Developing Machine Learning Algorithms for the Prediction of Early Death in Elderly Cancer Patients: Usability Study

Gabrielle Ribeiro Sena¹*, MSc; Tiago Pessoa Ferreira Lima¹²*, PhD; Maria Julia Gonçalves Mello¹*, PhD; Luiz Claudio Santos Thuler³*, PhD; Jurema Telles Oliveira Lima¹*, PhD

¹Department of Geriatric Oncology, Instituto de Medicina Integral Prof Fernando Figueira, Recife, Brazil
²Instituto Federal de Pernambuco - IFPE, Department os Computational Science, Recife, Brazil
³Research Center, Instituto Nacional do Cancer - INCA, Rio de Janeiro, Brazil

*all authors contributed equally

Corresponding Author:
Gabrielle Ribeiro Sena, MSc
Department of Geriatric Oncology
Instituto de Medicina Integral Prof Fernando Figueira
Rua dos Coelhos 300
Recife, 50070-902
Brazil
Phone: 55 81 21224100
Email: gabriellesena8@gmail.com

Abstract

Background: The importance of classifying cancer patients into high- or low-risk groups has led many research teams, from the biomedical and bioinformatics fields, to study the application of machine learning (ML) algorithms. The International Society of Geriatric Oncology recommends the use of the comprehensive geriatric assessment (CGA), a multidisciplinary tool to evaluate health domains, for the follow-up of elderly cancer patients. However, no applications of ML have been proposed using CGA to classify elderly cancer patients.

Objective: The aim of this study was to propose and develop predictive models, using ML and CGA, to estimate the risk of early death in elderly cancer patients.

Methods: The ability of ML algorithms to predict early mortality in a cohort involving 608 elderly cancer patients was evaluated. The CGA was conducted during admission by a multidisciplinary team and included the following questionnaires: mini-mental state examination (MMSE), geriatric depression scale-short form, international physical activity questionnaire-short form, timed up and go, Katz index of independence in activities of daily living, Charlson comorbidity index, Karnofsky performance scale (KPS), polypharmacy, and mini nutritional assessment-short form (MNA-SF). The 10-fold cross-validation algorithm was used to evaluate all possible combinations of these questionnaires to estimate the risk of early death, considered when occurring within 6 months of diagnosis, in a variety of ML classifiers, including Naive Bayes (NB), decision tree algorithm J48 (J48), and multilayer perceptron (MLP). On each fold of evaluation, tiebreaking is handled by choosing the smallest set of questionnaires.

Results: It was possible to select CGA questionnaire subsets with high predictive capacity for early death, which were either statistically similar (NB) or higher (J48 and MLP) when compared with the use of all questionnaires investigated. These results show that CGA questionnaire selection can improve accuracy rates and decrease the time spent to evaluate elderly cancer patients.

Conclusions: A simplified predictive model aiming to estimate the risk of early death in elderly cancer patients is proposed herein, minimally composed by the MNA-SF and KPS. We strongly recommend that these questionnaires be incorporated into regular geriatric assessment of older patients with cancer.

(JMIR Cancer 2019;5(2):e12163) doi: 10.2196/12163

KEYWORDS
geriatric assessment; aged; machine learning; medical oncology; death
Introduction

Background

Aging is a complex and personal, cumulative, and irreversible phenomenon that goes well beyond chronological age [1]. It involves several biological events associated with a great variety of molecular and cellular damage, leading to the gradual loss of physiological and immunological reserves and a greater risk for neoplasia-related death [1,2]. Assuming that the elderly population is heterogeneous, this population must be considered not only concerning their chronological age. Thus, an objective analysis of their living conditions as well as aspects related to oncological disease and its therapy is also required [3].

The International Society of Geriatric Oncology has recommended the use of the Comprehensive Geriatric Assessment (CGA) for the evaluation and follow-up of elderly cancer patients [4]. The CGA is a multidisciplinary tool that uses validated instruments to evaluate several elderly health condition domains, such as functional, cognitive, psychological, social, clinical, and nutritional aspects, as well as comorbidities and the use of medication, among others [5,6]. It is also strongly recommended by the geriatrics and gerontology fields in general because it is, in a complex and heterogeneous context, an objective, measurable, and reproducible form of evaluation, adding possibilities to standard clinical laboratory evaluations [7,8]. However, there is no consensus about what and how many instruments should be used. Employing CGA in practice, however, has become a huge challenge, and owing to its complexity and time spent in its application, it is often underutilized by oncologists and not judged as a completely satisfactory solution in practice, which has served as a stimulus for the construction of simpler tools that have the power to predict outcomes and guide clinical decisions [5,9].

