A data-driven approach for chemotherapy recommendation model based on deep learning for patients with colorectal cancer in Korea

Jinhyeok Park
Gachon University https://orcid.org/0000-0001-7471-9745
Kang-Yoon Lee
Gachon University
Jeong-Heum Baek
Gachon University Gil Medical Center
Youngho Lee (lyh@gachon.ac.kr)
https://orcid.org/0000-0003-0720-0569

Research article

Keywords: medical errors, Clinical Decision Support System (CDSS), learning-based chemotherapy recommendation,

Posted Date: July 27th, 2020

DOI: https://doi.org/10.21203/rs.2.21064/v2

Version of Record: A version of this preprint was published on September 22nd, 2020. See the published version at https://doi.org/10.1186/s12911-020-01265-0.

Abstract

Background: Recently, the Clinical Decision Support System (CDSS) has attracted attention as a method for minimizing medical errors. To overcome the limitation that existing CDSS does not reflect actual data, we proposed CDSS based on deep learning.

Methods: We proposed Colorectal Cancer Chemotherapy Recommender (C3R), a deep learning-based chemotherapy recommendation model. This supplements the limitation that the existing CDSS is difficult to support data-based decision making. It is configured to study the clinical data generated at Gachon Gil Medical Center and recommend appropriate chemotherapy. To validate the model, we compared the treatment concordance rate with the NCCN Guidelines, a representative cancer treatment guideline, and the results of the Gachon Gil Medical Center's Colorectal Cancer Treatment Protocol (GCCTP).

Results: The treatment concordance rates of the C3R model with the NCCN guidelines were 70.5% for the Top-1 Accuracy and 84% for the Top-2 Accuracy. Also, the treatment concordance rate with the GCCTP were 57.9% for the Top-1 Accuracy and 77.8 for the Top-2 Accuracy.

Conclusions: This model is meaningful in that it is Korea's first colon cancer treatment method decision support system that reflects actual data. In the future, if sufficient data is secured through multi-organization, more reliable results can be obtained.

Background

Generally, becoming a medical specialist takes approximately 10–15 years of training, starting from university entrance. A medical specialist determines the condition of the patient and makes an appropriate diagnosis based on medical and empirical knowledge acquired over a long time. However, despite this long period of training, many patients die every year from medical errors. According to a recent study performed by Johns Hopkins, >250000 people in the United States died because of medical error, which was the third leading cause of death after heart disease and cancer[1]. Medical error costs $20 billion annually; thus, minimizing it is important[2].

The Clinical Decision Support System (CDSS) has attracted attention as a method for minimizing medical errors[3]. The CDSS helps clinicians make rational decisions based on clinical information while diagnosing and treating diseases[4]. It can be applied to prevention, diagnosis, treatment, prescription, prognosis, etc. but is mainly used for diagnosis and treatment. With regard to technology, the CDSS is largely divided into knowledge-based CDSS and non-knowledge-based CDSS[5]. The knowledge-based CDSS provides rule-based decision making, on the basis of the knowledge base of medical data generated in clinical environments. In contrast, the non-knowledge-based CDSS provides decision-making by learning past experiences and patterns in clinical medical information through artificial-intelligence (AI) technologies, such as deep learning and machine learning. With the advancement of AI technology, significant developments are expected in the non-knowledge-based CDSS. However, there is a problem: it

is difficult to secure data and verify the integrity of the obtained data. In particular, in Korea, it is difficult to freely use high-quality medical data under the Personal Information Protection Act[6].

IBM's Watson for Oncology (WfO)—a leading non-knowledge-based CDSS—was developed in 2012 through collaboration with the Memorial Sloan Kettering Cancer Center (MSKCC), which is New York's largest private hospital[7]. WfO recommends a method for diagnosing and treating cancer using models trained by internalizing medical Big Data, including 25,000 patient cases, 290 medical journals, 200 literature, and 12 million pages of specialized data[8, 9]. In a study published by the American Society of Clinical Oncology (ASCO) in 2014, WfO evaluated treatment for 200 leukemia patients, with a consensus of 82.6%[10]. Additionally, MSKCC's 2014 study indicated a high treatment concordance rate for certain carcinomas, including colorectal cancer (98%) and cervical cancer (100%)[11]. However, according to data released in 2017 by Gachon Gil Medical Center, which was the first hospital in Korea to introduce WfO, the diagnosis concordance rate has decreased for most cancers[12]. In particular, the diagnosis concordance rate for colorectal cancer was approximately 65.8%, which was reduced by >25% compared with that when WfO was first introduced[12]. This is because the National Comprehensive Cancer Network (NCCN) guideline, which WfO refers to for diagnosing colorectal cancer, suggests only comprehensive treatment methods and does not consider individual characteristics of patients. In this regard, Strickland E [13] said : Watson for Oncology alleged that it provided useless and sometimes dangerous recommendations. Additionally, WfO is unable to link clinical medical data generated at the clinical site, e.g., electronic medical records (EMR) it is also criticized in terms of usability[14].

In many clinical fields, including WfO, research is being conducted to establish decision support systems. In the field of knowledge based CDSS, Rocha, H. A. L et al.[15] proposed a "Shared-decision making"-based CDSS system for the treatment of prostate cancer. This compares the results of the WfO with the results of the 'shared-decision making' process, which involves informed value-based selection with patients in the absence of the best treatment option. As a result, perfect match was found in 58%, partial match in 15%, and inconsistency in 31%. The main reason for the inconsistency was found to be because patients wanted more treatment than Surveillance. Krens, L. et al [16] have established a "CS rule" based Rule-based CDSS for the treatment of kidney failure in cancer patients. Clinical rules were defined for a total of 18 cytotoxic drugs, and only 112 of the 2681 prescriptions generated warnings. Similar studies also present a differential diagnosis of pulmonary fibrosis CDSS[17].

In the field of non-knowledge-based CDSS, Pyo, K. H. et al.[18] built a model for predicting anti-PD-1 cancer immunotherapy response using clinical and blood-based data from lung cancer patients. As a machine learning model, supervised learning models such as LASSO, Ridge, Elastic Net, SVM, ANN, and RF were used. Among them, Ridge regression model (AUC: 0.78) showed excellent performance in predicting anti-PD-1 response. Kenny H. Cha et al.[19] proposed CDSS based on Computed Tomography (CT) for the evaluation of response to treatment of muscle invasive bladder cancer. In order to confirm the degree of response before and after chemotherapy, they constructed "CDSS-T", a deep learning model based on the Convolutional Neural Network, using CT images and radioactive features. The mean AUC value of "CDSS-T" was 0.80, and the AUC value of 0.74 for doctors who did not use "CDSS-T". Various

studies have been conducted to establish the CDSS, but most of them were rule-based CDSS that does not reflect real-world data or CDSS that simply predicts the onset. According to the “2016 National Cancer Registration Statistics” released by the Central Registration Center in 2018, colorectal cancer is the second most common type of cancer (after gastric cancer) and is the fourth most common type of cancer in the United States[20, 21]. Additionally, because the recurrence rate (e.g., primary cancer recurring or new cancer) after the treatment of colorectal cancer is higher than that for other carcinomas, it is important to select an appropriate treatment method for chemotherapy recommendation.

