1 Patient Similarity Networks for Precision Medicine

1.1 Introduction

The increasing breadth and scope of patient-related data is redefining the traditional bounds of guideline-based patient evaluation and treatment. Precision medicine uses this new information landscape to better identify patient subgroups for prognosis and treatment. The availability of enhanced subgroups can contribute to better treatment decisions and outcomes for patients.

A central premise of precision medicine is that effective and transparent methods can be applied to partition patients into clinically relevant subgroups. A promising approach is to measure similarity between patients, since similar patients can reasonably be expected to have similar response to treatment, and similar outcomes. A natural way to do this is by quantitatively measuring similarity between data sets associated with patients.

We have developed a novel computational pipeline to assist clinicians in assigning the best therapy to individual patients, by integrating in a novel and transparent way multiple type of data from patients (including high dimensional genomics, EHR, treatment, imaging, etc.). The underlying method builds similarity clusters of patients: patient similarity networks (PSN) are an emerging paradigm for precision medicine, in which patients are clustered or classified based on their similarities in various features, including genomic profiles. We incorporated the netDx [5] computational framework, a novel approach to patient similarity networks that has been recently developed.

The target of this project is a clinical tool that will support prediction of clinical outcomes for colorectal cancer based on patient clinical and laboratory data. In clinical practice, decisions about alternative therapies for colorectal cancer patients are taken by the clinician on the basis of a quite broad and complex frame of data, which are weighted against each other in part with the support of guidelines and in part by experience. This complex integration of data and knowledge can benefit from a more structured algorithmic guidance.

We began this research with the following objectives:

- A BigMed dashboard tool based on patient similarity networks (PSNs) to help predict the clinical outcomes for colorectal cancer using patient clinical and laboratory data.

- An analysis of the performance of the PSN technique for this task, from both a clinical and technical perspective. This includes evaluating how well the PSN method fits into the domain workflow, and measuring how accurate and useful the results are.

- A roadmap for using the tool in different clinical diagnosis scenarios. We evaluate and document how data set properties affect performance (e.g., data set size, number and type of included variables), and develop a best-practices guide to select from alternative methods for structuring patient data for use with the tool.

- An analysis of our tool's performance compared to predictions made by an expert clinician, as well as predictions made by other computational methods.

- Implementing the method in a high-performance computing environment.

The remaining sections of this report describe our research and results, and indicate to what degree we met our starting goals. We also describe our evaluation of the patient similarity network approach, and what follow-on work would be appropriate to extend this project.

1.2 Introducing the Comet data and the classification task

Our project is focused on the OSLO-COMET data set, a patient database of clinical outcomes and data relating to colorectal cancer. We introduce the OSLO-COMET data set, and describe how we use this information to classify the outcomes of previously uncharacterized patients.

1.2.1 The Comet data

Liver is the most common site of colorectal cancer (CRC) metastases, and the only curative treatment is surgical intervention, like surgical resection [1]. The OSLO-COMET trial [2] studied outcomes of CRC patients who developed liver metastases, and underwent surgical resection.

One outcome of special interest is time to disease recurrence, because this variable can influence the use of preoperative chemotherapy and the sequence of treatment modalities. Our project evaluated a novel classification technique to predict the time to disease recurrence, based on a patient's clinical and genetic data.

Our goal is to develop a system that can predict the outcome of a patient with respect to disease recurrence – will the disease reappear within the first year after treatment (early, or short-term recurrence), or afterwards (late, or long-term recurrence)? We do this by using the information contained in the OSLO-COMET data set to train a *classifier*, which is a software program that can assign an uncategorized patient to one or more diagnostic or treatment categories of interest. We know, for each patient in the OSLO-COMET data set, whether recurrence is late or early. Using patient similarity networks, we can learn the patterns contained in the clinical and genetic data associated with each patient and so make a prediction for the new patient.

The OSLO-COMET data set contains 46 patients, all of whom have undergone liver resection for liver metastases. The time for to recurrence (or no recurrence) is known, and, as was noted above, the patients have been classified as **Long time to recurrence** (late) if disease recurrence was not observed 12 months after resection, and **Short time to recurrence (**early) otherwise.

Of the 46 OSLO-COMET patients, 28 were long-recurrence and 18 were short-recurrence.

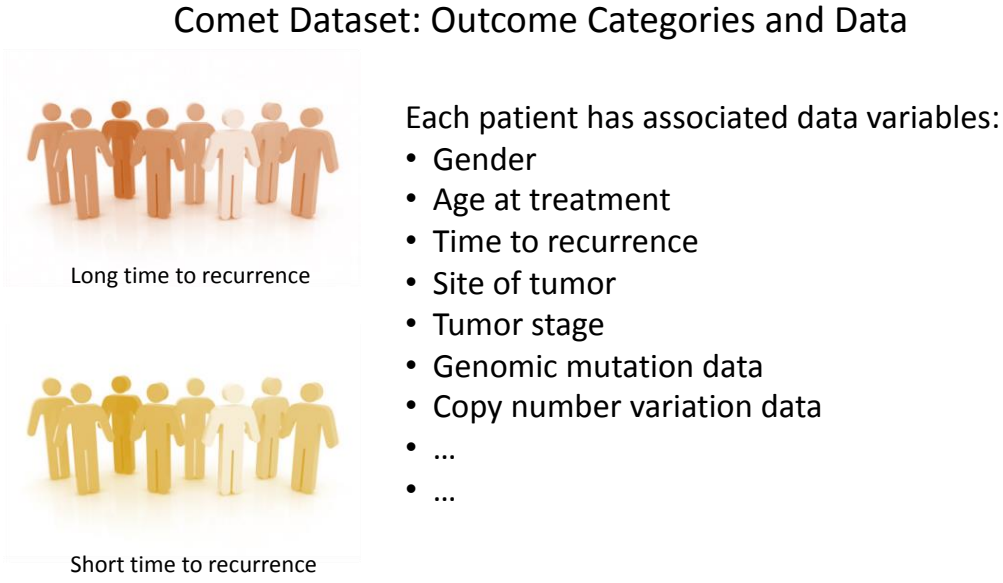
Each OSLO-COMET patient has associated clinical and genetic data which is relevant for the classification task. The following Comet variables are most relevant to the present work:

- Age at liver resection
- Site of primary tumor
- Tumor stage
- Lymph node stage
- Gender
- CEA level at liver resection
- ECOG (Performance status) at liver resection
- Chemotherapy prior to liver resection
- Mutations
- Copy Number Alterations

We now describe how this data, along with the associated outcome classification, can be used to create a classifier to predict the class of a new patient, based on that patient's data.

1.2.2 The classification task

We consider a patient data set, containing *variables* (also referred to as *features or covariates*) describing clinical, genetic, or other characteristics relating to the patients. The patients are also categorized according to a diagnosis or outcome of interest. For the OSLO-COMET data set, this category is categorized time to recurrence.

