



**HAL**  
open science

# Using Machine Learning and Deep Learning Methods to Predict the Complexity of Breast Cancer Cases

Akram Redjdal, Jacques Bouaud, Joseph Gligorov, Brigitte Seroussi

► **To cite this version:**

Akram Redjdal, Jacques Bouaud, Joseph Gligorov, Brigitte Seroussi. Using Machine Learning and Deep Learning Methods to Predict the Complexity of Breast Cancer Cases. Challenges of Trustable AI and Added-Value on Health, IOS Press, 2022, Studies in Health Technology and Informatics, 10.3233/SHTI220400 . hal-03681262

**HAL Id: hal-03681262**

**<https://hal.science/hal-03681262>**

Submitted on 12 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Using Machine Learning and Deep Learning Methods to Predict the Complexity of Breast Cancer Cases

Akram REDJDAL<sup>a,1</sup>, Jacques BOUAUD<sup>a</sup>, Joseph GLIGOROV<sup>b,c</sup> and  
Brigitte SEROUSSI<sup>a,c,d</sup>

<sup>a</sup>*Sorbonne Université, Université Sorbonne Paris Nord, INSERM, Laboratoire d'Informatique Médicale et d'Ingénierie des connaissances en e-Santé, LIMICS, F-75006 Paris, France*

<sup>b</sup>*Sorbonne Université, Institut Universitaire de Cancérologie, Paris, France*

<sup>c</sup>*AP-HP, Hôpital Tenon, Paris, France*

<sup>d</sup>*APREC, Paris, France*

**Abstract.** In many countries, the management of cancer patients must be discussed in multidisciplinary tumor boards (MTBs). These meetings have been introduced to provide a collaborative and multidisciplinary approach to cancer care. However, the benefits of MTBs are now being challenged because there are a lot of cases and not enough time to discuss all of them. During the evaluation of the guideline-based clinical decision support system (CDSS) of the DESIREE project, we found that for some clinical cases, the system did not produce recommendations. We assumed that these cases were complex clinical cases and needed deeper MTB discussions. In this work, we trained and tested several machine learning and deep learning algorithms on a labelled sample of 298 breast cancer patient summaries, to predict the complexity of a breast cancer clinical case. XGboost and multi-layer perceptron were the models with the best result, with an F1 score of 83%.

**Keywords.** Supervised Machine Learning, Deep learning, Binary classification, Breast Cancer.

## 1. Introduction

In most countries, therapeutic decisions for cancer patients have currently to be discussed in multidisciplinary tumor boards (MTBs). These meetings have been introduced to provide a collaborative and multidisciplinary approach to cancer care, bringing together surgery, oncology, radiology, and pathology specialists to optimize the decision-making process, and offer the best management to cancer patients. The benefits of MTBs, which have long been taken for granted, are recently being challenged. Positive outcomes from MTBs depend on the presence of qualified experts,

---

<sup>1</sup>Corresponding Author, Akram REDJDAL; E-mail: [redjdalakram300@gmail.com](mailto:redjdalakram300@gmail.com).

This work has been financed by a doctoral grant for AR from the University Institute of Health Engineering (UIS, Sorbonne University, Paris, France) and received the support of AP-HP health data warehouse

good preparation of patient cases, efficient leadership, time management, and good interactions among MTB clinicians [1].

DESIREE is a European project which aimed at developing web-based services for the management of primary breast cancer by MTBs. During the evaluation of the guideline-based clinical decision support system (CDSS) of DESIREE, we found that for some patient cases the system did not provide any therapeutic proposals or provided recommendations which were not followed by MTB clinicians [2]. These clinical cases were considered as complex cases, and we assume that such cases are not correctly handled by guidelines and need deeper MTB discussion.