Therefore, to resolve the limitation of the existing non-knowledge-based CDSS not reflecting the actual data, we developed an EMR data-based deep-learning model called the Colorectal Cancer Chemotherapy Recommender (C3R). In the process, the oversampling technique was used to solve the overfitting problem caused by class imbalance, which led to a significant performance improvement. In addition, the Deep-Surv model was used to support more accurate decision-making to check which factors influenced the chemotherapy recommendation in the deep learning process.

The structure of the manuscript is as follows: in “Background” section, we provide a brief background on clinical decision support system. In “Methods” section, we present our contributions, including an explanation of our implementation of data extraction & preprocessing and our proposed treatment recommendation model C3R. In “Results” section, we describe the experimental design and results. “Conclusion” and “Discussion” sections conclude the manuscript.

Methods

Dataset

In this study, the EMR data of Gachon Gil Medical Center, which launched IBM’s WfO in Korea for the first time in 2016, were used. Gachon Gil Medical Center has obtained reliable data through WfO and multi-disciplinary medical treatments involving face-to-face interaction between patients and three or more cancer treatment specialists. Data were collected from patients who had undergone colorectal cancer surgery between 2004 and 2012. The dataset included information such as the Demographic, Disease Characteristics, Cancer Characteristics, Tumor Characteristics, Treatment Characteristics, Survival Characteristics, and Genetic Characteristics. This standard information is based on the colorectal cancer Common Data Model (CDM) definition employed by five domestic hospitals, including Gachon Gil Medical Center.

The EMR data of the Gachon Gil Medical Center are divided into Scan EMR, XML EMR, and Database EMR according to the storage method. In Scan EMR and XML EMR, there is a possibility that the data are deleted or incorrectly entered while the medical record administrator checks the record. Therefore, to verify the reliability and integrity of the extracted dataset, several colon cancer specialists and medical record administrators collaborated to review the chart.

The chart review involved a detailed three-step process over a six-month period, involving several colorectal cancer specialists and medical record administrators. In the first step, the extracted data were checked to ensure that they were properly mapped with the code described in the colorectal cancer CDM definition document and extracted from the correct location through the normal method. In the second step, to ensure the reliability of the extracted data, an operation to identify and remove incorrect data, such as data redundancies and incorrect inputs, was performed. This chart review process was repeated at monthly intervals under the supervision of a colorectal cancer specialist. In the final step, to reduce unnecessary biases in the deep-learning model training, the colorectal cancer specialists selected first-priority variables that are highly related to survival. Table 1 presents six categories and the variables in each category.

Table 1. Dataset description

Input Variables	
Demographics	Age, Sex, ASA, BMI, Smoking History
Disease Characteristics	DM History, Pulmonary Disease, Liver Disease, Heart Disease, Kidney Disease
Cancer Characteristics	Prior Diagnosis Cancer, Initial CEA, Perforation, Obstruction, Emergency, Lymphovascular Invasion, Perineural Invasion, Distal Resection Margin, Radial Margin, Harvested Lymph Node, Positive Lymph Node, Early Complication, Recurrence
Tumor Characteristics	Hereditary Colorectal Tumor, Tumor Location (Pathology), Histologic Type, TNM Stage (Pathology)
Genetic Characteristics	K-ras, N-ras, BRAF
Treatment Characteristics	Postoperative Chemotherapy
Survival Characteristics	Overall Survival
Target Variables	
Chemotherapy	Postoperative Chemotherapy Regimen (5-FU/LV, XELODA, FOLFOX, FOLFIRI, Surveillance)

Data Preprocessing and Oversampling

Data Preprocessing

Data preprocessing is one of the processes that must be performed to obtain correct analysis results. If data preprocessing is not performed correctly, the relationship between the variables may be distorted; thus, accurate results may not be obtained[22]. Therefore, it is important to perform data preprocessing to generate a solid model. In this study, we focused on missing-value processing as well as categorical and continuous variable processing, prior to constructing deep-learning models.

First, if the missing-value ratio was determined to be >80% by checking the ratio of each variable, the variable was excluded, because sufficient data samples for training could not be obtained. Additionally, all the missing-value instances in the prediction target class “Post OP Chemo Regimen” were excluded. Continuous variables such as “Age,” “ASA,” and “CEA” have different ranges, and if training is performed without adjusting the ranges, overfitting may occur, and normal learning may not be able to proceed[23]. Therefore, the range of each variable was adjusted from –1 to 1 by applying the min-max normalization scaling method. In the case of categorical variables, the value was mostly composed of character data rather than numeric data; thus, it could not be automatically recognized and computed by the computer. Therefore, one-hot encoding was performed to vectorize each variable and represent 0 and 1. Fig 1 shows a data preprocessing process including a data oversampling process.

Data Oversampling

In most data generated in the real world, the proportions of the classes of the target variables have an imbalanced distribution[24, 25]. Data that have such a form are called “imbalanced data.” In particular, medical data generated in a clinical environment are severely imbalanced. Normally, we define a class with a relatively small proportion as a “Minor Class” and a class with a large proportion as a “Major Class.[26]” If model training is performed using imbalanced data, it is highly likely that the Minor Class will not be properly processed, and all data will be classified as Major Class[27]. Various methods, such as undersampling and oversampling, have been proposed to solve this problem. Undersampling involves adjusting the proportion of the class by removing some data of a Major Class, and oversampling involves proportioning the class by multiplying by data of a Minor Class. In general, when the amount of data is sufficient, undersampling is used. However, when the undersampling method is used, it is difficult to construct a normal learning model, because the dataset used in this study is not sufficiently secured. Therefore, in this study, we attempted to resolve the data imbalance by oversampling the data using the Bootstrap Resampling algorithm[28], which allows effective inference with a small amount of data. Oversampling data is only added to the minority class in train set to avoid affecting test performance. Fig 2 shows a bootstrapping based oversampling process.

Structure of Chemotherapy Recommender

To predict and recommend treatment methods, we developed a deep feed-forward neural network, called Colorectal Cancer Chemotherapy Recommender (C3R), which is the most basic implementation of the Deep Neural Network (DNN). The model is designed as a three-layer perceptron structure in the order of [Input Layer] – [Hidden Layer] – [Output Layer]. Detailed nodes composing each layer are designed as ([Input: 54] – [Hidden: 64] – [Hidden: 128] – [Hidden: 256] – [Hidden: 64] – [Output: 5]). We used Grid-

Search Algorithm to tune hyperparameters. Hyperparameter types and grid-search ranges are as follows. Layers ∈ {1, 2, 3, 4}, Batch size ∈ {32, 64, 128, 256}, Learning Rate ∈ {0.1, 0.01, 0.001, 0.0001}, Optimization Algorithm ∈ {Adam[29], Adadelta[30], RMSProp[31]}. The hyperparameters determined by grid-search algorithm are [Batch Size 64, Learning Rate 0.001, Optimization Algorithm: Adam Optimizer]. A dropout layer is added in the middle of each hidden layer to prevent overfitting. The ReLU[32] was used as the activation function of each layer except the output layer. Softmax[33] was used as the activation function of the output layer. The Softmax activation function calculates the input data and returns a probability value normalized to a value between 0 and 1; this can be expressed as follows:

See formula 1 in the supplementary files.

