Exploring variability of machine learning methods: first steps towards cancer biomarkers consensus signatures

Elsa Claude, Mickaël Leclercq, Arnaud Droit, Patricia Thebault, Raluca Uricaru

Exploring variability of machine learning methods: first steps towards cancer biomarkers consensus signatures

Elsa Claude[1,2], Mickaël Leclercq[1], Arnaud Droit[1], Patricia Thébault[2,⋆] and Raluca Uricaru[2,⋆]

[1] Centre de Recherche du CHU de Québec - Université Laval, Québec, Québec,G1V4G2, Canada
[2] CNRS, Bordeaux INP, LaBRI, UMR 5800, Univ. Bordeaux, 33400, Talence, France

Corresponding author: elsa.claude@u-bordeaux.fr

Abstract *Nowadays, the biomedical field takes advantage of computer science and biotechnologies development to look for potential new strategies in the fight against complex diseases like cancer. In cancer stage prediction and prognosis, researchers can look for biomarkers using machine learning (ML) approaches able to perform pattern recognition. Unfortunately, even with an expertise in ML field, it is difficult to know which algorithm will perform best on a specific type of data. ML-based strategies have multiple steps that can be difficult to set up as it implies the tuning of numerous parameters at every step. Thus, ML based studies usually focus on a unique model that is expected to fit their specific research question. However, based on such strategy, variations in the methods may lead to completely different results and model generalization is limited. In this study, we make a first contribution towards a large scale analysis meant to understand the behavior pattern of ML methods in the context of biomarker signature identification from omics cancer data. We present preliminary results of colorectal cancer stage prediction based on the intra- and inter-group comparisons of two types of ML methods: Bayes-based and Trees. We first estimate the robustness of various Bayes and Tree based models with respect to the tuning of their parameters. We then analyze the composition of the produced signatures in order to assess their level of confidence, by looking for consistent features between models. Preliminary results suggest that Bayes-based algorithms are promising as their performances and signatures seem to be consistent across various configurations.*

Keywords Machine Learning, Biomarker signature, Cancer stage prediction

1 Introduction

Significant advances have been made in medical research with the development of new methods to collect data at various biological levels, also known as omics (gen-, epigen-, transcript- etc.). Analyzing these types of data has become a recurrent challenge, as for addressing the complexity in diseases one needs to take advantage of the joint integration of various omics. In the last decade, machine learning (ML) methods have been widely applied on omics data, in both single- and multi-omics contexts, to identify potential biomarkers for cancer stage characterization (genes, metabolites etc.) or patient prognosis (microsatellite instability status, long non-coding RNAs etc.) [1, 2, 3, 4]. Overall, ML based methods are able to recognize patterns between classes of samples by going through large amounts of data but they come with a counterpart : the difficulty to choose between a wide panel of ML algorithms and parameter configurations. Unfortunately there are no clear guidelines regarding the strategy best suited to a specific dataset in this context. In addition to data pre-processing, multiple factors in ML training may impact the outcome of a ML-based analysis, such as the tuning of parameters, the feature selection strategy, the choice of the performance evaluation method, to name only some of them. This renders the use of such techniques, as well as the interpretation of the results highly complex.

In this context, some research studies have assessed the sensitivity to parameters changes using artificial, ecological or transcriptomics data [5, 6, 7]. Moreover, the variety of metrics allowing to evaluate ML-methods performance may have a high impact on ML model choices [8]. Multiple metrics have been employed over the years, such as the widely used accuracy ACC measure, the area under the curve AUC, but also new formulas like the Matthew's correlation coefficient MCC, which are giving

⋆. Contributed equally

new nuances to our way to assess models [9]. Interestingly, new studies have put in common multiple ML models to benefit from their various strengths to recognize patterns [10, 11], also allowing to highlight strong impact features.