The accurate prediction of a disease outcome is one of the most interesting and challenging tasks for physicians. As a result, a growing trend was noted in the studies published during the past years that applied machine learning (ML) algorithms for modeling cancer survival. This type of algorithms can discover and identify patterns and relationships between them, from complex databases, while they are able to effectively predict future outcomes of a cancer type [10]. On the basis of the study by Kourou et al [11], the accuracy of cancer prediction outcome has significantly improved by 15% to 20% in the previous years, with the application of ML techniques.

A study combining data from 4 cohorts involving the elderly, including elderly people with neoplasms, proposed to explore the performance of various ML classifiers (Naive Bayes [NB], k-nearest neighbors, artificial neural networks, random forest, and logistic regression) regarding death prediction in 6 months [12]. Another study used ML to predict mortality of patients in 3 to 12 months and to identify patients who could benefit from palliative care [13]. However, no ML application has been proposed using CGA to classify elderly cancer patients.

Objectives

Thus, the primary aim of this study was to propose and develop predictive models, using ML and CGA, to estimate the risk of early death in elderly cancer patients. The secondary aims were to optimize the CGA through the selection of the most appropriate instruments.

Methods

Comprehensive Geriatric Assessment

The ability of ML techniques to predict early mortality in a heterogeneous cohort was tested in 608 elderly cancer patients (aged over 60 years), admitted to the oncogeriatrics sector of the Instituto de Medicina Integral Prof. Fernando Figueira - IMIP, from January 2015 to July 2016. The IMIP is a teaching hospital and cancer center located in Recife, Pernambuco, Brazil. On admission to the cohort database, the patients were evaluated by CGA questionnaires presented in Table 1. The questionnaires were collected by a multiprofessional team, comprising a clinical oncologist, a geriatrician, a physiotherapist, a physical educator, a speech therapist, an occupational therapist, and a nutritionist. The project was approved by the IMIP Ethics Committee on Human Research on June 30, 2016, under number 58298316.5.0000.5201.
Predictive Models

Predictive modeling is the general concept of building a model that can make predictions. Typically, such a model includes an ML algorithm that learns certain properties from a database to make those predictions. We have presented below a brief summary of the commonly used supervised learning algorithms:

- Decision tree J48 (J48) [25]: They are tree-like graphs, where the nodes in the graph test certain conditions on a set of features and the branches split the decision toward the leaf nodes. The leaves represent the lowest level in the graph and determine the class labels.
- Multilayer perceptron (MLP) [26]: They are graph-like classifiers that mimic the structure of a human or animal brain where the interconnected nodes represent the neurons.
- Naive Bayes (NB) [27]: They are based on a statistical model (ie, Bayes theorem, calculating posterior probabilities based on the prior probability and the so-called likelihood).

The purpose of this work was not to introduce the highest accuracy prediction model. The goal was to designate the most relevant questionnaires to evaluate elderly health condition domains in CGA. Therefore, in the experiments, we always used the same configuration with the default parameter values in Weka (Waikato Environment of Knowledge Analysis) from The University of Waikato, version 3.8.3. The advantage of using default parameters is that it does not introduce optimistic bias by tuning the parameter to maximize performance on the test data. Figure 1 shows more details about the values used in each predictive model.

Preprocess of Database

The first step was to remove patients presenting redundancies and/or incomplete questionnaires/features. A total of 543 patients remained after that. Data normalization technique for equalizing the range of features, usually employed in the database before feature selection and learning phase, is of important concern in pattern recognition and computer-aided diagnosis [23]. The most common normalization method used during data transformation is the min-max (where the features are mapped into a predefined range, varying from 0 or -1 to 1). The main advantage of min-max normalization is that it preserves the relationships among the original data values [24]. In this work, all features were normalized in a (0,1) interval, calculated as in equation, where $v'$ is the value normalized, $v$ is the original value, $v_{\min}$ is the minimum value of corresponding feature, and $v_{\max}$ is the maximum value of corresponding feature: $v'=(v-v_{\min})/(v_{\max}-v_{\min}).$

Table 1. Questionnaires/features to evaluate elderly health condition domains in Comprehensive Geriatric Assessment.