The returned probability value is defined as the Chemotherapy Recommendation Index, and according to this value, priority can be determined for suggesting an appropriate treatment method to the patient. Fig 3 shows the detailed structure of C3R.

Model Verification and Evaluation Method

To evaluate the performance of the proposed C3R model, we used confusion matrix. After, we compared the diagnosis concordance rate between the C3R model and the Gachon Colorectal Cancer Treatment Protocol (GCCTP) & NCCN guidelines to verify the validity of C3R. The comparative indicators used were Top-1 Accuracy and Top-2 Accuracy, because the treatment methods proposed in each guideline were broken down by priority. The recommendations of the C3R model are included in the Top-1 Accuracy if they are included in the preferred treatment method proposed by each guideline, and they are included in the Top-2 Accuracy if they are included in the next treatment. Fig 4 shows the model verification process including model performance evaluation.

Gachon Colorectal Cancer Treatment Protocol

First, we used the GCCTP, which is used by colorectal cancer specialists to determine treatment options for patients at the Gachon Gil Medical Center. The GCCTP is a rule-based treatment recommendation system based on empirical knowledge from many colorectal cancer specialists. It allows a colorectal cancer specialist to diagnose a patient's condition and determine treatment options according to information such as the patient's demographics, TNM stage, and risk factor. Fig 5 shows an example of a colorectal cancer treatment protocol for a case without metastasis.

NCCN Guidelines

The NCCN guidelines, which were published by experts from 28 cancer centers in the United States, reflect the opinions of experts and serve as guidelines for international cancer standards. They cover the diagnosis, treatment decisions, and treatments for 97% of cancers in the United States and are updated annually with new medical grounds to provide the optimal clinical guidelines for treating cancer patients. NCCN guidelines are divided into rectal and colon cancer guidelines. Version 2 of 2019 was used for verification of colon cancer and version 2 of 2018 was used for verification of rectal cancer [34, 35].

Performance Evaluation Metrics

Various comparison indicators were used to evaluate the performance of the C3R model. we used the confusion matrix to evaluate the performance of the C3R model. The confusion matrix, which is typically used for evaluating the performance of an algorithm[36], compares the actual results with the model prediction results in a table form, with four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP predicted True and TN as False, but TN as False but True as True, FP predicted as True, but when True was False, FN predicted as False, but the actual result was If true. These results can be used to generate various evaluation indicators, such as accuracy, sensitivity, specificity, precision, recall, F1-Score and area under the curve (AUC)[37]. In this study, we used the precision, recall, F1-score, and AUC, which can be used regardless of class imbalance.

Results

Results of Data Extraction & Data Preprocessing

The initial EMR dataset consisted of 143 variables and 1511 instances After chart review and data preprocessing, the dataset consisted 59 variables and 1,169 instances. 5 variables in whole 59 variables were target classes that were ultimately predicted and recommended in this study. They consisted of "5-FU/LV," "XELODA," "FOLFOX," "FOLFIRI," and "Surveillance". We divided the final dataset into a training set for learning and a test set for model verification (ratio of 8:2) to construct a deep-learning model. Table 2 shows detail process of chart review and data preprocessing. Details of continuous and categorical variables can be found in the S1 and S2 Tables.

Table 2. Dataset changes owing to chart review and data preprocessing

	Process	Variables (+Target Classes)	Patients(*N*)
	1st CRC Dataset	142 (+1)	1511
Chart Review	1) Check extraction method and location	142 (+1)	1508
	2) Check inappropriate data	142 (+1)	1496
	3) Select priority variables (1st Processed CRC Dataset)	40 (+1)	1496
Data Preprocessing	1) Drop redundant variables	37 (+1)	1496
	2) Drop including 90% missing-value variables	32 (+1)	1496
	3) Drop instance containing missing value	32 (+1)	1169
	4) One-hot encoding (Final CRC Dataset)	54 (+5)	1169
Data Split	1) Data split (training/testing)	54 (+5)	935 / 234

Results of Data Oversampling

Through the Bootstrap Resampling process, we created minor-class data ("XELODA" and "FOLFIRI") to secure enough data for training. The minor class oversampled the existing data by a factor of approximately 5, and the major class did not oversample the data. The newly created data were added only to the training set, to avoid affecting the test results. The number of data for each target class after the oversampling is presented in Table 3.

Table 3. Results of oversampling for minor class

Method	Total	5-FU/LV	XELODA	FOLFOX	FOLFIRI	Surveillance
Original	1169	398	42	323	35	371
After Oversampling	**1454**	398	**206**	323	**156**	371

To check whether the distribution of the generated data was similar to that of the existing data, the "Age" mean and standard deviation and "OS" mean and standard deviation were calculated for each anticancer treatment method, according to gender. A T-test was conducted to check whether there was a difference between the two groups for "XELODA" and "FOLFIRI," which performed data oversampling (Table 3). The mean age of the "XELODA" male group was 68.00 years, and the standard deviation was 9.90 years, indicating no significant difference (95% confidence interval: 63.9–72.1 years). The mean OS of the "FOFIRI" female group was 35.04, and the standard deviation was 33.64. In addition to the aforementioned examples, all the indicators of "XELODA" and "FOLFIRI" exhibited no difference before and after oversampling ($p > 0.05$). Details of T-test can be found in S3 Table. This indicates that the data generated using the Bootstrap resampling technique had a similar distribution to the existing data, and the newly generated data were added to the existing minor class to enable effective learning about the minor class.

Performance of C3R

As mentioned previously, we used the precision, recall, F1-score, and AUC indices of the confusion matrix to evaluate the performance of the proposed anticancer treatment recommendation model, i.e., C3R. The AUC values for all classes were >0.95, indicating that the developed model generally had good performance. In particular, Surveillance correctly predicted all patient cases with 100% accuracy. In the case of oversampling, i.e., XELODA and FOLFIRI, the precision values were 0.80 and 0.89, respectively. The overall performance of the model, including all classes, was generally good, with a precision of 0.92, recall of 0.98, F1-score of 0.95, and AUC of 0.98 (Table 4, Fig 6).