Despite this, few large scale studies [7] were conducted to evaluate the impact of different training configurations (parameters, feature selection etc.) on ML-based strategies, for omics data analysis in cancer. In [5], it was shown that default parameters can provide in many cases the best accuracy score, while randomization of parameters may positively or negatively impact the performance relatively to the default set up. The authors suggest that testing several random configurations could be a strategy to eventually reach the highest accuracy. However, to our knowledge, no study explored the effect of such variations on the biomarker identification for cancer prediction and prognosis. Indeed, studies usually focus on global performance comparisons between some commonly used methods in the ML field, for a small number of configurations (most of them by default), without further examining the robustness of their output signatures. Note that a biomarker signature in ML context is a set of variables (*i.e.* features) such as genes or lncRNAs (depending on the input data) that is used to build a ML model and that it uses to differentiate groups in a dataset.

In this work we intend to address the lack of research on ML signatures variability, by implementing a proof-of-concept comparison strategy that focuses on the problem of biomarkers signature identification from transcriptomics data, for cancer prediction. We benefit from the training of several supervised ML algorithms with multiple configurations in order to evaluate their robustness with respect to their parameters, based on quantitative performance metrics, as well as on the relevance of their output signatures. Indeed, examining the signatures produced by different models could highlight interesting new targets that would have been overlooked if investigations were led using an unique model.

2 Materials and Methods

2.1 Data collection and processing

Here we use colorectal cancer data obtained from the GDC portal using the TCGAbiolinks R package [12]. We retrieved transcriptomics RNA-Seq data from the TCGA-COAD cohort [13]. In this study we use raw count data which have been filtered and normalized (removal of insufficient RNA counts through samples, library size and batch effect control) [14, 15].

The dataset includes 41 normal samples and 63 stage IV samples composed of counts of about 20 000 RNAs with their corresponding Ensembl IDs. Normal samples are identified as Solid Tissue Normal and Stage IV (IVa, IVb, IVc) have been gathered as Stage IV samples for classification, based on the American Joint Committee on Cancer (AJCC) stage labels. Samples with no clinical data and disease samples with less than 70% of tumor nuclei have been discarded.

2.2 Machine Learning strategy

Within the biomarker prediction context, where signature genes correspond to features selected by ML methods, we intend to study the impact of different ML models. Our comparison strategy, described below, is based on BioDiscML software [16], in order to benefit from its unified framework that allows sampling, parallel training of thousands of machine learning models and validation by numerous cross-validation and resampling steps.

2.2.1 Pre-processing stage In an attempt to make profit of the variability in the healthy and cancer datasets and to better reflect their intrinsic composition, here we implement a stratified sampling strategy, instead of the classical sampling step of ML and implemented in BioDiscML. We performed a hierarchical clustering analysis of our samples using the transcriptomics data and using the Ward's criterion and the euclidian distance. It revealed 5 main clusters in the healthy category. A cluster of size 1 was considered to be an outlier and removed from our study. Among the cancer samples, 6 clusters were retrieved. Based on this grouping of samples, we then applied a stratified sampling procedure to generate 3 representative samplings (2/3 of the dataset for training, and 1/3 for test purposes).

Moreover, to overcome the classical issue of the curse of dimensionality, when the number of features (around 20000 RNAs) is largely superior to the number of samples, a set of features was selected based on their predictive power (through information gain ranking). For our dataset and for each sampling around 62% of the features were removed.

2.2.2 Training process ML-methods can be organized into two main categories : supervised and unsupervised learning methods. The former use labeled classes to train a model in order to highlight patterns among the classes, while the latter are left to discover inherent grouping in the data. As the scientific question we address in this work concerns a classification problem (normal versus stage IV colorectal cancer), to develop this proof of concept, we selected ML algorithms widely used in biomedical studies [17] from two groups of supervised ML-methods, Bayes-based and Tree-based. Among the Bayes probabilistic classifiers, here we focus on Naive Bayes [18], Bayesian Network [19] and Averaged 1-Dependence Estimators (A1DE) [20]. The first two are often applied on biological data while the third was developed to address the attribute-independence issue of Naive Bayes' method. Among tree classifiers, which are based on decision trees [21], some are commonly used in a clinical context, such as C4.5 and Random Forest, and were selected in our study. Additionally, two less common tree methods were selected: Naive Bayes Tree (NB Tree) [22], and Simple Classification And Regression Trees (Simple CART) [21].