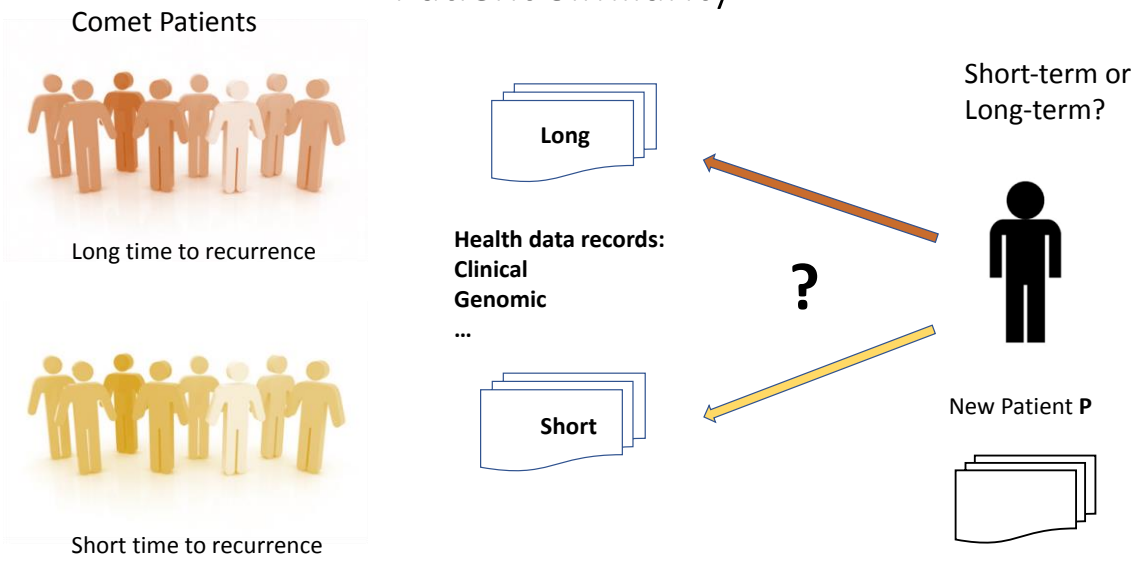


A central premise of precision medicine is that effective and transparent methods can be applied to partition patients into clinically relevant subgroups, such as length of time to recurrence of disease [3]. A natural way to do this is to discover patterns of similarity based on the clinical and genetic data associated with the patients, which correlate with the outcome category of interest.

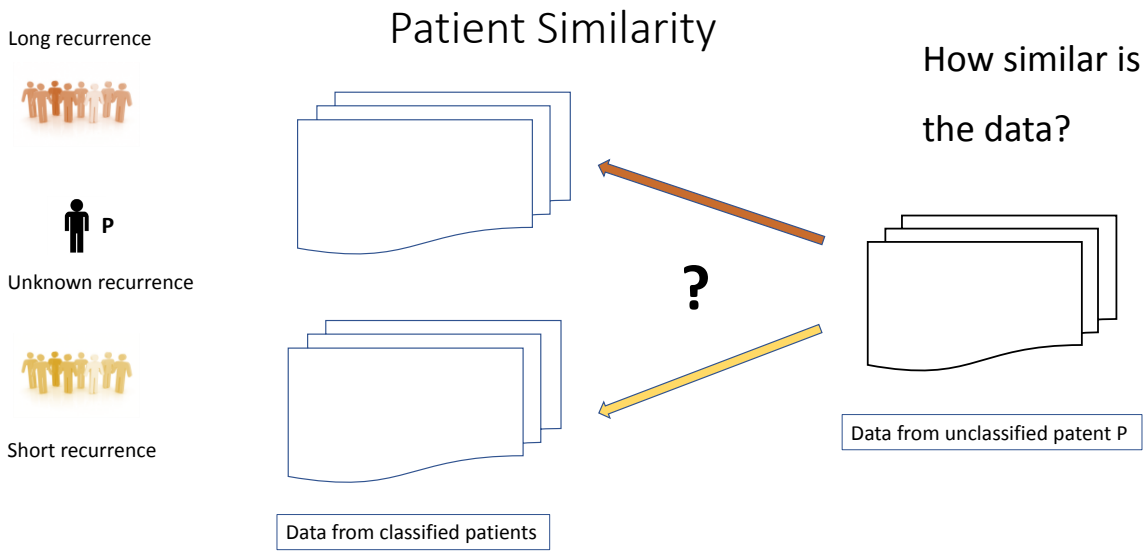
Patient Similarity

If we have these two groups of patients, categorized by outcome, and the associated patient data, we can now consider classifying a previously untreated patient P, where we have access to P's data. The task is to assign P to the most likely category (corresponding to the most likely clinical outcome). In our case, we wish to assign P to either the long-term or short-term to recurrence category.

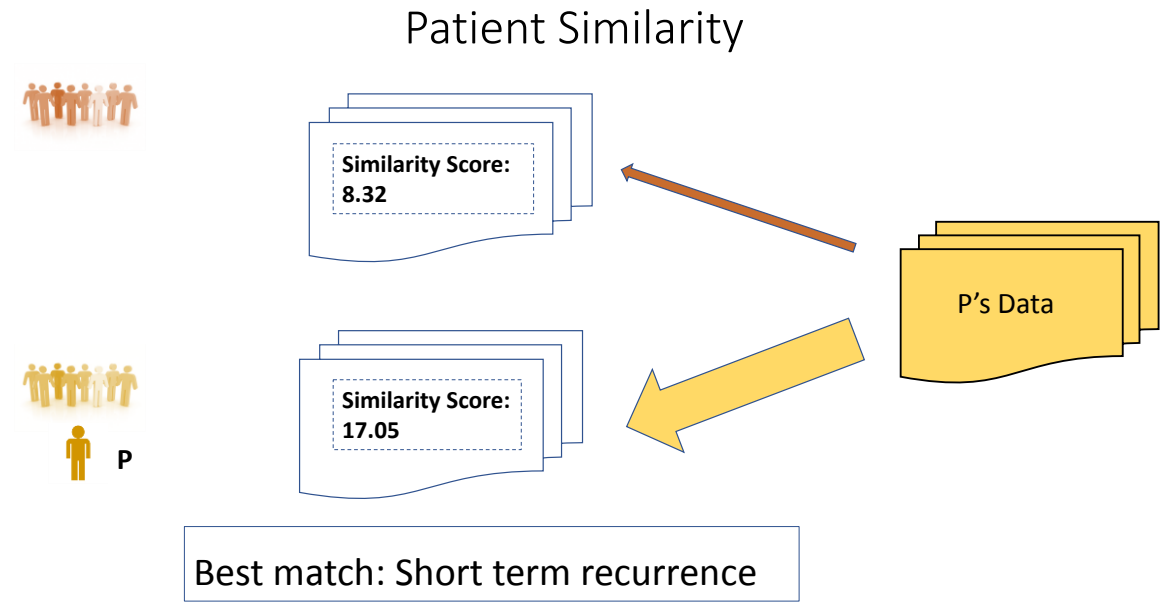
Patient Similarity



We do this classification using a recently developed technique called PSNs [4], which we describe in more detail below. Patient similarity assigns the new patient P to one of the two categories by sequentially measuring how similar P is to each category, and choosing the category to which P is closest. The measure of 'closeness' is based on the quantitative similarity of P's to the patients in each category.