Table 4. Performance of the proposed model for each chemotherapy method

Class	Precision	Recall	F1-score	AUC
5-FU/LV	0.99	0.96	0.97	0.97
XELODA	0.80	1.00	0.89	0.99
FOLFOX	0.95	0.94	0.95	0.96
FOLFIRI	0.89	1.00	0.94	0.99
Surveillance	1.00	1.00	1.00	1.00
Total	0.92	0.98	0.95	0.98

Additionaly, to evaluate the performance of the C3R objectively, we compared the performance with machine learning algorithms. The C3R compared with SVM, Decision Tree, K-NN, and Random Forest. As a result, it was confirmed that Decision Tree shows the best performance except the proposed model. This was possible because Decision Tree is the best algorithm for classifying binary data. Table 5 shows performance of apply

Table 5. Performance of the proposed model for each machine learning algorithms

Method	Precision	Recall	F1-score	AUC
Proposed	**0.92**	**0.98**	**0.95**	**0.98**
SVM	0.80	0.90	0.89	0.85
Decision Tree	0.90	0.94	0.93	0.93
K-NN	0.82	0.83	0.80	0.82
Random Forest	0.91	0.93	0.92	0.92

Model Verification For a verification of C3R, we randomly extracted 200 data from a test set that was not involved in model training. However, 24 data included in the part where the GCCTP treatment protocol did not reflect the variables used in the C3R model were excluded from the comparison evaluation. Table 6 presents a comparison between the chemotherapy treatment methods recommended by C3R and those recommended by GCCTP & NCCN. XELODA and FOLFIRI, for which there were insufficient test samples, exhibited fluctuations in the Top-1 and Top-2 Accuracy.

Table 6. Comparison of the Top-1 and Top-2 Accuracy between the proposed model and the GCCTP & NCCN guidelines

Group		*N*	Top-1 Accuracy (%)	Top-2 Accuracy (%)
GCCTP	5-FU/LV	55	23.63	78.18
	XELODA	5	80.00	80.00
	FOLFOX	37	83.78	91.89
	FOLFIRI	4	0	75.00
	Surveillance	76	71.05	71.05
	Total	**176**	**57.95**	**77.84**
NCCN	5-FU/LV	59	47.45	83.05
	XELODA	6	50.00	100.00
	FOLFOX	50	92.00	94.00
	FOLFIRI	5	100.00	100.00
	Surveillance	80	80.00	80.00
	Total	**200**	**70.50**	**84.00**

Starting from the GCCTP treatment concordance rate, the Top-1 Accuracy was 57.95%, and the Top-2 Accuracy was 77.84%. In particular, in the cases of 5-FU/LV and FOLFIRI, the Top-1 Accuracy was 23.63% and 0%, respectively. However, in the cases of XELODA, FOLFOX, and Surveillance, the Top-1 Accuracy ranged from 70% to 80%, and the Top-2 Accuracy for FOLFOX was 91.89%, which was the highest treatment concordance rate among the chemotherapy methods.

The treatment concordance rate with the NCCN guidelines was higher than that of GCCTP, with a Top-1 Accuracy of 70.50% and a Top-2 Accuracy of 84%. Except for 5-FU/LV and XELODA, a treatment concordance rate of >80% was achieved for all chemotherapy treatment methods. In particular, FOLFIRI was limited to only five samples, but both the Top-1 and Top-2 Accuracy achieved a 100% treatment concordance rate.

Model Explanation

To explain why the C3R model recommends treatment options, we try to use the SHAP(SHapley Additive exPlanations) model. SHAP[38] is a game theoretic approach to explain the output of any machine learning model. It connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions.

The Fig. 8 shows variables each contributing to push the model output from the each variables to the model output. In general, pathologic variables such as TNM Stage and tumor location were found to have a significant effect on the model. In addition, demographics such as age, smoking history, and histologic type were also found to influence the results.

Discussion

In this study, we propose a DNN-based deep-learning model called C3R to provide chemotherapy recommendations for colorectal cancer patients after surgery. There is a limitation in this study that the model was built using specific data from a single institution. It may be said that this is generalized to the data collected at certain hospitals. However, in order to minimize this generalization problem and to ensure the scalability of the model developed in this study, we extracted data based on the 'Colorectal Cancer Data Dictionary' built through linkage with various domestic hospitals. 'Colorectal Cancer Data Dictionary' is a type of Common Data Model (CDM) of colorectal cancer, and was created to unify data variables and formats that occur in hospitals. Currently, it is constructed using only single hospital data, but it is also possible to reflect data collected from various hospitals in the future. In order to expand the model proposed in this study, a test will be conducted for patients with colorectal cancer from the Gachon Gil Medical Center.

The treatment concordance rates of the C3R model with the NCCN guidelines were 70.5% for the Top-1 Accuracy and 84% for the Top-2 Accuracy. This is an increase of >10% compared with the GCCTP, but Korea’s special medical insurance system hard to use of chemotherapy methods recommended by NCCN. In Korea, chemotherapy is proposed to patients on the basis of guidelines provided by the Health Insurance Review and Assessment Service (HIRA). If patients select a treatment that does not satisfy the guidelines of the HIRA, they are required to pay a substantial fee, because they will not be eligible for health insurance. Therefore, most patients select chemotherapy methods that satisfy the HIRA guidelines, which may have reduced the diagnosis agreement in the present study, because the treatment methods proposed by the C3R model and the NCCN guidelines may differ.

Katzman JL et al.[39] proposed DeepSurv. was briefly covered in the Model Explanation section in the process of identifying variables affecting the model. DeepSurv proposed a deep learning-based recommendation model based on patient survival data. The DeepSurv model requires accurate tracking to determine the patient's prognosis after chemotherapy. In order to determine the patient's exact prognosis, continuous follow-up, typically over 3 or 5 years, is required. However, realistically, it is not easy to follow the patient during that period. In addition, there are limitations in that if a patient dies within the follow-up period, it is difficult to determine the exact cause of death because several factors are involved in the cause of death. On the other hand, Our proposed method differs in that it has developed a model for recommending chemotherapy using objective data that can be extracted from patients.

In order to expand the model proposed in this study, tests will be conducted on colon cancer patients visiting Gachon Gil Hospital. During the test, we will analyze the match rate of chemotherapy with clinicians to confirm that we are supporting decision making properly. Afterwards, the model's performance can be further enhanced through a series of processes that expand the model to multiple connected hospitals to collect refined data. With more research at scale, it can be used by clinicians to select personalized treatment options.

Conclusions

The C3R model is a CDSS based on data generated at Gachon Gil Medical Center. It learns clinical cases of various patients to recommend treatment methods according to personalized treatment. The AUC of the C3R model was approximately 0.98, which indicated excellent performance with the EMR data, but the treatment concordance rate with the GCCTP & NCCN guidelines dC3Reased. The GCCTP—a treatment protocol established through the cooperation of several colorectal cancer specialists—has the limitation that all the data generated in the clinical environment are not reflected in the rule-based system. While >40 variables are used in the C3R model, the number of variables used in constructing the GCCTP was approximately 20, indicating that the treatment method is recommended.

The model compensates for the limitation that existing non-knowledge-based CDSSs do not reflect actual data. It is meaningful in that it is Korea's first colon cancer treatment method decision support system that reflects actual data. From a clinical viewpoint, even if a CDSS built with sufficient data outputs different results from existing treatment protocols and guidelines, the CDSS is based on a vast amount of data; thus, the results can be accepted as a new opinion in the treatment protocol rather than treated as an algorithmic error. In addition, a explainable CDSS can be built by providing deep learning models and deep learning interpretation model.

