Short-Term Mortality Prediction in Advanced Cancer Patients Eligible for End-of-Life (EOL) Care Processes Using Electronic Health Records

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:22837768">http://nrs.harvard.edu/urn-3:HUL.InstRepos:22837768</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
inpatient consults, etc. Classifiers tested included Naïve Bayes, support vector machine (SVM), K nearest neighbor (k-NN), artificial neural nets (ANN), random forest (RF), and logistic regression. Each disease cohort was analyzed using the same training and validation samples to compare the different classifiers. Area under receiver operating curve (AUC) was used as the performance measure for all classifiers.

Results
Each of the classifiers trained using the augmented features i.e. ‘cumulative impact’ and ‘novel’ features performed better than their ‘traditional’ model counterparts. For the prostate cancer cohort, the best performing model was the RF which had an AUC of 0.895 (SD 0.011) using the augmented features and AUC 0.782 (SD 0.011) using the traditional features. For the bladder cancer cohort, the best performing model was also the RF which had an AUC of 0.934 (SD 0.011) using the augmented features and AUC 0.817 (SD 0.010) using the traditional features.

Conclusion
The incorporation of an individual patient’s augmented and time-stratified feature sets properly curated from the electronic medical record allowed for the construction of better performing classification models that can accurately predict cancer mortality within three months. Given that most relevant biomedical and clinical variables are heterogeneous, multidimensional, and non-linearly related their integration within a disease- and patient-specific clinical context through modern machine learning and informatics techniques provides opportunity for real-time point of care knowledge representation of an oncologist’s clinical intuition and complement to their clinical inference. Next steps include further integration of curated data based on palliative care expertise, such as changes in pain meds over time, interventional procedures, etc. Larger implications for this work include guiding end-of-life process improvements, policy, and resource utilization.

INTRODUCTION:
With respect to oncology, high-throughput technologies continue to provide insight into the genetic and molecular underpinnings of cancer along the disease trajectory. In so doing, it has advanced the promise of ‘precision medicine’ (Katsnelson, 2013) with the idea that this information allows for more accurate categorization and treatment of patients. However, personalized application of a patient’s disease-specific insights at the point-of-care remains a challenging task which ideally requires integration of multi-dimensional data measurements from heterogeneous sources into patient-centric models. Such sources may include demographics, tissue histology findings, imaging and biomarker diagnostics, and
conventional clinical parameters. Moreover, as technology continues to better facilitate the management of cancer as a chronic albeit complex health condition, it has prompted a focus on the personalized dimensions of care delivery. From a healthcare system perspective, as the second leading cause of death in the United States (Siegel et al., 2014), the continually rising costs of cancer care, ironically effected in part by progress of novel diagnostic and therapeutic paradigms, has been burdening health systems across the country. As such, it is within an evolving cancer care landscape that precision medicine provides promise to contain costs of treatment and resource inefficiencies (Porter, 2007; Porter 2010).

Data Heterogeneity and Informatics-driven Patient Insights: Reflecting a dynamic clinical entity, cancer is a heterogeneous disease state comprising a complex interplay of individual factors (Bielas et al, 2006; Salk et al., 2010; Loeb, 2011). Besides highlighting the complexity of available data, integration studies such as Genome-wide association studies (GWAS) as well as efforts to couple genetic data to electronic health records (EHRs) for phenome-wide association studies (PheWAS) still provide only a limited clinical context. In the post-genomic era, this has underscored the need to develop more comprehensive models that provide a more holistic networked systems view of the patient over their individual clinical course. Necessarily, this includes the wealth of data being accumulated from personal spheres such as social media and wearable monitoring devices which offer a functional perspective on patients which can complement ‘omic’ data while offering singular insights into an individual. This realization has been the foundation for the notion of predictive, preventive, personalized, and participatory (P4) Medicine (Hood et al., 2011; Price et al., 2009). In current parlance, precision medicine as endorsed by the US National Research Council comprehensively engenders this goal of bridging translational gaps and integrating the spectrum of data into point-of-care, actionable models [Figure 1] (Younesi and Hofmann-Apitius., 2013).