The training of the different methods was made with BioDiscML software [16] under multiple configurations. BioDiscML performs an iterative training process to select a signature (feature subset) that optimizes the global performance of our ML models (see below for the performance evaluation metrics used in the training process). For each iteration, a set of features is selected with stepwise methods, here the Forward stepwise selection and Backward stepwise elimination (FB) strategy that implies that features with the highest information gain score will be integrated in priority. More details on the iterative training can be found in [16].

2.2.3 Evaluation Metrics The feature selection step in the training process involves an optimization procedure based on an evaluation of the models after each iteration. Here the Matt

hews Correlation Coefficient (MCC) is used, as it is known to be a good compromise, with respect to the ACC or the F-1 score, when evaluating ML model performance in presence of unbalanced data and more specifically in binary classification [9].

The models generated in the training step were further evaluated in order to assess their potential overfitting to the input data. This was done with several cross-validation procedures on data including the test set (10 cross-validation, repeated holdout and bootstrapping [23]). For each model, the average MCC (AVG MCC), as well as the associated standard deviation MCC (STD MCC) were then computed using the MCC scores obtained from the resampling techniques. A MCC value close to 1 along with a STD MCC close to 0 indicate an efficient and robust model.

3 Results

3.1 Model comparisons

The 7 classifiers cited above (Naive Bayes, Bayesian Network, A1DE, C4.5, Random Forest, NB Tree and Simple CART) have been trained with numerous parameter values, including the default ones, on the 3 stratified samplings of our dataset. The various parameters available for each classifier are binary, discrete, continuous or to pick from a determined list and can be inter-dependant. Parameters were randomly tuned, thus leading to the dropout of training of models in case of parameter values not adapted to the data, or in case of impossibility to compute the MCC score.

For each stratified samplings (as a proof of concept, we used in this study k=3), various models were trained as following. In the Bayes group of classifiers, Naive Bayes having 2 parameters led to the training of only 3 models (in each sampling). The Bayes Network method has 4 parameters but few are appropriate for our data so an average of 9 models were trained. The Averaged 1-Dependence Estimators (A1DE) on the other hand, has 4 parameters to tune thus leading to an average of 157 A1DE models being trained. In the Tree class, NB Tree had no parameters so a unique model was