Declarations

Ethics approval and consent to participate

The study was approved by the institutional review board of hospital (IRB No. GCIRB2018-259). All data were fully anonymized before we accessed. 'Gachon Gil Medical Center' but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of 'Gachon Gil Medical Center'.

Consent for publication

Not applicable

Availability of data and material

'Gachon Gil Medical Center' but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of 'Gachon Gil Medical Center'

.

Competing interests

The authors declare that they have no competing interests

Funding

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (2018-2-00861, Intelligent SW Technology Development for Medical Data Analysis)

Authors' contributions

JHP conducted an overall study design and data analysis. JHB has made a significant contribution to the review and provision of EMR data. YHL and KYL were major contributor in methodologies validation and writing the manuscript.

Acknowledgements

Not applicable

Abbreviations

AI: Artificial intelligence; ASCO: American society of clinical oncology; AUC: Area under the curve; CDM: Common data model; CDSS: Clinical decision support system; CT: Computed tomography; C3R: Colorectal cancer chemotherapy recommender; DNN: Deep neural network; EMR: Electronic medical records; FN: False negative; FP: False positive; GCCTP: Gachon colorectal cancer treatment protocol; HIRA: Health insurance review and assessment service; MSKCC: Memorial sloan kettering cancer center; NCCN: National comprehensive cancer network; SHAP: Shapley additive explanations; TN: True negative; TP: True positive; WfO: Watson for oncology;

References

1. Sipherd R. The third-leading cause of death in us most doctors don’t want you to know about. CNBC. 2018. Feb 22.
2. Rodziewicz TL, Hipskind JE. Medical Error Prevention. [Updated 2019 May 5]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2019 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK499956/
3. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. Bmj, 2005;330(7494):765.
4. Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. Jama, 2005; 293(10): 1223-1238.
5. Berner ES, La Lande TJ. Overview of clinical decision support systems. In Clinical decision support systems. New York: Springer; 2007; 3-22.
6. Personal Information Protection Act. [Internet] Available from: http://www.pipc.go.kr/cmt/english/functions/infoCommunication.do

7. Somashekhar SP, Sepúlveda MJ, Puglielli S, Norden AD, Shortliffe EH, Rohit Kumar C, et al. Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. Annals of Oncology, 2018; 29(2): 418-23.
8. IBM Watson Hard At Work: New Breakthroughs Transform Quality Care for Patients [Internet] IBM News releases [cited 2017 July 20]. Available from: http://www-03.ibm.com/press/us/en/pressrelease/40335.wss?i=1360645029661
9. Gil Hospital adopts IBM WFO for the first time in South Korea. [Internet] IBM News Release (KR) [cited 2017 July 20]. Available from: http://www03.ibm.com/press/kr/ko/pressrelease/50591.wss#release
10. Koichi T, Kantarjian HM, Garcia-Manero G, Stevens RJ, Dinardo CD, Allen J, et al. MD Anderson's Oncology Expert Advisor powered by IBM Watson: a web-based cognitive clinical decision support tool. 2014 ASCO Annual Meeting; 2014 May 30-Jun 3; Chicago, IL, USA: J Clin Oncol; 2014. 32(5s)
11. Koichi T, Kantarjian HM, Garcia-Manero G, Stevens RJ, Dinardo CD, Allen J, et al. MD Anderson's Oncology Expert Advisor powered by IBM Watson: a web-based cognitive clinical decision support tool. 2014 ASCO Annual Meeting; 2014 May 30-Jun 3; Chicago, IL, USA: J Clin Oncol; 2014: 32(5s).
12. Lee WS, Ahn SM, Chung JW, Kim KO, Kwon KA, Kim Y, et al. Assessing Concordance With Watson for Oncology, a Cognitive Computing Decision Support System for Colon Cancer Treatment in Korea. JCO clinical cancer informatics. 2018;2:1-8.
13. Strickland E. IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum; 2019; 56(4): 24-31.
14. Zauderer MG, Gucalp A, Epstein AS, Seidman AD, Caroline A, Granovsky S, et al. Piloting IBM Watson Oncology within Memorial Sloan Kettering's regional network. 2014.
15. Rocha HAL, Dankwa-Mullan I, Juacaba SF, Willis V, Arriaga YE, Jackson GP, et. al. Shared-decision making in prostate cancer with clinical decision-support. 2019:e16576.
16. Krens L, Damhof M. Evaluation of a clinical decision support system (CDSS) to optimize cytotoxic drug dosing in cancer patients with renal insufficiency. 2016:1499.
17. Lee Y, Chae Y, Jeon S. Integration and Evaluation of Clinical Decision Support Systems for Diagnosis Idopathics Pulmonary Fibrosis (IPF). Healthcare informatics research. 2010;16(4):260-72.
18. Pyo KH, Ahn BC, Xin CF, Jung D, Kim CG, Hong MH, et al. A machine learning based clinical decision support system (CDSS) for anti-PD-1 therapy using non-invasive blood marker and clinical information for lung cancer patients. 2019:683.
19. Cha KH, Hadjiiski LM, Cohan RH, Chan HP, Caoili EM, Davenport M, et al. Diagnostic accuracy of CT for prediction of bladder cancer treatment response with and without computerized decision support. Academic radiology. 2019;26(9):1137-45.
20. Cancer occurrence and death. [Internet] Ministry of Health and Welfare (KR) [cited 2019 January 28]. Available from: http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2770
21. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RG, Barzi A. Colorectal cancer statistics, 2017. CA: a cancer journal for clinicians. 2017;67(3):177-93.

22. Kotsiantis SB, Kanellopoulos D, Pintelas PE. Data preprocessing for supervised leaning. International Journal of Computer Science. 2016;1(2):111-7.
23. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167 [Preprint]. 2015.
24. Poedjiastoeti W, Suebnukarn S. Application of Convolutional Neural Network in the Diagnosis of Jaw Tumors. Healthcare informatics research. 2018;24(3):236-41.
25. Wiharto W, Kusnanto H, Herianto H. Interpretation of clinical data based on C4. 5 algorithm for the diagnosis of coronary heart disease. Healthcare informatics research. 2016;22(3):186-95.
26. Chawla NV, Japkowicz NKotcz. Editorial: special issue on learning from imbalanced data sets. SIGKDD Exploration Newsletter. 2004;6(1):1-6.
27. Thanathamathee P, Lursinsap C. Handling imbalanced data sets with synthetic boundary data generation using bootstrap re-sampling and AdaBoost techniques. Pattern Recognition Letters. 2013;34(12):1339-47.
28. Efron B. Bootstrap methods: another look at the jackknife. In Breakthroughs in statistics. New York: Springer; 1992:569-93.
29. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980 [Preprint]. 2014.
30. Zeiler MD. Adadelta: an adaptive learning rate method. arXiv:1212.5701 [Preprint]. 2012
31. Tieleman T, Hinton G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural networks for machine learning. 2012;4(2):26-31.
32. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th international conference on machine learning (ICML-10); 2010.
33. Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press. 2016.
34. National Comprehensive Cancer Network. (2019). Colon cancer (version 2.2019). Available from: https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf
35. National Comprehensive Cancer Network. (2018). Rectal cancer (version 2.2018). Available from: https://www.nccn.org/professionals/physician_gls/pdf/rectal.pdf
36. Stehman SV. Selecting and interpreting measures of thematic classification accuracy. Remote sensing of Environment. 1997;62(1):77-89.
37. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2011.
38. Lundberg SM. Lee SI. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems. 2017:4765-74
39. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC medical research methodology, 2018;18(1):24.