<table>
<thead>
<tr>
<th>Questionnaire/feature</th>
<th>Perspective</th>
<th>Range/cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charlson comorbidity index [14]</td>
<td>A prospective method for classifying comorbid conditions that might alter the risk of mortality</td>
<td>0 to 37 points, with an increase of up to 5 points per age range</td>
</tr>
<tr>
<td>Geriatric depression scale-short form [15]</td>
<td>A self-report measure of depression in older adults. Users respond in a yes/no format</td>
<td>score 0 to ≤5 is normal; score &gt;5 is depression</td>
</tr>
<tr>
<td>International physical activity questionnaires-short form [16]</td>
<td>A set of questionnaires to obtain international comparative data on physical activity</td>
<td>0 is sedentary, 1 is insufficiently active, 2 is active, 3 is active, and 4 is very active</td>
</tr>
<tr>
<td>Karnofsky performance scale [17]</td>
<td>Used to quantify patients’ general well-being and activities of daily life</td>
<td>0 to 100, the lower the score, the worse the survival for most serious illnesses</td>
</tr>
<tr>
<td>Katz index of independence in activities of daily living [18]</td>
<td>Was developed to study results of treatment and prognosis in the elderly and chronically ill. Grades of the index summarize overall performance in bathing, dressing, going to toilet, transferring, continence, and feeding</td>
<td>0 to 6, high score means patient is independent and low score means patient is very dependent</td>
</tr>
<tr>
<td>Mini-mental state examination [19]</td>
<td>A method for grading the cognitive state of patients for the clinician</td>
<td>0 to 30, the lower the score rate, the worse the cognitive impairment</td>
</tr>
<tr>
<td>Mini nutritional assessment-short form [20]</td>
<td>A screening tool used to identify older adults who are malnourished or at risk of malnutrition. Comprises 6 questions on food intake, weight loss, mobility, psychological stress or acute disease, presence of dementia or depression, and body mass index</td>
<td>0 to 14, scores of 12-14 are considered normal nutritional status; 8-11 indicate at risk of malnutrition; and 0-7 indicate malnutrition</td>
</tr>
<tr>
<td>Polyparmacy [21]</td>
<td>Refers to the regular use of a greater number of medicines (5 or more drugs)</td>
<td>0 is no; 1 is yes</td>
</tr>
<tr>
<td>Timed up and go [22]</td>
<td>The patient is observed and timed while he rises from an arm chair, walks 3 m, turns, walks back, and sits down again</td>
<td>0 is low risk of falling (less than 20 seconds), 1 is average risk of falling (20-29 seconds), and 2 is high risk of falling (30 seconds or more)</td>
</tr>
</tbody>
</table>
K-Fold Cross-Validation

Cross-validation (CV) [28] is one of the most widely used methods to assess the generalizability of predictive models [29] and is subject to ongoing active research [30]. K-fold CV comprises dividing the database into K parts (folds) of equal sizes. For this study, a 10-fold CV is used, and each part is held out in turn and the predictive model (J48, MLP, or NB) is trained on the remaining nine-tenths; then, its error rate is calculated on the holdout set. Thus, the learning procedure is executed a total of K times on different training sets (each of which have much in common). Finally, the K error estimates are averaged to yield an overall error estimate. In this work, the folds are made by preserving the percentage of samples for each class.

Imbalanced Learning

The learning procedure and the subsequent prediction of predictive models can be affected by the problem of imbalanced database [31]. The balancing issue corresponds to the difference in the number of samples in the different classes. The resulting database presented 92 deaths within 6 months of admission to the service and 451 patients alive at the end of that period. All deaths were attributed to cancer (treatment complications or disease progression). With a greater imbalanced ratio, the decision function favors the class with the largest number of samples, usually referred as the majority class. The way to fight this issue was to generate new training sets on 10-fold CV by random sampling so that the proportion between classes remained at one-to-one.

Metrics

The area under receiver operating characteristics curve, or simply area under curve (AUC), has recently been proposed as an alternative single-number measure for evaluating the generalization of learning algorithms [32]. This measure is far better than classification accuracies when the 2 classes are unbalanced and the cost of misclassification is unspecified [33]. An area of 1.0 represents a model that made all predictions perfectly, and an area of 0.5 represents a model as good as random. AUC can be broken down into sensitivity and specificity:

- Sensitivity is the true positive rate, and for this study, it is the percentage of patients with early death that are predicted correctly.
- Specificity is also called the true negative rate, for example, the percentage of patients without early death that are predicted correctly.