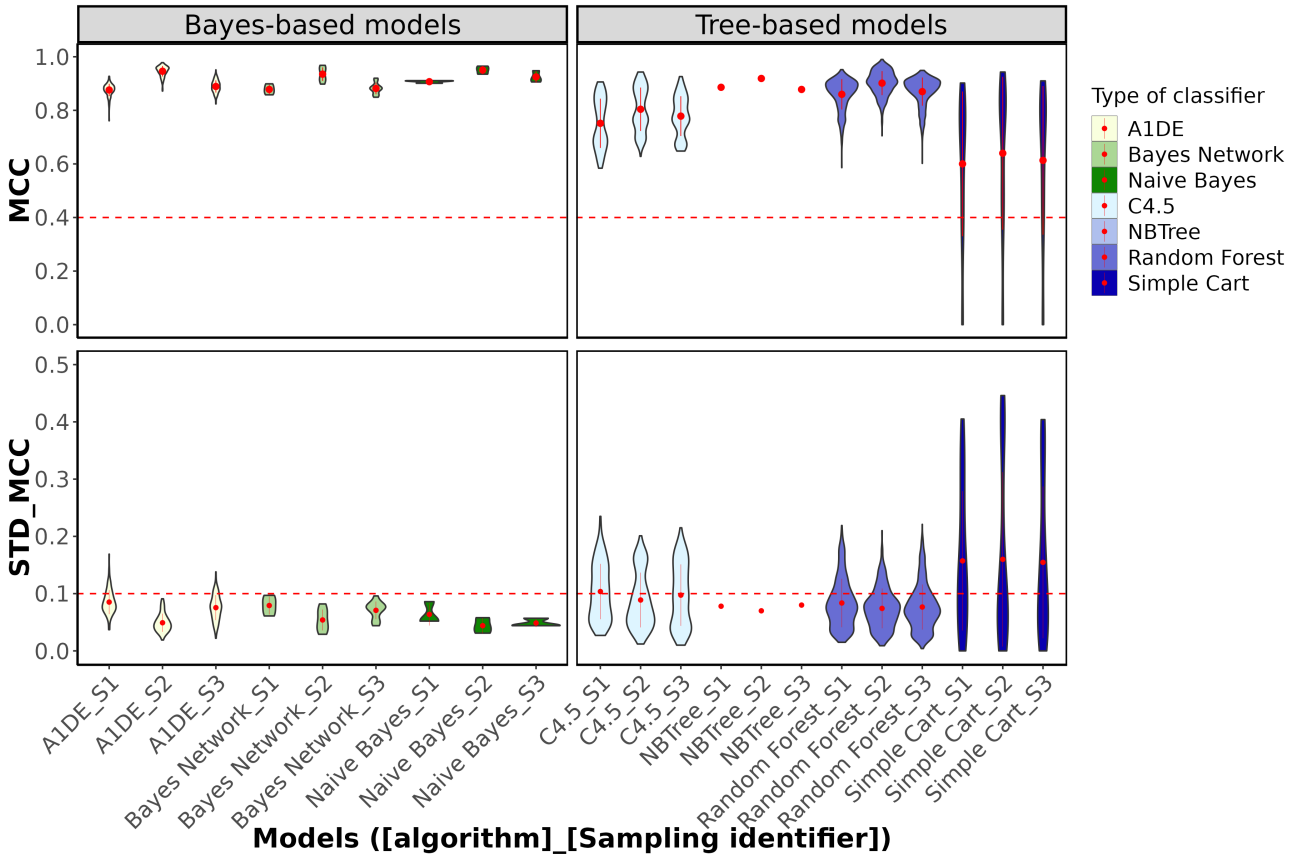

Fig. 1. Performance metrics distribution among the k=3 samplings for Bayes and Tree models trained with various parameter configurations. MMC = Matthew's Correlation Coeff., STD MCC = Standard Deviation of the MCC. Red dashed line indicates MCC and STD MCC filter.

trained, while Random Forest had 8 parameters allowing the training of 900 models. Finally, given the 4 Simple CART parameters and the 6 of C4.5, it resulted in the training of 60 and 250 models respectively.

3.1.1 Evaluation by performance metrics We evaluated the robustness of our models with performance metrics described in the 2.2.3 section, namely the MCC and STD MCC. Figure 1 shows the variation of MCC and STD MCC values among the different configurations of the classifiers on the 3 different samplings.

Both Bayes-based and Tree-based models were found to be consistent across the samplings as their MCC and STD MCC values do not seem to vary from one sampling to another. However, when looking at the dispersion of MCC and STD MCC values between models produced for a given method among the 7, Bayes-based models were found to be more robust regarding the variations of their own parameters than Tree-based models (note that this cannot be stated for NB Tree, which produced an unique model per sampling). Indeed, for Bayes models the MCC score roughly varies from 0.76 to 0.98, while for Tree methods it may even drop to 0 for Simple Cart (0.58 for C4.5 and Random Forest) and on the other hand reach a maximum of 0.99. Moreover, Figure 1 highlights models passing a 0.4 MCC threshold (above the dashed red line), with a large majority of Simple CART models that do not meet this threshold. When looking at the STD MCC in Figure 1, the same tendencies can be observed (here a 0.1 threshold was used and models should be below the dashed red line in order to pass the filter): models are globally consistent across the samplings but the 3 Bayes methods seem to be more robust with respect to the STD MCC (from 0.02 to 0.17) than Tree-based ones (ranging from 0 to 0.41). Moreover, around 88% of Bayes models across all samplings pass the MCC and STD MCC combined filter, while only 69% of the trees.

3.1.2 Evaluation by signature length and composition Classifiers are commonly evaluated by performance metrics but rarely based on their signatures. In Figure 2 we examine the size of the signatures output by the different models on the k different samplings. The results indicate that signature lengths both for Bayes and Tree models tend to be stable across the 3 different samplings. Moreover, the signatures have comparable lengths between models obtained with the same method (A1DE models give signatures with lengths varying from 2 to 7, Bayes Network from 3 to 6, Naive Bayes from 4 to 8, C4.5 from 1 to 6, NB Tree 4 to 6), except maybe for Simple CART (range from 1 to 7) and for Random Forest models (from 2 to 14). Additionally, the 3 NB Tree models (one per sampling) give similar signature lengths between samplings.