The PSN method assigns a *similarity score* to each category, which summarizes the similarity between P and the patients in the category. P is assigned to the category that scores highest. Both scores are reported, which allows the confidence or uncertainty or strength of the categorization to be assessed by measuring the difference between the two scores.



1.3 How to classify with Patient Similarity Networks

We now present a stepwise description of how we address the classification task. We first introduce the netDx framework [5], a recently-developed software framework from the University of Toronto, which forms the basis for our system.

We then outline the steps involved in constructing a PSN model for classification, and then show how the model is used to predict the category of a new patient.

We then discuss the accuracy and stability of the model.

Finally, we present a user interface design for reporting the results to the user.

1.3.1 The netDx patient classification framework

netDx is a framework for building PSN-based patient classification systems, and it is the first to use the PSN approach [5]. netDx was designed to classify a patient of unknown diagnosis or outcome based on their similarity to patients for which this outcome has been established.

netDx is not a single user-ready program, but rather a *framework* of software components that is used to create a user-ready program. Our work has focused on developing and evaluating an end-user system for CRC outcome prediction, using the functions and procedures contained in the netDx framework.

As we have described, our outcome of interest has two categories – short term to recurrence, and long term to recurrence, corresponding to early and late recurrence. The netDx framework is capable of handling data sets having more than two categories. We present and discuss the classification procedure in the context of the two-category task, but the results and issues generalize to more complex multi-category cases.

There are many different methods for predicting clinical risk outcome, including logistic regression, random forests, support vector machines, deep neural networks, and others. Patient Similarity Networks (PSNs) adds a complementary approach to the existing techniques, and may offer advantages when handling specific domains and/or data sets.

1.3.2 Preliminary evaluation

The netDx papers discuss positive results obtained when applied to breast cancer and asthma, which we found relevant to our area of interest. Additionally, we carried out a pilot study of our own to evaluate netDx prior to applying it to the OSLO-COMET data. We replicated the results of a previous automated classification of data from the Oslo2 multicenter breast cancer study [6]. We found netDx performed with a high degree of accuracy when applied to the data and classifications from this paper.

1.4 Building a PSN-based patient classifier with netDx

We now outline the process of building a classifier for predicting time to recurrence. These steps are a general approach, which would also be used for other data types and classification tasks similar to this one.

1. **Build Similarity Networks**: A similarity network contains a quantitative measure of similarity between each pair of patients. This is done independently for each variable, resulting in one similarity network for each variable. Conceptually, similarity measures the 'distance' between the two patients, according to the variable. For example, when considering gender, a pair of male patients are closer than a male and female patient. The measure of similarity is described more fully below.

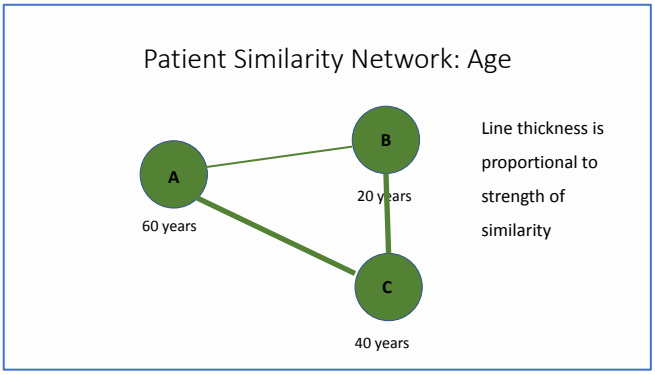
2. **Find the Most Informative Networks**: Next, netDx uses a train/test procedure is used to discover which of the variable-specific PSNs work best to classify with high accuracy. It is not expected that all of the patient data variables will have equal classification power, and this step eliminates variables found to be uninformative or which add noise or uncertainty to the process.

3. **Build the Integrated Network**: The subset of PSNs discovered in the previous step are integrated into a single PSN. The integrated similarity between a pair of patients is a weighted combination of the similarities in the composite networks.

The result is a classifier that can then be used to predict the outcome of new patients. In practice, the data set used to train the classifier might grow or change over time. It might be expected, then, that the classifier might need to be rebuilt using the updated data set, to reflect this new information.

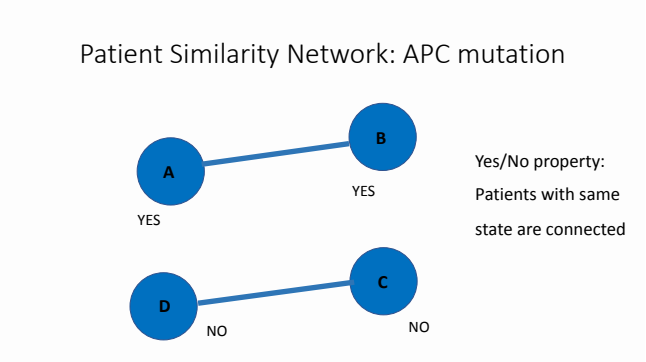
1.4.1 Building Similarity Networks

The PSN approach is conceptually intuitive, and supports straightforward methods to prepare data for processing. The fundamental concept is that, for each pair of patients (P and Q, say) one can measure and encode the similarity between P and Q. For example, we might measure how close P and Q are in age, or whether P and Q were diagnosed with tumors in the same physiological location. The similarity measure is embedded in a mathematical graph, which is a natural way to embed, measure, and process 'relatedness' between objects [7]. The following figure shows a simple graph showing similarity in age between patients.



Here, similarity is shown by the thickness of the connecting line between pairs of patients. The heavier the line, the more similar (closer in age) the two patients are.

Similarity networks are not limited to numeric or quantitative data. Categorical data can be represented by networks as well. The following similarity network shows the presence/absence of a mutation in the *APC* gene



This network contains two subnetworks, one for each category. Here, there are two categories, but there can be an arbitrary number. The *APC* similarity network as a whole consists of the set of subnetworks.

Networks are a good data structure for this task because they allow similarities based on different variables to be combined. This combining is a necessary step in in the process, as will be shown.