Figure 1: Aggregation and integration of various data into a networked systems model. (Source: Younesi and Hofmann-Apitius., 2013)
Precision in End-of-Life (EOL) Cancer Care

In the context of terminal cancer patients where treatment is by definition palliative as opposed to curative, there is an unmet need to better quantify, characterize, and anticipate each patient’s clinical trajectory at the most sensitive phase of their care, namely end-of-life (EOL) (Lorenz et al., 2004; Johnson et al., 2005; Lorenz et al., 2008). Complementing the above networked systems view, given that factors such as disease phenotypes, treatment response, patient tolerability, and clinical outcomes are a function of the combined effects of various factors, the drive of precision medicine can be further used to develop an updated disease and clinical taxonomy that can guide treatment choices and care coordination, including at EOL, based on causal rather than correlative relationships. The derivation of a new taxonomy based on a comprehensive ‘information commons’ containing raw information about individual patients and from which meaningful relationships are derived and represented within a knowledge network [Figure 2] (NRC, 2011) can help to address the existing gap in optimal EOL care. Fundamentally, however, it is crucial to first be able to define and identify those patients that are in clear need of dedicated EOL care coordination.

**Figure 2**: Integration of data spectrum into a knowledge network enabling tailored taxonomic classification. *(Source: Committee on A Framework for Developing a New Taxonomy of Disease, 2011)*

Predictive Modeling of EOL

At present, mortality prediction tools are not routinely used in the care setting. In part this is due to the fact that identification of high-risk patients in the clinical literature relies on logistic regression to predict mortality, using small datasets and small feature sets (Yourman et al., 2012). However,
biomedical and clinical applications alike require more sophistication given competing needs. Taken from a practical, actionable perspective, classifiers used for predictive modeling are ideally feature-sparse, modeled to nonlinear feature interaction, accurate, and scalable. Assumptions of linear regression models do not capture or appropriately model the multi-dimensionality and complexity of latent features, thereby resulting in biased coefficients and models that may be too complex and overfit or too simple and underfit, either way giving poor predictions (King and Zeng, 2001). Conversely, more technically reliable classifiers employing machine learning methodologies in the analysis of biomedical and clinical data have been applied to robust data such as that obtained from intensive care unit admissions; however, these either do not appropriately represent the majority of encounter data which characterize typical patients’ data footprint, or they use data which are otherwise not readily collected for patients outside of a monitored ICU setting (Slaughter et al., 2012; McMillan et al., 2012; Ghassemi et al., 2014).

The three pillars of an attractive machine learning model which can be generalized to the clinical setting are interpretability, accuracy, and efficiency. Laboratory results, medications, and diagnoses, generally have a clear structured format which contribute to a knowledge network. Moreover, certain structured data can be characterized to form metadata to enrich an algorithm’s pillars. In the same vein, unstructured data such as contained in clinic notes, radiology and pathology reports, and procedure notes contain a wealth of information which can be extracted, structured and characterized to also contribute to a knowledge network. Meeting the inherent challenge of knowledge curation and management of unstructured data speaks to the parallel roles of natural language processing techniques, careful pre-processing, and engineering of features from clinical medicine and contextual subject matter expertise.