Figures

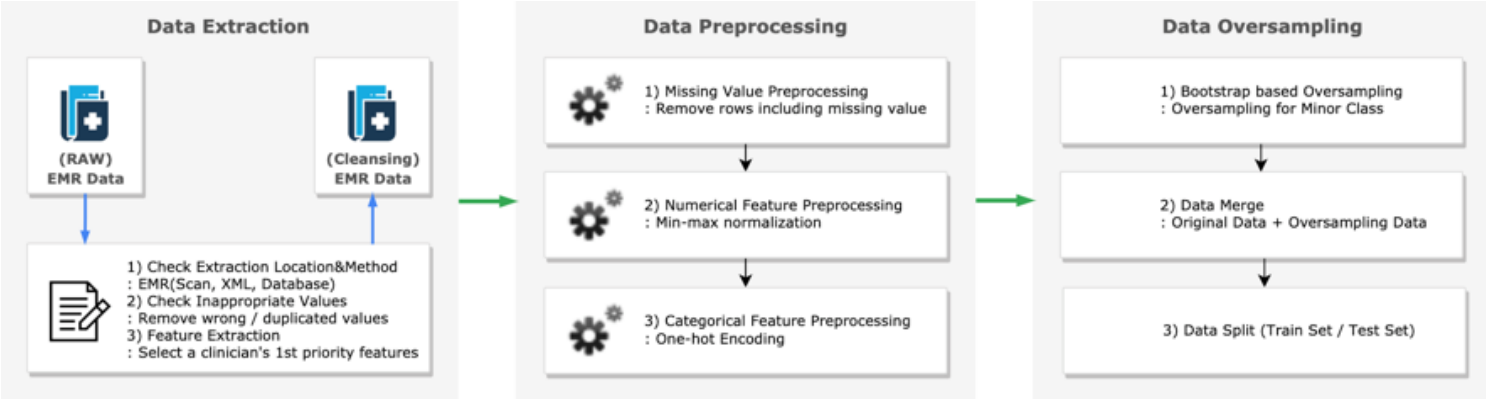

Figure 1

Process of data preprocessing and data oversampling

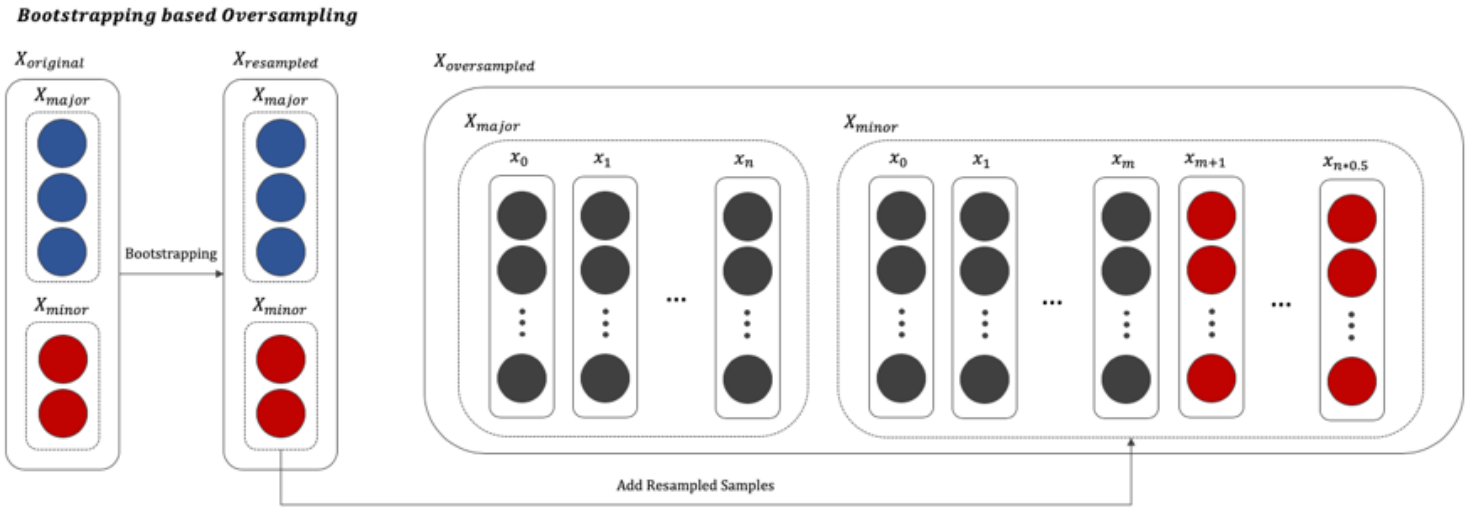

Figure 2

Process of oversampling algorithm based bootstrap (The blue node means majority class and the red node means minority class in target variable. Oversampling is limited to Minority Class. The oversampled data is added to the minority class.)