Note that, in the simplest case, each PSN represents a single data type or variable from the data set. This enables a clear and simple way to encode the desired information. However, the PSN approach does not require these simple encodings. Many data types can be encoded in PSNs in alternative ways. More refined methods of encoding the data can be used, which may enhance the power of the model.

As an example of a more structured encoding, consider a data set that contains mutation information for a large number of genes, scored as yes/no as was done for APC above. For various reasons, such as computational performance, or the risk of introducing noise or overfitting into the model, constructing one PSN per gene may not be desirable. A wiser alternative may be to reduce the number of variables by creating PSNs based on genetic pathways, with one PSN would be constructed per pathway. For each PSN, patient similarity would be scored by looking at the mutational impact on the pathway.

The choice of how to construct PSNs from the patient data is the most consequential part of building the system, and the one to which the most time and effort is devoted.

1.4.2 Finding the Most Informative Networks

The OSLO-COMET patient data consists of a large range of different variable types, drawn from a variety of sources. Some types are descriptive (age, gender), some describe treatment (chemotherapy/no chemotherapy), and others describe molecular genetic tumor state (mutations, copy number alterations). It is reasonable to expect that not all of the variables will contribute equally to discriminating between the long-term and short-term categories. netDx uses an iterative test procedure to find those variables – as encoded by the similarity networks – to rank the variable-specific similarity networks according to predictive power. This ranking is based on the performance of the networks in the test procedure, and is

reported as a numeric score. Networks that are above a user-defined score cutoff are used for further analysis.

It is important to note that this ranking and cutoff selection process is done independently for the two categories, resulting in a high-reliability network set for each category. Independent sets are used because each category may be best predicted by its own set of networks (corresponding to different variables). As a result, categories may have identical sets, the sets may overlap, or the sets might be disjoint.

1.4.3 Building the Integrated Network

Those PSNs passing feature selection are used to compute a single integrated PSN. Because, as just described, each category has its own set of high-performing variable PSNs, a separate integrated network is built for each category.

An integrated PSN is built by 'adding' together the variable PSNs. Since some of the variable PSNs have a higher importance, *weights* are assigned to each PSN before addition. Using weights gives each variable PSN more or less weight according to its importance. These weights are computed by netDx as part of the processing.

In practice, adding PSNs in this fashion reduces to adding together the similarity scores for each pair of patients.

As an example, suppose we had three variable similarity networks, 1,2, and 3. For every patient pair (P,Q), the integrated network will contain a similarity score PQ_{int} calculated from PQ_1 , PQ_2 , and PQ_3 , the PQ scores from networks 1, 2, and 3. The calculation would look like

$$PQ_{int} = ß_1PQ_1 + ß_2PQ_2 + ß_3PQ_3$$

Where $ß_1$, $ß_2$, and $ß_3$ are weights computed by netDx quantifying the relative importance of the PQ_i in the integrated similarity.

1.4.4 Classifying a New Patient

We now have two integrated PSNs, one optimized for the long-term to recurrence patient subset, and for the short-term to recurrence patient subset.

We classify the new patient P by using the integrated PSNs to find which of the two subsets most resembles P. This is done as follows.

1. P is inserted into each integrated PSN, and is similarity-linked to each patient in the PSN.
2. A computational method called *label propagation* finds the aggregated similarity between P and all of the patients in the PSN. This is done for each integrated PSN
3. The label propagation algorithm returns a score for each of the two PSNs. P is assigned to the PSN scoring highest.

1.5 Predicting time to recurrence: Results

We have described how we used the OSLO-COMET data set to build a classifier that assigns an unknown patient to the long-recurrence or short-recurrence categories. Two important stages of this process are finding good encodings of the data variables, and evaluating the performance of the classifier.

As discussed above, a central task of the patient similarity network approach to classification is to find a good encoding of the data variables as input PSNs to the model building process. We tested a large range of variable combinations, where we selected subsets of the data, and encoded them in PSNs in different. We found that the accuracy was best using the following Comet variables

- Age at liver resection
- Site of primary tumor
- Tumor stage
- Lymph node stage
- Gender
- CEA level at liver resection
- Chemotherapy prior to liver resection

We measured the effectiveness of the classifier by testing it on patients drawn from the OSLO-COMET data set itself. We did this by extracting patients from the data set before training, and using the extracted data as a test case. This technique is called *leave-one-out cross validation,* and is a standard method in statistical learning. Briefly, this method iteratively selects one patient as a 'test', builds a model with the remaining patients, and tests if the model can correctly classify the test patient. A patient is correctly classified if the model correctly labels it with the known class (long-recurrence or short-recurrence). The model accuracy is then defined as the percentage of the OSLO-COMET patients correctly classified by the model.

The cross-validation technique iteratively tested all 46 OSLO-COMET patients. The classifier correctly labeled 76% of the patients. Of these 35 correctly classified cases, 21 of the 28 long-recurrence patients were correctly classified (75%), and 14 of 18 short-recurrence patients were correctly classified (78%).

One unanticipated result of our work was that model performance was dependent on selecting a good subset of variables to use. We had expected that adding uninformative variables would not affect the outcome, since netDx selects the best variables to use. Adding extra variables appears to add noise, or make the modeling process subject to overfitting. This issue is worth further investigation.

We also feel that it would be informative to analyze the correctly classified (N=35) and incorrectly classified (N=11) patients, to discover what aspects or properties of these patient groups might explain the performance results. Having knowledge of what makes a patient 'difficult' to classify might aid in redesigning the classifier and improving the results.

The details of how the PSNs were created is described in section 1.11.

1.6 Classification accuracy compared to predictions by expert clinicians

One measure of the performance and practical utility of a classification system is how it compares with the performance of expert clinicians. We have designed a survey consisting of balanced subsets of the OSLO-COMET data, which will be sent to about fifteen clinicians. Each participant will be sent a set of data which is composed of cases from the OSLO-COMET data set that is balanced with respect to outcome (early or late recurrence) and model performance (correctly/incorrectly predicted). This report will be updated with the results and analysis from the survey.