Our aim is to optimize the organization of MTBs by identifying complex cases prior to MTBs so that the way clinical cases are discussed could be adapted to allow for a longer discussion for complex cases. We consider the problem of assessing the complexity of a breast cancer clinical case as a binary classification task. We used machine and deep learning models and assessed the performance of such approaches to learn the complexity of breast cancer cases on the basis of breast cancer patient summaries (BCPSs) which are textual documents summarizing the description of the patient condition used as the basis for MTB discussion.

## 2. Material and Methods

### 2.1. Breast cancer patient summaries

Breast cancer patient summaries (BCPS) are free text documents used during MTBs. They provide a portrait of a patient case with all relevant information clinicians need to know to make the best patient-specific therapeutic decision. BCPSs contain different types of information, but usually have a common structure with the reason for presentation, tumor type, biometric data, patient personal history, patient family history, TNM staging, etc. Assistance Publique-Hôpitaux de Paris (AP-HP) is the greatest healthcare public university hospital with 39 sites. AP-HP collects patient data, with their consent, in a data warehouse from which BCPSs used in this study were collected.

### 2.2. Extraction of structured data from BCPSs

We previously used semantic annotators to extract structured data from textual BCPSs [3]. ECMT and SIFR [4] are two annotators that work for the French language. To take advantage of two annotators widely used in the biomedical field for the English language, MetaMap and cTAKES [5], we automatically translated French BCPSs in English. Since BCPSs are textual documents containing a lot of abbreviations, acronyms, and specialized terms related to the oncological field (e.g., “Her2”, “IRM”, “TEP”), a first step was to disambiguate the texts. To solve this issue, we created a local dictionary with medical acronyms and their expansion based on online available dictionaries. Then, we replaced acronyms in BCPSs by their expansion, and finally used the pre-trained Opus-MT translation model (<https://opus.nlpl.eu>). As a result, all BCPSs were available in both French and English. As a previous work [3] concluded that the application of all four annotators gave the best set of annotations, including complexity-related concepts, we executed the four annotators, and processed the output of each annotator to generate a semantic representation of a BCPS as two vectors, a vector of UMLS concepts (CUI) extracted by SIFR, cTAKES, and MetaMap. And a

second vector containing the labels of the concepts extracted by ECMT (ECMT does not extract UMLS CUIs). For each concept, we associated information about negation as attached to the concept provided by the annotators (e.g., in “*absence d’adénopathie*”, the adenopathy concept was present but identified as negated). Figure 1 depicts the whole sequence implemented for the extraction of structured data from BCPSs.

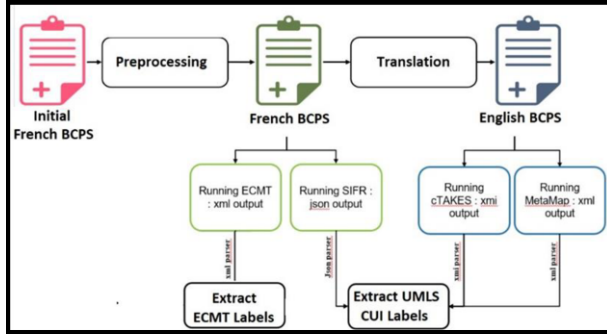


Figure 1. Structured data extraction

### 2.3. Data preprocessing

In order to get a BCPS representation consumable by all learning models starting from the two vectors obtained from annotators, we converted each BCPS into a row of features that represented the clinical concepts. We included all the labels of the concepts extracted and the value for each feature was 1 if the concept was present, 0 if the concept was not present, and -1 if it was present and negated. We preserved the order of concepts as expressed in a BCPS by using an index column to specify the order in which they appeared in the text.

### 2.4. Learning algorithms

Three supervised machine learning algorithms were tested (Decision trees, Random forests, XGboost). For deep learning algorithms, we previously tested Multilayer Perceptron (MLP) on a similar task [6], and we wanted to compare MLP performance to models such as Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs).

The learning dataset was made of a sample of BCPSs which were blindly annotated as “complex” or “non complex” by two expert clinicians. When the experts disagreed, we considered the case as complex.