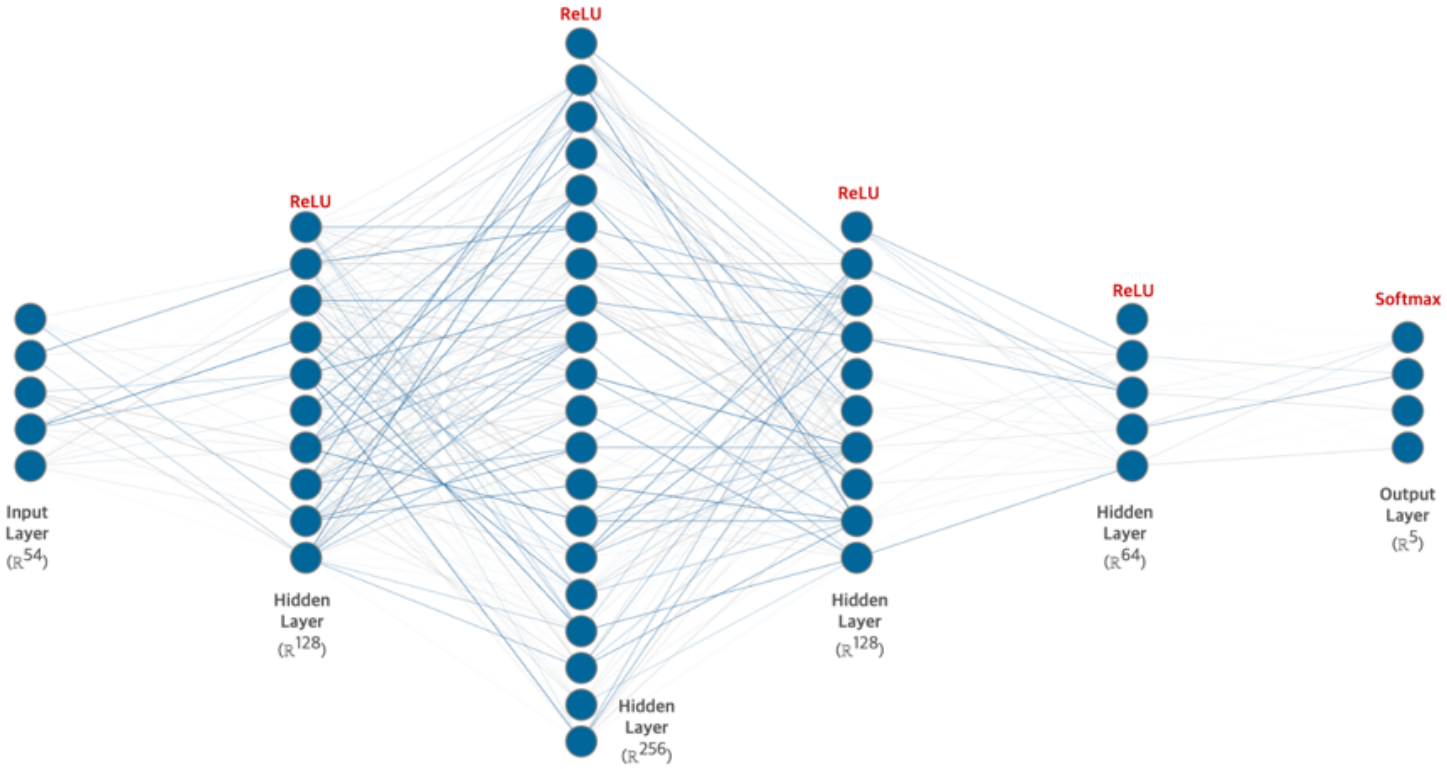

Figure 3

Structure of deep learning model for chemotherapy recommendation

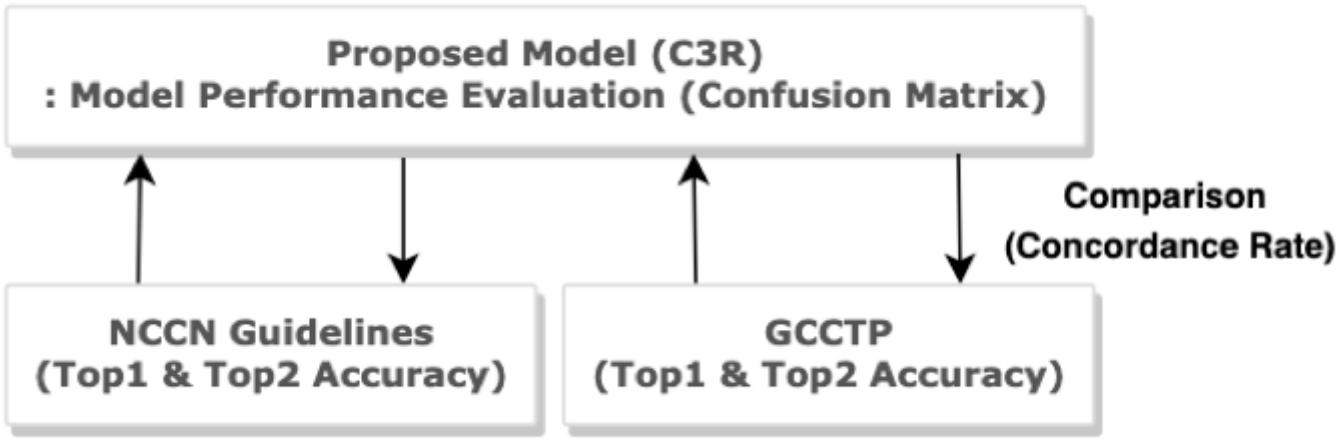

Figure 4

Process of model evaluation and verification

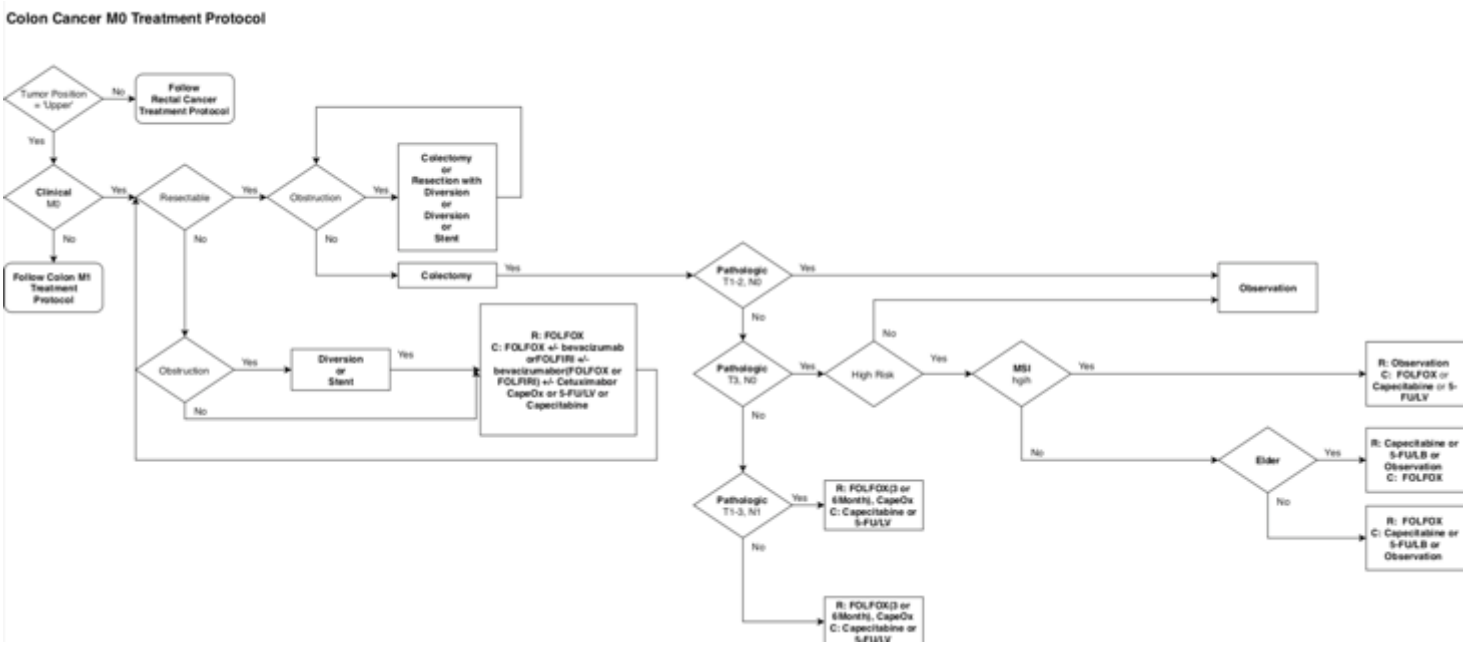

Figure 5

Example of a CRC treatment protocol: Colon Cancer M0 Treatment Protocol (Protocol is an algorithm used when administering chemotherapy to patients with colorectal cancer at Gachon Gil Medical Center. The protocol is largely divided into rectal cancer and colon cancer, and is divided into M0 case and M1 case without metastasis in detail.)