1.7 User interface and reporting

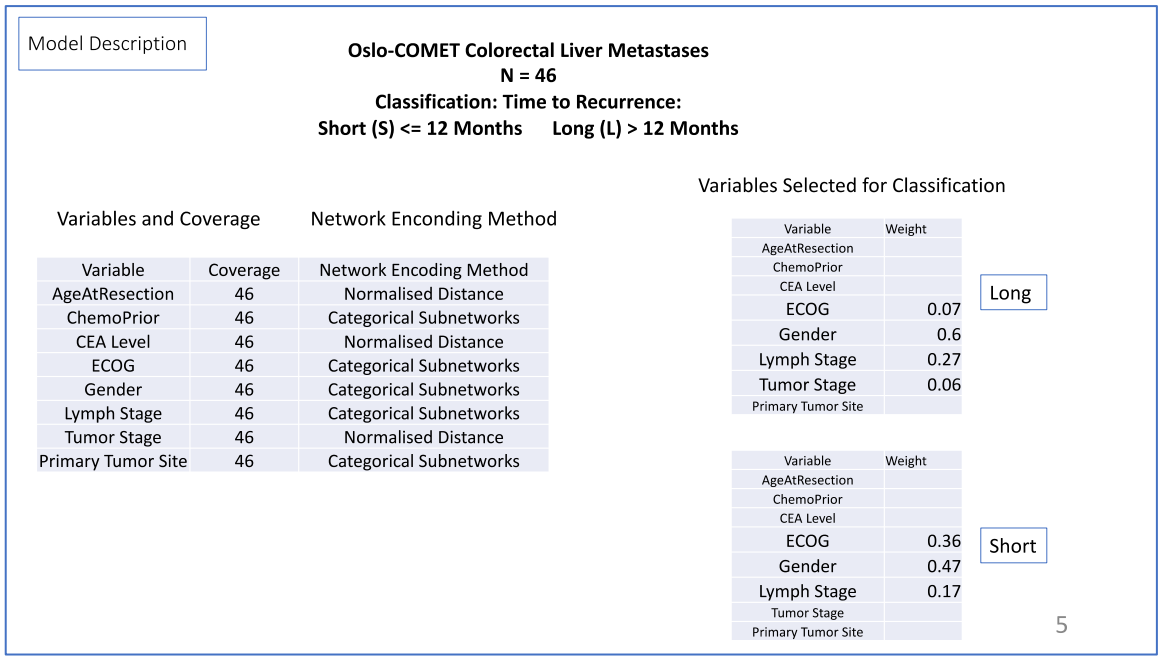
The preceding sections have shown how we build a classifier and evaluate its performance. We now discuss how the properties and results of the classification might be presented to a user. These designs could be the basis for a component in the dashboard for clinical decision support, or an independent tool.

We assume that users of the classifier will have different degrees of familiarity with the OSLO-COMET data set and the classification procedure. We can expect users who use the system infrequently or are only casually acquainted with the classifier and the underlying data.

1.7.1 Model description and metrics

The classification workflow results in a class prediction for a patient. Before this stage, though, the user must become familiar with the data and assumptions used to build the classifier. It is important to have a display or report that informs the user about the underlying design. This information is needed to provide context for the analysis.

1.7.2 . Model and data summary



A summary of the data set and classification categories is in the top heading. This is followed (on the left side) by a listing of the

- The formal name of the data set, and the number of patients it contains
- A description of the classification categories
- Data variables and coverage. Coverage indicates how many patients have values for this item. The OSLO-COMET data has full coverage – all patients have values for each variable. If this were not the case, the coverage value would be less than N, the number of patients in the data set.
- The **Network Encoding Method** shows the encoding method for each variable. There are a small number of common schemes that were used to encode data as PSNs. These encoding schemes guide how the similarity between a pair of patients (P, Q) is calculated. The encoding methods are discussed in more detail in section 1.11.
- **Variables Selected for Classification** indicates which highly informative PSNs were selected for the final model, as described earlier. The selected PSNs have a non-empty weight value, indicating the relative impact of that network in the integrated network. Recall that PSNs are selected for each category, so there are two lists shown.

1.8 Summary of model performance on the comet data.

Recall that the classifier is tested by sequentially holding out each patient from the data set, training a model on the remaining patients, and then using the reserved patient as a test case. This is done for each patient in the OSLO-COMET data set. A useful indication of how the classifier performs on patients of varying types can be found by examining the classification of each patient.

ID	True Class	Short Score	Long Score	Predicted Class
C1	S	1.00	0.61	S
C3	S	0.79	0.67	S
C4	L	0.54	0.79	L
C5	L	0.89	0.63	S
C8	S	0.79	0.28	S
C10	L	0.82	0.47	S
C11	L	0.93	0.63	S
C12	L	0.21	0.84	L
C13	L	0.57	0.89	L

ID	True Class	Short Score	Long Score	Predicted Class
C33	L	0.68	0.84	L
C34	L	0.71	0.84	L
C35	S	0.83	0.22	S
C36	L	0.75	0.95	L
C38	S	0.79	0.11	S
C40	L	0.25	0.95	L
C42	L	0.39	0.53	L
C43	L	0.29	0.95	L
C44	S	1.00	0.94	S

C14	S	0.03	0.78	L
C15	L	0.04	0.84	L
C16	S	0.83	0.72	S
C17	S	0.79	0.17	S
C18	L	0.61	0.95	L
C20	L	0.07	0.84	L
C23	S	1.00	0.89	S
C24	S	0.83	0.39	S
C26	S	1.00	0.94	S
C27	L	1.00	0.53	S
C28	L	0.64	0.95	L
C29	S	0.83	0.44	S
C30	L	0.36	0.53	L
C31	S	0.41	0.67	L
C32	L	0.96	0.68	S

C46	L	0.43	0.84	L
C47	L	0.86	0.58	S
C49	S	0.28	1.00	L
C51	L	0.32	0.84	L
C52	S	0.07	0.83	L
C53	L	0.11	0.84	L
C54	L	0.79	0.63	S
C57	S	0.79	0.28	S
C59	L	0.46	0.89	L
C61	L	0.14	0.84	L
C62	S	0.79	0.22	S
C65	L	0.50	0.89	L
C71	L	0.18	0.84	L

True Class is the known category of the patient, and **Predicted Class** is the model's prediction. **Short Score** and **Long Score** are the similarity scores that the model predicted for each category. The highest-scoring class is listed as predicted class.