Results

Evaluating All Possible Combinations of Comprehensive Geriatric Assessment Questionnaires

Feature selection is an important and frequently used technique for dimension reduction by removing irrelevant and/or redundant information from the database to obtain an optimal feature subset. A 10-fold CV was used to evaluate all possible combinations of CGA questionnaires, presented in Table 2, to estimate the risk of early death in elderly cancer patients. Thus, in each fold, the combination of questionnaires with highest AUC is selected. The same folds are applied to all 511 combinations. Tiebreaking is handled by choosing the smallest set of questionnaires. The occurrence of questionnaires selected on the 10-fold CV, using predictive models, is presented in Table 2. In Figure 2, the flowchart of our methodology is shown.
Table 2. Occurrence of the Comprehensive Geriatric Assessment questionnaires in the 10-folds using decision tree (J48), multilayer perceptron, and Naive Bayes.

<table>
<thead>
<tr>
<th>Model</th>
<th>Charlson co-morbidity Index</th>
<th>Geriatric depression scale-short form</th>
<th>International physical activity questionnaire-short form</th>
<th>Katz index of independence in activities of daily living</th>
<th>Karnofsky performance scale</th>
<th>Mini-mental state examination</th>
<th>Mini nutritional assessment-short form</th>
<th>Polypharmacy</th>
<th>Timed up and go</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision tree</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Multilayer perceptron</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2. Flowchart of methodology.

Evaluating combinations of occurrences

Tables 3-5 show the sensitivity, specificity, and AUC values expressed as mean (SD) on the 10-fold CV for the NB, J48, and MLP. The subsets of CGA questionnaires, presented in these tables, consider the occurrences of Table 2. The subset of questionnaires with occurrence ≥0, for example, uses all set of CGA questionnaires, as it considers all occurrences greater than or equal to 0. The other subsets use the same logic and are detailed in the footnotes under the tables. In each metric, according to the paired t test, the P value is calculated considering the subset of questionnaires with occurrence ≥0. The experimental results demonstrate that the feature selection can discard questionnaires and finally find out subsets that reduce the dimensionality of data to make the predictive models more efficient and the results more accurate. Thus, a simplified predictive model aiming to estimate the risk of early death in elderly cancer patients is proposed herein, minimally composed by the Mini Nutritional Assessment-Short Form (MNA-SF), accompanied or not by the Karnofsky performance scale (KPS) and/or the Mini-Mental State Examination.
### Table 3. Metrics considering Comprehensive Geriatric Assessment questionnaire subsets on Naive Bayes classifier.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subsets of questionnaires with occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≥10 occurrences</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>P value</td>
<td></td>
</tr>
<tr>
<td>Sensibility</td>
<td>81.61 (4.62)</td>
</tr>
<tr>
<td>Specificity</td>
<td>65.89 (14.72)</td>
</tr>
<tr>
<td>AUC e</td>
<td>82.43 (6.35)</td>
</tr>
</tbody>
</table>

<sup>a</sup>≥10 occurrences: All comprehensive geriatric assessments (Charlson comorbidity index, geriatric depression scale-short form, international physical activity questionnaire-short form, Katz index of independence in activities of daily living, Karnofsky performance scale, mini-mental state examination, mini nutritional assessment-short form, polypharmacy, and timed up and go).

<sup>b</sup>≥6 occurrences: Charlson comorbidity index, geriatric depression scale-short form, international physical activity questionnaire-short form, Karnofsky performance scale, mini-mental state examination, and mini nutritional assessment-short form.

<sup>c</sup>≥9 occurrences: Charlson comorbidity index, Karnofsky performance scale, mini-mental state examination, and mini nutritional assessment-short form.

<sup>d</sup>10 occurrences: Karnofsky performance scale, mini-mental state examination, and mini nutritional assessment-short form.

<sup>e</sup>AUC: area under curve.

### Table 4. Metrics considering comprehensive geriatric assessment questionnaire subsets on decision tree (J48) classifier.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subsets of questionnaires with occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≥10 occurrences</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>P value</td>
<td></td>
</tr>
<tr>
<td>Sensibility</td>
<td>70.34 (16.79)</td>
</tr>
<tr>
<td>Specificity</td>
<td>62.89 (15.11)</td>
</tr>
<tr>
<td>AUC e</td>
<td>67.55 (10.27)</td>
</tr>
</tbody>
</table>

<sup>a</sup>≥10 occurrences: all comprehensive geriatric assessments (Charlson comorbidity index, geriatric depression scale-short form, international physical activity questionnaire-short form, Katz index of independence in activities of daily living, Karnofsky performance scale, mini-mental state examination, mini nutritional assessment-short form, polypharmacy, and timed up and go).