Within the context of an oncology clinic treating advanced cancer patients, a fundamental question is whether appropriate patient cohorts are receiving the tailored EOL care they require. This may include palliative care or hospice consultations, discussion of goals of care, or selection of treatment alternatives based on changing lab parameters, multiple lines of chemotherapy, or specific symptomatology, all of which are contained in the EHR and registry. As such, the objective of this project was to explore the performance of several machine learning (ML) classifiers to develop an algorithm that improves upon prior studies utilizing ML methodologies to answer similar questions (Hosseinzadeh et al, 2013) by predicting metastatic solid-tumor oncology patients receiving palliative chemotherapy who are anticipated to expire within three months.
METHODS:

Data Segmentation (as per methodology in Makar et al. (2015)):

Data was compiled on 22,700 and 7,300 adult prostate and bladder cancer patients with Stage IV disease, respectively. The patients had received care at Dana-Farber Cancer Institute (DFCI) and Brigham and Women’s Hospital (BWH) between 2004 and 2014. The three data sources included the Oncology Data Retrieval System (Onco DRS) which is a data repository containing outpatient clinic information; the institutional cancer registry, and the Research Patient Data Registry (RPDR) which is a centralized clinical data registry which gathers data from various hospitals within the Partners HealthCare network; the data for this study used BWH specific patient data. Patients were identified based on being alive at time $t_0$, which was defined as 3 months prior to documented date of decease for each patient, selected based on a conservative clinical expectation of a time point where EOL services are necessary and based on the fact that patients must have a life expectancy of no more than six months to be eligible for hospice benefits.

Each case was distributed into non-mutually exclusive groups based on presence of major comorbidities as conventionally used in health services literature (Gagne et al., 2011; Makar et al., 2014). Inclusion in a disease group was based on whether or not a patient had been diagnosed with the disease in the year prior to $t_0$. Besides cancer, my choice of significant comorbidities was limited to congestive heart failure (CHF), chronic kidney disease (CKD), chronic obstructive pulmonary disease (COPD), and cerebrovascular accident (CVA) which were used to assess mortality rate prediction across a spectrum of pathologies. For each disease group, cases were equally divided into training and validation datasets. The ‘severity’ of conditions including co-morbidities and functional performance was documented in time segments using a one-year look-back period before documented patient death, in order to more accurately characterize patients’ clinical state at $t_0$ i.e. at 3 months before death. To compare the relative performance improvement, a recently developed and well performing model for mortality prediction was used (Gagne et al., 2011) based on the combination of conventional Elixhauser and Charlson indices using administrative and demographic data, in addition to a 20-variable comorbidity set. The ‘traditional’ feature set which was extracted based on Gagne criteria was provided by a tally of International Classification of Disease (ICD-9) codes accounting for any of these variables over the $t_0$-12month look-back period (Wennberg et al., 2013; Makar et al., 2014), indicated by Boolean indices of ‘0’ or ‘1’ for ‘absent’ or ‘present’, respectively.

A ‘cumulative-impact’ feature set representing severity of the above comorbidities was compiled using a time-segmented tally of encounters in the 1-3 and 3-12 months prior to $t_0$. The comorbidity variables were represented using count variables based on a tally of the number of documented encounters referencing each comorbidity in the two designated time periods: $\sum\{1-3\}$ months, and $\sum\{3$-
12) months prior to t0. Note: these time periods have been empirically validated showing marked increase in the number of total diagnoses at three months before death (Makar et al., 2014).

Lastly, the ‘novel’ features used to augment the above data to better classify patients included cancer stage at initial diagnosis, cancer grade, functional status (Inouye et al, 1998; Lee et al., 2006), meta-data accounting for therapeutic-line of cancer treatment, ED visits, non-elective hospitalizations, inpatient consults, visiting nurse services, skilled nursing facilities placement, and medical equipment claims (e.g., walker, wheelchair, home oxygen) (Table 1). Classifiers tested included Naïve Bayes, support vector machine (SVM), K nearest neighbor (k-NN), artificial neural nets (ANN), random forest (RF), and logistic regression. Each disease cohort was analyzed using the same training and validation samples to compare the different classifiers. Area under receiver operating curve (AUC) was used as the performance measure for all classifiers.

The data were extracted from unstructured sources using NLP tools of regular expression matching and negation detection (regexm and negex in Python, R). R packages used included tm (general text mining) and RWeka (tokenization and stemming) and word cloud is a to visualize text frequencies, including relative frequencies.