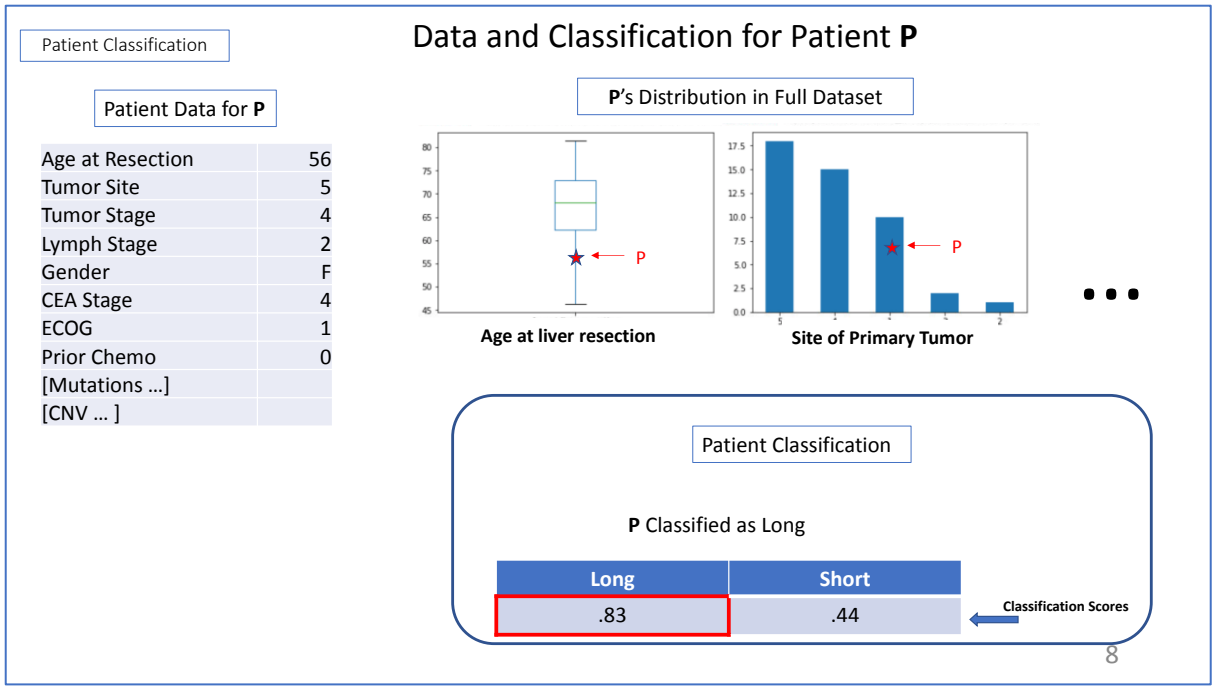
In addition to showing the predicted class, this table, with additional processing, can offer some further insights into classification performance. There is a slight difference in classification performance between short-term (78%) and long-term (75%) patients. One can also find patterns in the numeric difference between the long-term and short-term scores when considering false positives and false negatives.

1.9 Result of patient classification

The previous section gave an idea of how the model and data properties might be presented to the user. We now look at how classification results might be presented.

After reviewing the summary information describing the model and the OSLO-COMET data, the user can submit patient data for classification using a suitable interactive interface. The system will then perform classification, and report the results.

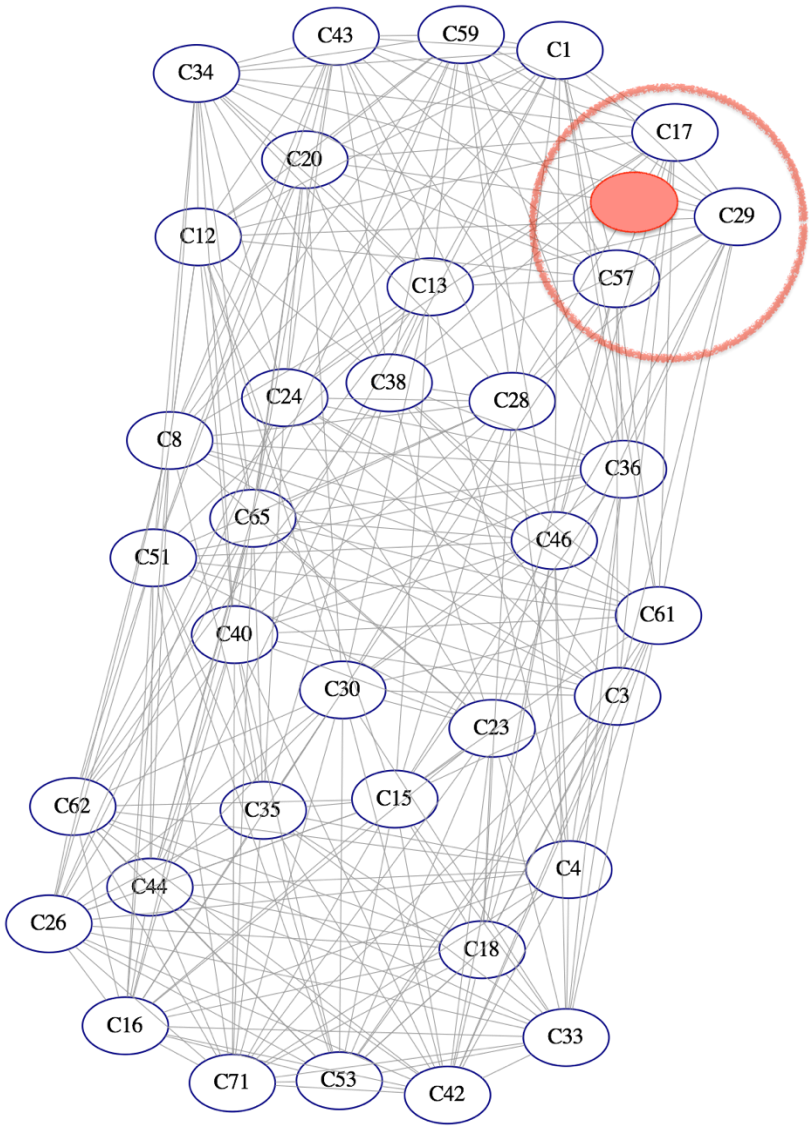
The results report shows information about the classification of a new patient P, and how the data associated with P compares with the OSLO-COMET data set.



- **Patient Data for P** summarizes the data associated with P
- **P's Distribution in Full Data set** summarizes where P's data variables fall in relation to the summary of the full data set.
- **Patient Classification** shows the relative scores for each category, with the assigned category highlighted. Scores for all the classes, and not just the top-scoring class, are presented, as this can provide a sense of how confident the classification is.

1.10 Insights into patient classification

The patient classification process has at its core the notion of similarity between patients. Classification is based on finding, for an unclassified patient P, those patients of known classification that P is most similar to. In a network context, these nearest neighbors of P can be seen as those patients with the strongest links to P.



The user will naturally be interested in examining the patients closest to P.
A report component will help analyze data associated with the K nearest neighbors to P.