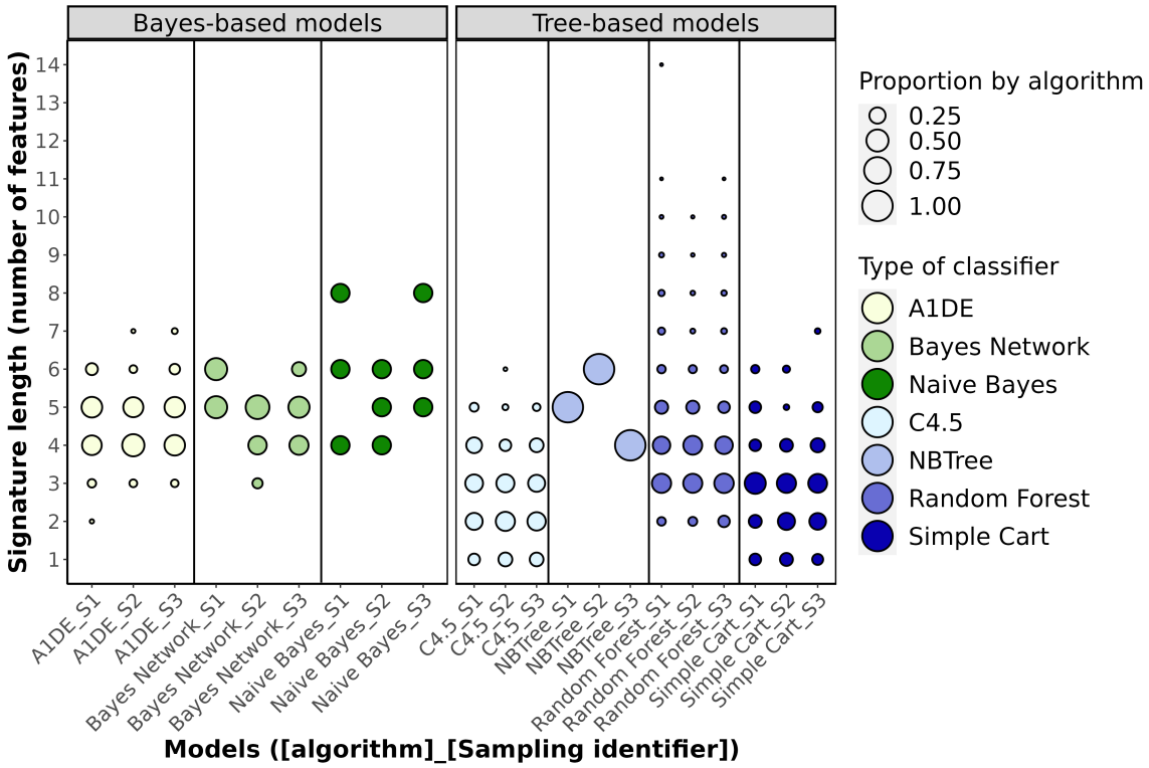

Fig. 2. Signature size variation for A1DE, Bayes Network, Naive Bayes, C4.5, NB Tree, Random Forest and Simple Cart algorithms across 3 different samplings without filtering on the MCC or STD MCC values.

For further investigations, given the great variations in performance according to MCC and STD MCC scores, models were filtered based on their MCC and STD MCC values using the following thresholds : ≥ 0.4, respectively ≤ 0.1. Next, we focus on the most relevant features for a given sampling meaning, for each method, those features that are output in the signatures of at least 20% of the associated trained models. This gives a consensus signature that is representative of a group of models corresponding to a classifier. Figure 3 depicts the presence of common features between the various algorithms, for a given sampling that was randomly picked to illustrate our remarks (the second sampling among the 3 presented in this paper). For example, one may note that Naive Bayes consensus signature has 6 specific features (that do not appear in the other consensus signatures), while the other 5 features are shared with at least two other algorithms and with up to 5 others algorithms. Apart from these features, the remaining 8 features are either unique to a given method (4 features), or shared between 2 (3 features) or 3 methods (1 feature). Below, we discuss 5 features that were found in the consensus signature for at least 2 methods.