Table 1: Traditional and Novel Feature Sets (for Prostate and Bladder cancer cohorts)

<table>
<thead>
<tr>
<th>Disease State</th>
<th># of Cases</th>
<th>Traditional Feature Set: *documented with dummy variable</th>
<th>Cumulative Impact and Novel Feature Set: *documented with two-count variable for each group: one for 1-3 and another for 3-12 months</th>
</tr>
</thead>
</table>
| Prostate Cancer       | 22,700     | ○ Age  
○ Sex 
○ Marriage Status 
○ Comorbidity Disease groups: Alcohol abuse, deficiency anemia, cardiac arrhythmias, coagulopathy, complicated diabetes, dementia, fluid & electrolyte disorders, hemiplegia, HIV/AIDS, hypertension, liver | ○ Race  
○ Functional status: Aerodigestive-GI/GU, aerodigestive-nutrition, aerodigestive-respiratory, performance status (ECOG score), cognition, sensorium and speech.  
○ Durable medical equipment: bed, cane/walker, O2 and wheelchair. |
| Bladder Cancer        | 7,300      |                                                         |                                                                             |
disease, metastatic cancer, CHF, psychosis, COPD, chronic pulmonary disease, peripheral vascular disorder, renal failure, tumor and weight loss.

- **Utilization:** Home health days, inpatient days, skilled nursing facility admissions, inpatient admissions, clinic and ER visits.
- **Cancer stage at initial diagnosis:** Stage IV vs. other
- **Cancer grade**
- **Therapeutic line:** 1st-line vs. 2+ lines.

**ANALYSIS (as per methodology in Makar et al., 2014):**

Six classifiers were implemented using R version 3.1.0. Naïve Bayes (klaR package), support vector machines (SVM) (kernlab package), K nearest neighbors (k-NN) (caret package), artificial neural nets (ANN) (nnets package), random forests (RF) (randomForest package), and logistic regression (stats package), both with L1 regularization (‘lasso’) (glmnet package). The prostate and bladder cancer disease cohorts were analyzed separately using the same training and validation samples to compare between the above classifiers. Area under receiver operating curve (AUC) was used as the performance measure for all classifiers. Results are provided for the validation dataset. A greedy search algorithm was applied to each cohorts training sample to identify optimal tuning parameters. Default parameters were used first, then each was changed at a time. A final parameter was chosen based on the value that maximized AUC of the tuning sample. This optimized value was then used in tuning each successive parameter. For each classifier and each disease cohort an optimized model was trained using the training dataset then applied to the testing dataset to predict outcomes calculating AUC and a bootstrapped standard deviation.

**RESULTS:**

Mortality rates in the training samples were 37.8%, for the bladder cancer, and 23.4% for prostate cancer cohorts respectively. By comparison, the validation sample rate were 37.3% and 22.8% respectively. The AUC for the validation sets using the augmented cumulative impact and novel feature sets are listed in Table 2 and Table 3 for the bladder and prostate cancer cohorts, respectively. As anticipated, each of the classifiers trained using the updated features performed better than their traditional model.
counterparts. While not necessarily anticipated, the best performing model was the RF using the novel and cumulative impact features.

### Table 2: AUC and bootstrapped SD of the Bladder cancer validation dataset for the best performing combination of tuning parameters and sampling techniques for each family of classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Augmented Variables</th>
<th>Traditional Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>SD</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.934</td>
<td>0.011</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.865</td>
<td>0.010</td>
</tr>
<tr>
<td>Naive bayes</td>
<td>0.882</td>
<td>0.011</td>
</tr>
<tr>
<td>Neural nets</td>
<td>0.819</td>
<td>0.012</td>
</tr>
<tr>
<td>SVM</td>
<td>0.837</td>
<td>0.011</td>
</tr>
<tr>
<td>Log Regression</td>
<td>0.789</td>
<td>0.013</td>
</tr>
</tbody>
</table>