1.11 Encoding the Comet data as patient similarity networks

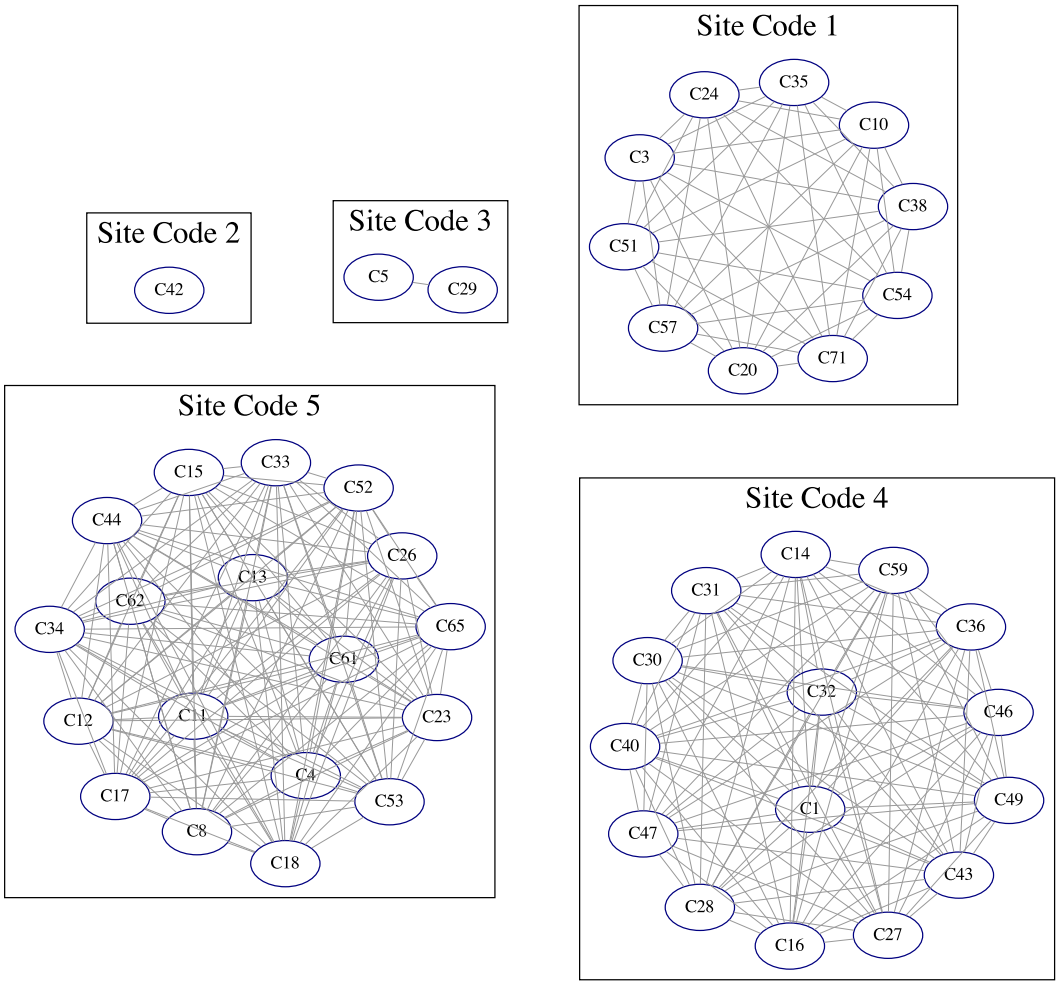
The OSLO-COMET data is encoded with two network types – categorical and normalized distance. These are described below.

1.11.1 Categorical data

Categorical data features assign patients to one of a number of distinct classes or types. Chemotherapy can be such a data type, for example – a patient has had chemotherapy, or not. Similarly, tumor localization can be expressed as a categorical feature, where the patient is given one of a small number of labels depending on the site of the tumor.

The category labels can partition the patients into groups, each group containing patients having that label. These partitions can be used to encode the variable as a PSN. This is done by recognizing that all patients in a partition share a similarity relationship, which is not shared by patients outside the partition.

The PSN is created by making a subnetwork from each partition. The patients in a partition are connected by links with identical similarity weight (1). Since there is no similarity (with respect to the variable) with patients outside the partition, there is no similarity between them. The end result is a set of subnetworks like this



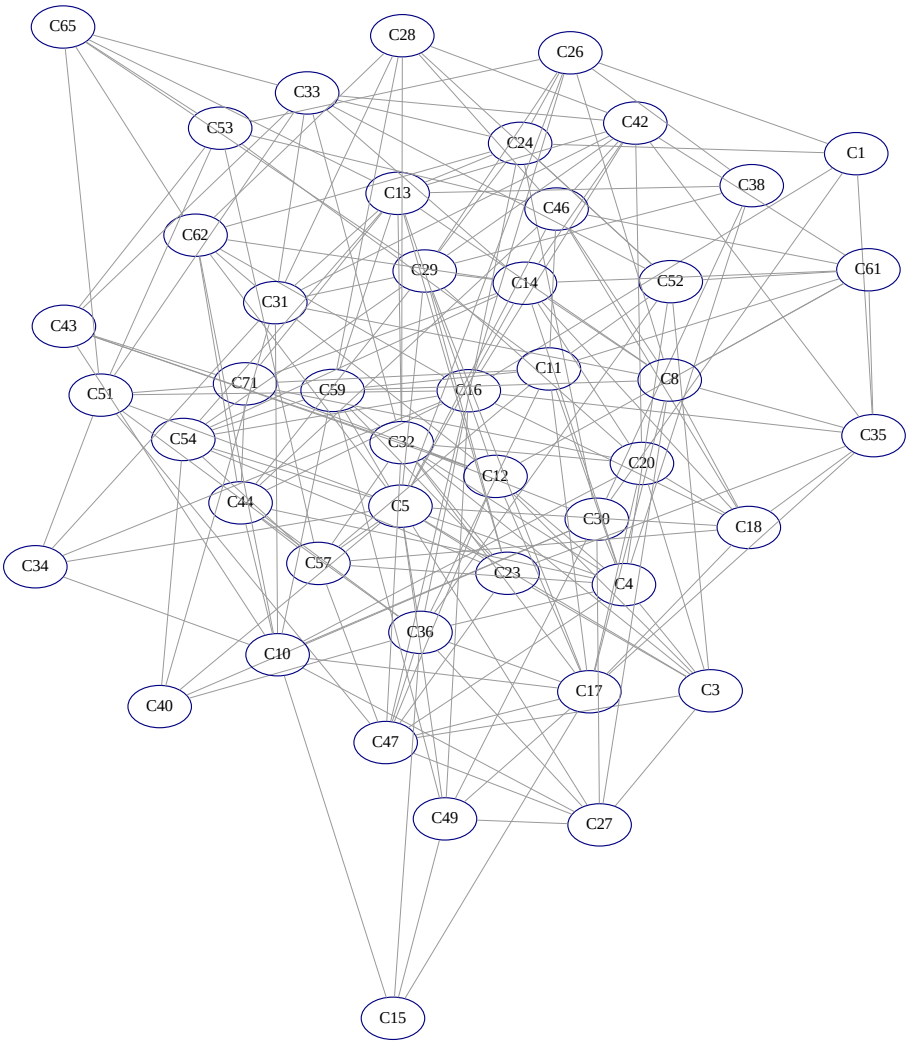
1.11.2 Numeric data

A natural way to compare numerical values is to take the difference. The similarity of two values is then proportional to the inverse of the distance. PSN similarity is measured on a normalized scale. Here, the similarity between any two items P,Q ranges from 0.0 to 1.0, and is calculated as

$$1 - \frac{dist_{P,Q}}{maxdist}$$

Where $dist_{P,Q}$ is the difference between P and Q (ignoring sign), and *maxdist* is the maximum value of $dist_{P,Q}$ in the data set.