<sup>b</sup>≥4 occurrences: Charlson comorbidity index, geriatric depression scale-short form, international physical activity questionnaire-short form, and mini nutritional assessment-short form.

<sup>c</sup>≥6 occurrences: Charlson comorbidity index and mini nutritional assessment-short form.

<sup>d</sup>10 occurrences: mini nutritional assessment-short form.

<sup>e</sup>AUC: area under curve.

### Table 5. Metrics considering comprehensive geriatric assessment questionnaires subsets on multilayer perceptron classifier.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Subsets of questionnaires with occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≥10 occurrences</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>P value</td>
<td></td>
</tr>
<tr>
<td>Sensibility</td>
<td>68.75 (8.34)</td>
</tr>
<tr>
<td>Specificity</td>
<td>62.67 (17.84)</td>
</tr>
<tr>
<td>AUC e</td>
<td>69.64 (9.83)</td>
</tr>
</tbody>
</table>

<sup>a</sup>≥10 occurrences: all comprehensive geriatric assessments (Charlson comorbidity index, geriatric depression scale-short form, international physical activity questionnaire-short form, Katz index of independence in activities of daily living, Karnofsky performance scale, mini-mental state examination, mini nutritional assessment-short form, polypharmacy, and timed up and go).

<sup>b</sup>≥1 occurrence: Karnofsky performance scale, mini-mental state examination, and mini nutritional assessment-short form.

<sup>c</sup>10 occurrences: Karnofsky performance scale and mini nutritional assessment-short form.

<sup>e</sup>AUC: area under curve.
Discussion

Principal Findings

Results indicate that the MNA-SF has greater predictive power to estimate the risk of early death in elderly cancer patients as it was selected on the 10-folds. MNA-SF is a rapid test validated for screening for nutritional risk and malnutrition in the elderly population. The predictive value of MNA-SF for early death may be related to the fact that the 6 MNA-SF questions cover areas other than just nutrition, which are frequently included in the CGA, such as mobility, neuropsychological disorders, and self-reported health, in addition to nutrition aspects, including weight loss, reduced food intake, and body mass index. In fact, low MNA-SF may reveal the effects of advanced disease in the overall health of patients, which also affects cancer-related mortality. A Brazilian study showed that abnormal nutritional status was an independent factor associated with hospital death among older patients with various chronic diseases, including cancer [34]. A similar association was also demonstrated in elderly Asian cancer patients who would receive first-line chemotherapy [35]. Finally, a French multicenter study with 348 elderly cancer patients aged 70 years and above also found that low MNA scores were associated with increased risk of premature death [36].

The results also indicated that KPS questionnaire has proven itself a valuable tool to estimate the risk of early death in elderly cancer patients. In the past decades, various studies have demonstrated the prognostic value of the KPS not only primarily for various cancers [37-40] but also for other disease entities [41]. It can also be considered as a significant indicator of hospitalization and survival time, in addition to identifying risk groups to assist in the orientation of patients to geriatric outpatient [42].

Limitations

The efforts of this paper are a starting point. They provide solid evidences and some clinical recommendations. We proposed and developed simple ML models for the prediction of early death in elderly cancer patients. These models are accurate and precise and could be possibly used by clinicians to make proper treatment plans. However, additional research is needed to continue to strengthen the evidence base.

Conclusions

The results showed that the MNA-SF and KPS have the highest predictive power to identify elderly patients at risk for early death. We strongly recommend that these questionnaires be incorporated into regular geriatric assessment of older patients with cancer.

The MNA-SF and the KPS requires only a few minutes to be completed. In addition, both can be easily managed by any member of the multidisciplinary team to help in the early identification of patients at risk, providing information that assists in the planning of interventions and improving the adherence to CGA in daily clinical oncology practice.

This study also has limitations that should be considered. This is a nonrandomized, single-center, exploratory study of a heterogeneous patient population similar to a real-life population of older patients with cancer. Conversely, some of its weaknesses could be considered the main strengths of the study: this is one of the few studies in Brazil that, in the clinical practice context of a Unified Health System oncology unit, investigated the use of ML algorithms in the prediction of early death in elderly cancer patients.

Conflicts of Interest

None declared.

References


Abbreviations

AUC: area under curve
CGA: Comprehensive Geriatric Assessment
CV: cross-validation
J48: Decision Tree
KPS: Karnofsky performance scale
ML: machine learning
MLP: multilayer perceptron
MNA-SF: Mini Nutritional Assessment-Short Form
NB: Naïve Bayes