4 Discussion and Conclusion

In our study, several classifiers from two important classes of supervised methods : Bayes and Tree-based methods were able for the most part to label our various samples in normal and Stage IV colorectal cancer. When looking at performance metrics obtained on 3 samplings (generated with a stratified sampling strategy), like the MCC and the STD MCC, Bayes-based models appears to be less

affected by the variation of their parameters compared to the Tree-based methods. Moreover, Bayes Network and Naive Bayes models, along with a majority of A1DE ones give the best results with respect to these metrics (above a minimal MCC threshold of 0.4 and below a maximal STD MCC of 0.1). On the other hand, around 69% of the Tree-based models did not pass the STD MCC threshold (and also MCC threshold for Simple CART models), thus suggesting that they are more impacted than the Bayes ones with respect to parameter tuning.

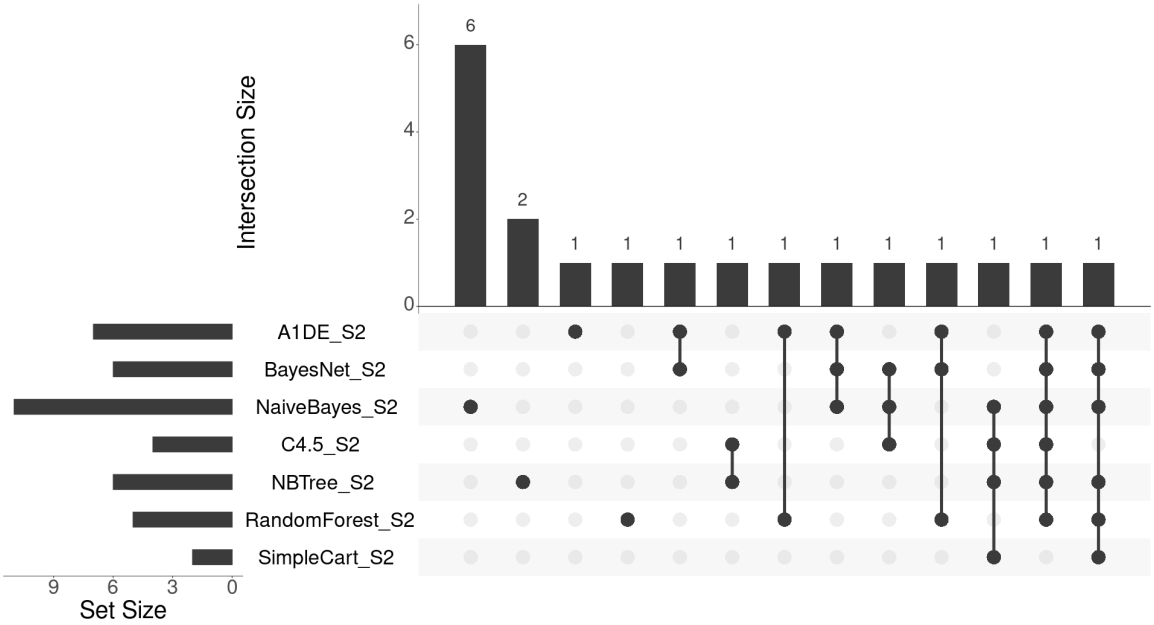

Fig. 3. Number of features being specific or shared among consensus signatures for the 7 classifiers.

When analyzing the lengths of the signatures, except from Random Forest models, the range of the length values is similar among the different methods. Additionally, when observing the prevalent features between the different methods (present in more than 20% of the models for a given method), half of them are specific to classifier, while the other half are shared between at least two algorithms. Interestingly, the ones being shared are potential candidates for biomarkers of stage IV colorectal cancer following the assumption that a prediction obtained by several methods is more likely to be robust or meaningful. Among these, the PLUT regulator (an antisens RNA), stands out as being shared across many algorithms and for all samplings. One of its targets, PDX1, has been characterized as having a major role in glucose-dependent regulation of insulin gene expression and associated to the early development of pancreatic cancers [24]. Looking at the other top genes, given by more than 50% of the algorithms, two additional RNA regulators are proposed, EDIL3-DT and LINC02418 that are related to cancer development. Moreover, the latter has already been reported as a potential biomarker for colorectal cancer [25, 26]. Finally, the SP8 transcription factor and the KLK7 gene (member of the kallikrein gene family), were identified as biomarkers for cancer [27, 28] .