The following diagram shows the patient similarity network for age at resection. Stronger similarity is shown by shorter links (some links have been hidden for clarity).



1.12 Evaluation of the PSN approach to patient classification

1.12.1 Modeling the data: choice of variables and similarity measures

We earlier presented a subset of 7 OSLO-COMET variables that we found most accurately classified patients. This subset was discovered by an iterative process of testing and

evaluating groups of variables, and we found this subset to be the most accurate of the variables we tested. It is not feasible to exhaustively test all combinations of OSLO-COMET variables, so we have no guarantee that the 7 chosen variables have the best performance. There may be some other combination that performs better.

This search for the best-performing subset of variables was by a large margin the most time- and labor-intensive part of the model construction process. A systematic, automated search for the optimal variable set would be a good area to for further investigation.

1.12.2 Stability and variance in the model

As described earlier, netDx selects highly informative networks for each classification category, and assigns different weights to them when constructing the integrated PSN. The internal algorithm introduces a stochastic element to this processing, which means that the classification outcome when considering a new patient can vary slightly in consecutive runs of the model.

We found that this variation is minimized in our final model consisting of our best-performing variable subset. It is still an aspect of the model-building process that needs to be measured, and presented to the user as a part of the reporting process.

We described earlier how we measured the accuracy of our model on the OSLO-COMET data by classifying each of the 46 patients in turn, and finding the number of patients correctly classified (35 out of 46). To account for the stochastic variation, we replicated this process 20 times. This gave us in 20 classifications for each of the 46 patients.

We examined the 20 classifications for each patient. We expected one of three outcomes: All 20 replicates were correctly classified, all 20 were incorrectly classified, or a mix of correct and incorrect classifications. We found that no patient fell into this last category. The 20 replicates were either all correct or all incorrect.

This is not to be expected in general. An alternative model, using a slightly different candidate feature subset, classified 4 patients with mixed results.

1.12.3 Need for high-performance computing

The iterative nature of testing candidate feature subsets for modeling, combined with the cross-validation accuracy testing and handling the stochastic variance, requires a significant amount of computing resources. We therefore developed a parallel pipeline, suitable for execution on the Colossus compute cluster hosted by the Services for Sensitive Data (TSD) at the University of Oslo. We found this a satisfactory solution, and believe it is suitable to incorporate into a user-facing pipeline.

1.13 Future tasks and directions

We have identified some areas which would make good follow up topics for exploring the potential of patient classification using patient similarity networks.

- The netDx framework performance can be modified or tuned by adjusting some parameters that affect the computations. Discovering how best to adjust these would boost performance and cut development time when building new models. Good candidate parameters for this are
    - PSN weights for network integration. User domain knowledge could be used to modify the numerical weights used to build the integrated network. Individual weights might be *a priori* increased or decreased, to adjust the influence of specific data variables.
    - Score cutoff for selecting informative PSNs. netDx selects PSNs for inclusion in the integrated network based on a numerical cutoff, which can be set by the user

- Explore relation between reported per-category similarity scores. netDx assigns a category to a new patient P by calculating P's similarity to patients in each category, and selecting the highest-scoring category. As with any predictive system, assignment accuracy is not 100%. One possible indication of performance might be the relative relation between the computed similarity scores.
    - Scores could have the same ratio, but different magnitudes. Consider the two categories short-term (S) and long-term (L). If test patient P scores 50 L, and 100 S, how does this compare in reliability to 5 L, and 10 S? The ratio is ½ in each case, but the absolute values of the scores differ.
    - Do lower magnitudes of scores imply less confidence, even if relative difference is the same? 20 S, 25 L differ by 5, as does 40 S, 45 L. Do the two pairs of scores imply the same confidence?
    - There are classification tasks in cancer studies where there are more than two classes, in contrast to the OSLO-COMET classifier, which has only two. In this case, what is the interpretation of the distance between the highest scoring category and the others? Is it significantly higher than the second highest, or is there only a modest difference between this selected category and the others? Would this observation bet useful for calculating the confidence that the classification is correct?

References

1. Østrup O, Dagenborg VJ, Rødland EA, Skarpeteig V, Silwal-Pandit L, Grzyb K, Berstad AE, Fretland ÅA, Mælandsmo GM, Børresen-Dale AL, Ree AH, Edwin B, Nygaard V, Flatmark K. Molecular signatures reflecting microenvironmental metabolism and chemotherapy-induced immunogenic cell death in colorectal liver metastases. Oncotarget. 2017 Jul 18;8(44):76290-76304.

2. Fretland ÅA, Dagenborg VJ, Bjørnelv GMW, Kazaryan AM, Kristiansen R, Fagerland MW, Hausken J, Tønnessen TI, Abildgaard A, Barkhatov L, Yaqub S, Røsok BI, Bjørnbeth BA, Andersen MH, Flatmark K, Aas E, Edwin B. Laparoscopic Versus Open Resection for

Colorectal Liver Metastases: The OSLO-COMET Randomized Controlled Trial. Ann Surg. 2018 Feb;267(2):199-207.

3. Parimbelli E, Marini S, Sacchi L, Bellazzi R. Patient similarity for precision medicine: A systematic review. J Biomed Inform. 2018 Jul;83:87-96.

4. Pai S, Bader GD. Patient Similarity Networks for Precision Medicine. J Mol Biol. 2018 Sep 14;430(18 Pt A):2924-2938.

5. Pai S, Hui S, Isserlin R, Shah MA, Kaka H, Bader GD. netDx: interpretable patient classification using integrated patient similarity networks. Mol Syst Biol. 2019 Mar 14;15(3):e8497.

6. Aure MR, Vitelli V, Jernström S, Kumar S, Krohn M, Due EU, Haukaas TH, Leivonen SK, Vollan HK, Lüders T, Rødland E, Vaske CJ, Zhao W, Møller EK, Nord S, Giskeødegård GF, Bathen TF, Caldas C, Tramm T, Alsner J, Overgaard J, Geisler J, Bukholm IR, Naume B, Schlichting E, Sauer T, Mills GB, Kåresen R, Mælandsmo GM, Lingjærde OC, Frigessi A, Kristensen VN, Børresen-Dale AL, Sahlberg KK; OSBREAC. Integrative clustering reveals a novel split in the luminal A subtype of breast cancer with impact on outcome. Breast Cancer Res. 2017 Mar 29;19(1):44.

7. https://en.wikipedia.org/wiki/Graph_(discrete_mathematics)