proposed scheme. A random forest method was applied for the classification with the following parameters: `rf = RandomForestRegressor(n_estimators = 100, n_jobs = -1, oob_score = True, bootstrap = True, random_state = 42)`. A decision tree was developed to partition the data space into cluster regions.

3. Results

The AUC of ROC was calculated for PFS = 0.93 and for OS = 0.94. The decision tree is presented in Figure 1.

The model makes different decisions about outcomes for different ages, in particular physical activity for patients over 78 years of age (patients with low physical activity are more at risk of recurrence. Patients between 78 and 83 years of age with low physical activity were more likely to die (OS event) if prunes were present in their diet. For patients with normal exercise, sunscreen was used as a protective cream. For patients between 60 and 78, depending on age and physical activity, the following factors affected cancer progression: type of diet (consumption of prunes increased the risk, consumption of liver pate increased the probability of survival).

For patients younger than 60, depending on age and physical activity, the following factors influenced the cancer progression: carbohydrate intake, exposure to sun. Both factors increased the risk of cancer progression.

Random forest regression showed 89 % AUC of ROC. The R^2 Training Score was 0.88, the OOB Score came up with 0.06, and the R^2 Validation Score with 0.10.

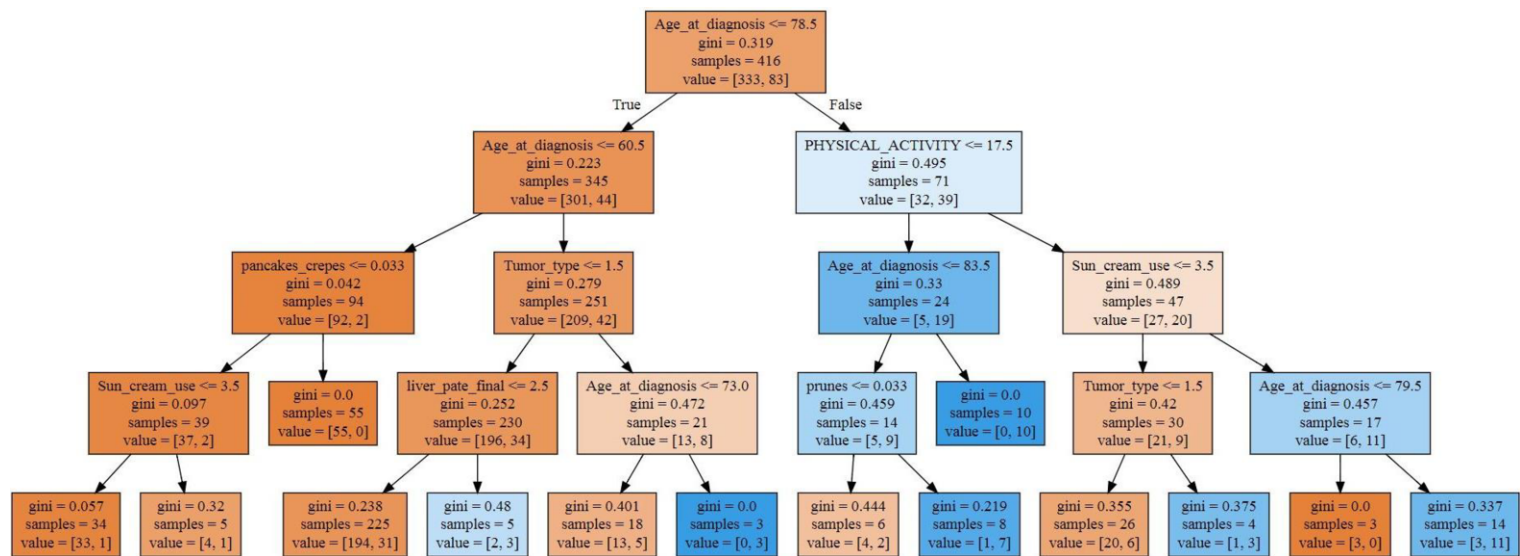


Figure 1. PFS Decision tree

The correlation analysis showed the existence of Relapse and a Diet Type (diet_type). It is an interesting conclusion on Diet Type, because the variable is encoded as follows. 0=mixed, 1=vegetarian, 2=vegan. It turns out that vegans have a higher risk of recurrence. This may be due to a higher consumption of soybeans and plant products that contain phytoestrogens, potential endometrial cancer agents.