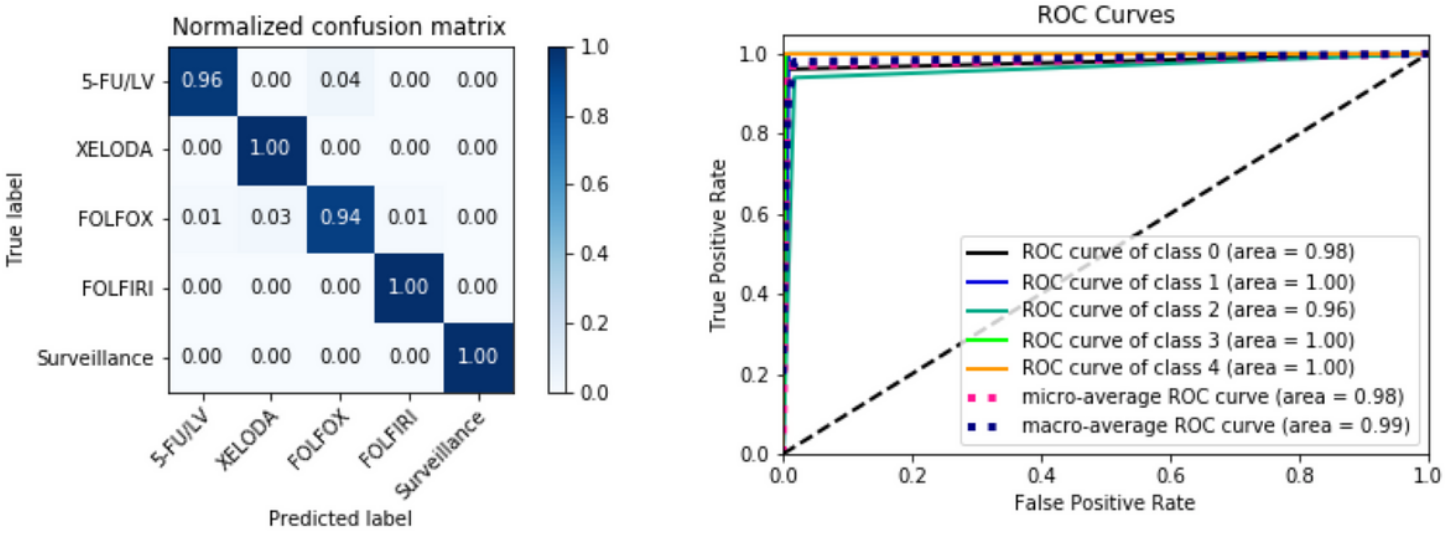

Figure 6

ROC curve and confusion matrix of the proposed model

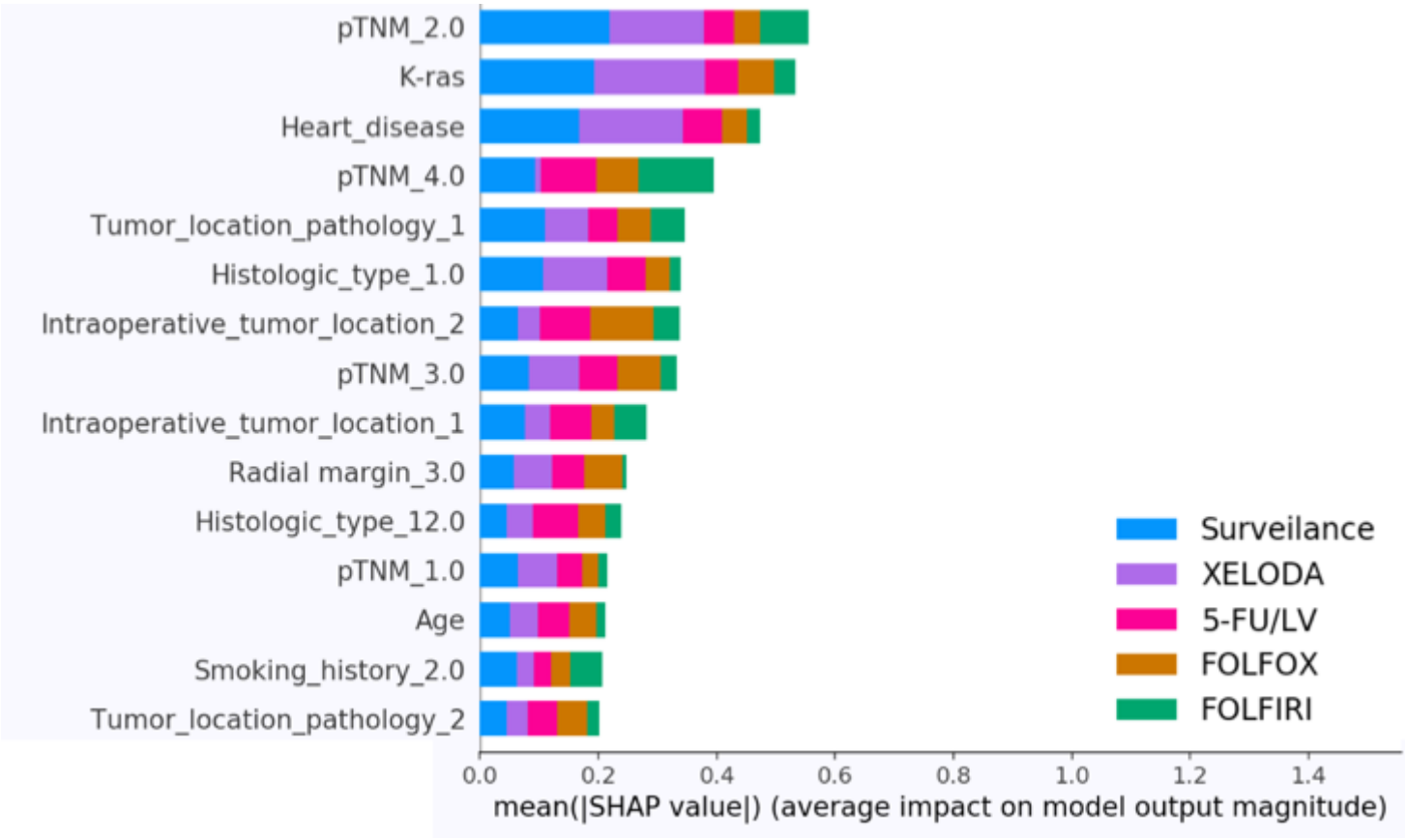

Figure 7

Overview of which variables are most important for a model

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingInformationFINAL.docx
- formula.docx