This first study is a proof of concept and needs to be generalized to explore the diversity of ML methods used. Additionally, the impact of samplings will be also analyzed (by increasing the k value) while integrating additional cohorts. The comparison of the signatures produced by the classifiers will be central to define robustness criteria, based on the idea that a consensus prediction between different methods is more relevant to the expert.

References

[1] Nathan Wan et al. “Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA”. In: *BMC Cancer* 19.1 (Aug. 2019), p. 832. ISSN: 1471-2407. DOI: 10.1186/s12885-019-6003-8. URL: https://doi.org/10.1186/s12885-019-6003-8 (visited on 03/17/2022).

[2] Lin Huang et al. "Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma". eng. In: *Nature Communications* 11.1 (July 2020), p. 3556. ISSN: 2041-1723. DOI: 10.1038/s41467-020-17347-6.

[3] Xinyin Han et al. "MSIsensor-ct: microsatellite instability detection using cfDNA sequencing data". eng. In: *Briefings in Bioinformatics* 22.5 (Sept. 2021), bbaa402. ISSN: 1477-4054. DOI: 10.1093/bib/bbaa402.

[4] Meng Zhou et al. "Computational recognition of lncRNA signature of tumor-infiltrating B lymphocytes with potential implications in prognosis and immunotherapy of bladder cancer". eng. In: *Briefings in Bioinformatics* 22.3 (May 2021), bbaa047. ISSN: 1477-4054. DOI: 10.1093/bib/bbaa047.

[5] Diego Raphael Amancio et al. "A systematic comparison of supervised classifiers". eng. In: *PloS One* 9.4 (2014), e94137. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0094137.

[6] A.K. Jain et al. "Statistical pattern recognition: a review". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.1 (Jan. 2000). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 4–37. ISSN: 1939-3539. DOI: 10.1109/34.824819.

[7] Reinel Tabares-Soto et al. "A comparative study of machine learning and deep learning algorithms to classify cancer types based on microarray gene expression data". In: *PeerJ Computer Science* 6 (Apr. 2020), e270. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.270. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7924492/ (visited on 03/16/2022).

[8] Nathalie Japkowicz. "Why Question Machine Learning Evaluation Methods? An Illustrative Review of the Shortcomings of Current Methods". en. In: (), p. 6.

[9] Davide Chicco and Giuseppe Jurman. "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation". en. In: *BMC Genomics* 21.1 (Jan. 2020), p. 6. ISSN: 1471-2164. DOI: 10.1186/s12864-019-6413-7. URL: https://doi.org/10.1186/s12864-019-6413-7 (visited on 03/16/2022).

[10] Niamh Errington et al. "A diagnostic miRNA signature for pulmonary arterial hypertension using a consensus machine learning approach". English. In: *eBioMedicine* 69 (July 2021). Publisher: Elsevier. ISSN: 2352-3964. DOI: 10.1016/j.ebiom.2021.103444. URL: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(21)00237-1/fulltext (visited on 03/16/2022).

[11] Zaoqu Liu et al. "Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer". eng. In: *Nature Communications* 13.1 (Feb. 2022), p. 816. ISSN: 2041-1723. DOI: 10.1038/s41467-022-28421-6.

[12] Antonio Colaprico et al. "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data". In: *Nucleic Acids Research* 44.8 (May 2016), e71. ISSN: 0305-1048. DOI: 10.1093/nar/gkv1507. URL: https://doi.org/10.1093/nar/gkv1507 (visited on 03/16/2022).

[13] Katarzyna Tomczak et al. "The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge". In: *Contemporary Oncology* 19.1A (2015), A68–A77. ISSN: 1428-2526. DOI: 10.5114/wo.2014.47136. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4322527/ (visited on 03/16/2022).

[14] Mark D. Robinson et al. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". eng. In: *Bioinformatics (Oxford, England)* 26.1 (Jan. 2010), pp. 139–140. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btp616.