Nut consumption can also play its role in the vegetarian diet due to cadmium levels in them. Cadmium in the body has an estrogen-like effect, which is a known EC risk factor.

4. Discussion

Based on our results, the risk of relapse and death due to EC correlated with stage, tumor type, its differentiation and patient's age. These findings are consistent with common knowledge, as it is well-known that prognosis worsens with advanced stage and age. Type 2 EC is also associated with lower survival rate.

Interestingly, relapse and death rate inversely correlated with physical activity. These data complement previous studies that have shown the benefits of regular physical activity in the prevention of endometrial cancer [8-10].

Soft drinks consumption was also correlated with risk of relapse and even stronger with risk of death due to EC. This is probably due to the fact, that high-sugar foods and beverages stimulate insulin production. This is related to increased BMI and type II diabetes and most probably directly to endometrial cancer progression [11,12]. However, further studies are needed due to contradictory results from other works.

Less explained fact relates to sun cream use, which is correlated with increased risk of both relapse and death due to EC. People have different skin phototypes, which affect their sun cream use. Since phototype is inherited the underlying mechanism behind our finding might be potentially hidden in various genetic features.

Other findings related to dietary habits are to be thoroughly studied in well-designed case-control studies.

The patients were followed-up for more than eight years, which is relatively long, while medical records were retrieved from the well-established national database.

The retrospective design of the study has inherent disadvantages. Besides, since patients completed the questionnaires on their own, bias cannot be excluded entirely.

5. Conclusion

The study demonstrates how dietary factors effect risk of EC development. Especially their influence on the relapse of the cancer.

To conclude, here we demonstrate the feasibility of using machine learning in the analysis of the outcome predictors in EC patients.

Acknowledgement

This work was financially supported by the Government of the Russian Federation through the ITMO fellowship and professorship program. This work was supported by the Russian Foundation for Basic Research, grant number 20-37-70047.

References

- [1] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global cancer Observatory: cancer today. Lyon, France: international agency for research on cancer. Cancer Today. 2018.
- [2] The Centers for Disease Control and Prevention's Inside Knowledge: Get the Facts about Gynecologic Cancer Campaign. Series: Inside Knowledge: Get the Facts about Gynecologic Cancer. Centers for Disease Control and Prevention 2012, June 9.
- [3] Colombo N, Creutzberg C, Amant F, Bosse T, Gonzalez-Martin A, Ledermann J, et al. ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and follow-up. *Ann Oncol*. 2016 Jan;27(1):16-41. doi: 10.1093/annonc/mdv484. Epub 2015 Dec 2.
- [4] Lortet-Tieulent J, Ferlay J, Bray F, Jemal A. International Patterns and Trends in Endometrial Cancer Incidence, 1978-2013. *Journal of the National Cancer Institute*. 2018;110(4):354-61.
- [5] Lee NK, Cheung MK, Shin JY, Husain A, Teng NN, Berek JS, et al. Prognostic factors for uterine cancer in reproductive-aged women. *Obstetrics and gynecology*. 2007;109(3):655-62.
- [6] Safe S. Cadmium's disguise dupes the estrogen receptor. *Nat Med*. 2003;9(8):1000-1.
- [7] Endometrial cancer treatment Physician Data Query (PDQ) 2018 [Available from: <http://www.cancer.gov/cancertopics/pdq/treatment/endometrial/healthprofessional>].
- [8] Huijgens A, Mertens H. Factors predicting recurrent endometrial cancer. Facts, views & vision in ObGyn. 2013;5(3):179.
- [9] Collaborators M, Beral V, Bull D, Reeves G. Endometrial cancer and hormone-replacement therapy in the Million Women Study. *Lancet*. 2005;365:1543-51.
- [10] Furness S, Roberts H, Marjoribanks J, Lethaby A, Hickey M, Farquhar C. Hormone therapy in postmenopausal women and risk of endometrial hyperplasia. *Cochrane Database of Systematic Reviews*. 2004(3).
- [11] Henderson BE. The cancer question: an overview of recent epidemiologic and retrospective data. *Am J Obstet Gynecol*. 1989;161(6 Pt 2):1859-64.
- [12] Onstad MA, Schmandt RE, Lu KH. Addressing the Role of Obesity in Endometrial Cancer Risk, Prevention, and Treatment. *J Clin Oncol*. 2016 Dec 10;34(35):4225-4230. doi:10.1200/JCO.2016.69.4638.