### 2.5. Model training pipeline

We trained the six selected models using a k-fold cross-validation strategy, where models’ hyperparameters tuning process was executed using Grid Search. Additionally, we selected the most relevant variables for each model using feature selection. Resulting classification models were classically evaluated using precision, recall, and F1-score, both before and after feature selection. Figure 2 displays the whole sequence of treatments for training machine and deep learning model, starting from UMLS CUIs and ECMT labels.

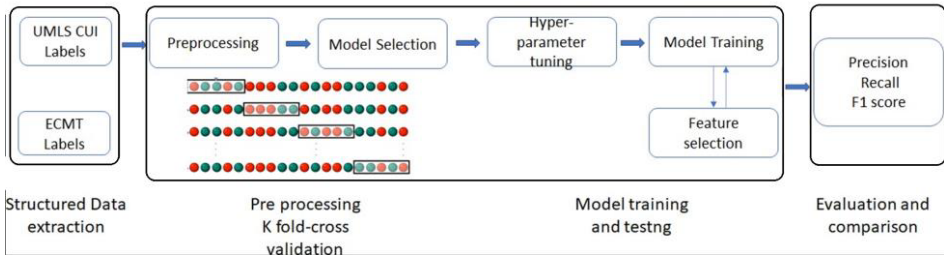


Figure 2. Model training pipeline

### 3. Results

We worked on a sample of 298 BCPSs corresponding to clinical cases discussed in MTBs between April and July 2021 at the Tenon Hospital in Paris (France). From the 298 BCPSs, we extracted a total number of 10,719 concepts (either negated or not). The presence and negation of these variables in BCPSs were then used as a feature in the inputs. So, for each model, we had a dataset of 10,719 features and 298 patient cases. Following the analytic plan, we generated an initial training dataset and a testing dataset. The training set was divided into 5 splits (stratified-k-fold with  $k=5$ ) on which we trained each model. All combinations were submitted to Grid Search hyperparameters tuning. The results of each model, with its best parameters are shown in table 1.

Table 1. Evaluation of machine and deep learning algorithms

Models	<i>Before feature Selection</i>			<i>After feature Selection</i>		
	Precision	Recall	F1	Precision	Recall	F1
<b>XGboost</b>	<b>0.82</b>	<b>0.83</b>	<b>0.83</b>	<b>0.82</b>	<b>0.83</b>	<b>0.83</b>
Random Forest	0.51	0.71	0.59	0.51	0.71	0.59
Decision Trees	0.51	0.71	0.59	0.51	0.71	0.59
<b>MLP</b>	<b>0.84</b>	<b>0.83</b>	<b>0.83</b>	<b>0.81</b>	<b>0.83</b>	<b>0.83</b>
CNN	0.51	0.71	0.59	0.51	0.71	0.59
RNN	0.55	0.65	0.58	0.55	0.65	0.58

### 4. Discussion and Conclusions

The main goal of this work was to assess the capacity of machine learning and/or deep learning algorithms to predict the complexity of breast cancer clinical cases in order to optimize MTB organization. For instance, a guideline based CDSS recommendation might be followed for non-complex cases, whereas it would be discarded for complex cases.

Starting from a set of textual unstructured BCPSs priorly tagged as complex or not by experts, we were able to extract structured data using semantic annotators and use these structured data as input for learning algorithms. Results showed that XGboost and MLP performed the best with an equal F1 score of 83%. For all models, it is noticeable that feature selection did not improve the results. So, without feature selection, MLP is slightly better than XGboost in terms of precision (0.84 vs 0.82) and comparable in terms of recall (0.83). From this set of models and on this dataset, there was no

evidence of deep learning outperformance over classical machine learning. In our analyses, the MLP model was a multi-layer feed-forward neural network with two layers. More complex network architecture might improve the MLP approach, but this was not tested yet.