[15] Matthew E. Ritchie et al. "limma powers differential expression analyses for RNA-sequencing and microarray studies". In: *Nucleic Acids Research* 43.7 (Apr. 2015), e47. ISSN: 0305-1048. DOI: 10.1093/nar/gkv007. URL: https://doi.org/10.1093/nar/gkv007 (visited on 03/16/2022).

[16] Mickael Leclercq et al. "Large-Scale Automatic Feature Selection for Biomarker Discovery in High-Dimensional OMICs Data". eng. In: *Frontiers in Genetics* 10 (2019), p. 452. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00452.

[17] Shahadat Uddin et al. "Comparing different supervised machine learning algorithms for disease prediction". eng. In: *BMC medical informatics and decision making* 19.1 (Dec. 2019), p. 281. ISSN: 1472-6947. DOI: 10.1186/s12911-019-1004-8.

[18] George H. John and Pat Langley. "Estimating Continuous Distributions in Bayesian Classifiers". In: *arXiv:1302.4964 [cs, stat]* (Feb. 2013). arXiv: 1302.4964. URL: http://arxiv.org/abs/1302.4964 (visited on 03/17/2022).

[19] Jorge López Puga et al. "Bayesian networks". en. In: *Nature Methods* 12.9 (Sept. 2015). Number: 9 Publisher: Nature Publishing Group, pp. 799–800. ISSN: 1548-7105. DOI: 10.1038/nmeth.3550. URL: https://www.nature.com/articles/nmeth.3550 (visited on 03/17/2022).

[20] Geoffrey I. Webb et al. "Not So Naive Bayes: Aggregating One-Dependence Estimators". en. In: *Machine Learning* 58.1 (Jan. 2005), pp. 5–24. ISSN: 1573-0565. DOI: 10.1007/s10994-005-4258-6. URL: https://doi.org/10.1007/s10994-005-4258-6 (visited on 03/17/2022).

[21] Leo Breiman et al. *Classification And Regression Trees*. New York: Routledge, Oct. 2017. ISBN: 978-1-315-13947-0. DOI: 10.1201/9781315139470.

[22] Ron Kohavi. "Scaling up the accuracy of naive- Bayes classifiers: A decision-tree hybrid". en. In: (), p. 7.

[23] Ron Kohavi. "A study of cross-validation and bootstrap for accuracy estimation and model selection". In: *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*. IJCAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Aug. 1995, pp. 1137–1143. ISBN: 978-1-55860-363-9. (Visited on 03/23/2022).

[24] I Akerman et al. "Human Pancreatic Cell lncRNAs Control Cell-Specific Regulatory Networks". English. In: *Cell Metabolism* 25.2 (2017), pp. 400–411. ISSN: 1550-4131. DOI: 10.1016/j.cmet.2016.11.016.

[25] Yinghui Zhao et al. "Long noncoding RNA LINC02418 regulates MELK expression by acting as a ceRNA and may serve as a diagnostic marker for colorectal cancer". eng. In: *Cell Death & Disease* 10.8 (July 2019), p. 568. ISSN: 2041-4889. DOI: 10.1038/s41419-019-1804-x.

[26] Jun Tian et al. "LINC02418 promotes colon cancer progression by suppressing apoptosis via interaction with miR-34b-5p/BCL2 axis". In: *Cancer Cell International* 20.1 (Sept. 2020), p. 460. ISSN: 1475-2867. DOI: 10.1186/s12935-020-01530-2. URL: https://doi.org/10.1186/s12935-020-01530-2 (visited on 03/30/2022).

[27] Alexandra Elisabeth Wagner et al. "SP8 Promotes an Aggressive Phenotype in Hepatoblastoma via FGF8 Activation". In: *Cancers* 12.8 (2020). ISSN: 2072-6694. DOI: 10.3390/cancers12082294. URL: https://www.mdpi.com/2072-6694/12/8/2294.

[28] Francine Walker et al. "Kallikrein-related peptidase 7 (KLK7) is a proliferative factor that is aberrantly expressed in human colon cancer". In: *Biological Chemistry* 395.9 (2014), pp. 1075–1086. DOI: doi:10.1515/hsz-2014-0142. URL: https://doi.org/10.1515/hsz-2014-0142.