Globally, the used algorithms showed good results as compared to previous breast cancer complexity detection research work that we performed in a different context [6]. A similar research work used machine learning from free text to identify breast cancer receptor status with F1 scores reaching 0.89 to 0.92 [7]. Such results are better than ours, but the task looks to be simpler, and the training set used was larger (n=1300). Another work using learning algorithms to predict lung cancer recurrence from EHRs demonstrated a F1 score of 0.72 on a cohort of 2,442 patients [8]. This suggests that our approach is feasible and that our results are acceptable (they could certainly be improved with more annotated BCPSs). This confirms the hypothesis that the use of such algorithms could be useful for this kind of predictions [9]. We wanted to use the same input for both machine learning and deep learning algorithms and the input was a 2-dimension matrix with the presence or absence of concepts retrieved in BCPS texts by semantic annotators. Going further, encoders such as umlsBERT could also be used on BCPSs to classify the data in addition to work on a larger dataset, but classical machine learning algorithms could not directly be compared on such a textual input.

## References

- [1] El Saghir NS, Keating NL, Carlson RW, Khoury KE, Fallowfield L. Tumor boards: optimizing the structure and improving efficiency of multidisciplinary management of patients with cancer worldwide. *Am Soc Clin Oncol Educ Book*. 2014:e461-6. doi: 10.14694/EdBookAM.2014.34.e461. PMID:24857140
- [2] Redjdal A, Bouaud J, Guezennec G, Gligorov J, Seroussi B. Reusing Decisions Made with One Decision Support System to Assess a Second Decision Support System: Introducing the Notion of Complex Cases. *Stud Health Technol Inform*. 2021 May 27; 281:649-653. doi: 10.3233/SHTI210251. PMID: 34042656.
- [3] Redjdal A, Bouaud J, Gligorov J, Seroussi B. Are Semantic Annotators Able to Extract Relevant Complexity-Related Concepts from Clinical Notes? *Stud Health Technol Inform*. 2021 Nov 18; 287:153- 157. doi: 10.3233/SHTI210836. PMID: 34795101.
- [4] Sakji S, Gicquel Q, Pereira S, Kergourlay I, Proux D, Darmoni S, Metzger MH. Evaluation of a French medical multi-terminology indexer for the manual annotation of natural language medical reports of healthcare-associated infections. In *MEDINFO 2010 2010* (pp. 252-256). IOS Press.
- [5] Reátegui R, Ratté S. Comparison of MetaMap and cTAKES for entity extraction in clinical notes. *BMC Med Inform Decis Mak*. 2018 Sep 14;18(Suppl 3):74. doi: 10.1186/s12911-018-0654-2. PMID:30255810; PMCID: PMC6157281.
- [6] Le Thien MA, Redjdal A, Bouaud J, Seroussi B. Deep Learning, a Not so Magical Problem Solver: A Case Study with Predicting the Complexity of Breast Cancer Cases. *Stud Health Technol Inform*. 2021 Nov 18; 287:144-148. doi: 10.3233/SHTI210834. PMID: 34795099.
- [7] Pironet A, Poirel HA, Tambuyzer T, De Schutter H, van Walle L, Mattheijssens J, et al. Machine Learning-Based Extraction of Breast Cancer Receptor Status from Bilingual Free-Text Pathology Reports. *Front Digit Health*. 2021 Aug 17 ;3:692077.
- [8] Mohamed SK, Walsh B, Timilsina M, Torrente M, Franco F, Provencio M, et al. On Predicting Recurrence in Early Stage Non-Small Cell Lung Cancer. In *Proc. AMIA 2021*. To appear.
- [9] Sekercioglu N, Fu R, Kim SJ, Mitsakakis N. Machine learning for predicting long-term kidney allograft survival: a scoping review. *Ir J Med Sci*. 2021 May;190(2):807-817. doi: 10.1007/s11845-020-02332-1. Epub 2020 Aug 6. PMID: 32761550.