### Table 3: AUC and bootstrapped SD of the Prostate cancer validation dataset for the best performing combination of tuning parameters and sampling techniques for each family of classifiers

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Augmented Variables</th>
<th>Traditional Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>SD</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.895</td>
<td>0.011</td>
</tr>
<tr>
<td>Lasso</td>
<td>0.840</td>
<td>0.010</td>
</tr>
<tr>
<td>Naive bayes</td>
<td>0.879</td>
<td>0.011</td>
</tr>
<tr>
<td>Neural nets</td>
<td>0.781</td>
<td>0.012</td>
</tr>
<tr>
<td>SVM</td>
<td>0.802</td>
<td>0.011</td>
</tr>
<tr>
<td>Log Regression</td>
<td>0.790</td>
<td>0.013</td>
</tr>
</tbody>
</table>

**DISCUSSION:**
In a large cohort of advanced bladder and prostate cancer patients, this study was able to show the significance of applying machine learning methodologies to integrated patient data extracted from readily available clinical data repositories i.e. EHR and tumor registries. Intuitively, oncologists are able to roughly estimate patient survival within 2-3 months based on a host of patient-specific parameters, disease-specific characteristics, and treatments rendered. Importantly, there are subtle, yet consequential clinical insights which are correlated directly with factors such as the stage of disease at time of presentation, and the initial choice of treatments, and subsequently the specific sequencing of treatments. Such factors speak to the aggressiveness of a disease course, as well as the expected prognosis. Notably, these insights are not captured in structured fields within the clinical databases. However, they are embedded within clinical notes and treatment summaries. Cancer, with its complex and evolving disease states, represents a prime example of where systems’ modeling is necessary to provide accurate disease insight at any given time. Of course, machine learning algorithms are fundamentally dependent on the quality of the data analyzed, and the context in which it is applied. The preprocessing phase of this study was the most crucial and time consuming. Given the anticipation that nuanced patient-, disease-, and
treatment-data would allow for more accurate prediction of mortality, the main task was to apply domain-specific subject matter expertise to create a knowledge framework which in turn would allow for novel knowledge generation based on metadata.

In the application of various machine learning algorithms, the RF classifier performed the best at using the data insights provided by the augmented data set containing cumulatively tallied and novel features. In considering this outcome, it is necessary to appreciate the differences between classification algorithms and the possibility that different types of models may perform better given different tasks. Linear classifiers (Yuan et al., 2012) are highly scalable (Fan et al, 2008), naturally interpretable, and feature-sparse when paired with L1 regularization (Efron et al., 2004). However, given that linear classifiers cannot navigate nonlinear associations, they have limited accuracy when applied to more complex, lower-dimensional data (Chen et al., 2014). Linear classifiers would not be expected to adequately serve the purpose of this study given that they separate events and non-events using a hyperplane, which are often not linearly separable. The kernel trick can be applied to linear classifiers to learn nonlinear decision boundaries, but this is at the sacrifice of their natural interpretability and scalability (Schölkopf, 2002). In contrast, nonlinear classifiers are not limited in the same way. Deep neural nets are highly nonlinear and scalable, but their interpretability is problematic and they do not naturally incorporate feature sparsity (Hinton et al., 2006). Tree classifiers, such as Random Forests (Breiman, 2001) may have similar issues although it performed well using this study’s datasets due to its facility with large feature sets as tallied over time. A specific facet of RF algorithms as applied to our relatively rare target outcome of death, is that each tree can be ‘grown’ on alternate data subsets which impact the outcome of interest. This is important to prevent bias or loss of data when down sampling or over-fitting when up-sampling (Makar et al., 2014).

On the whole, the use of the augmented features containing the cumulatively tallied as well as the extracted novel features improved the performance of each classifier. Not surprising, this provides validation for the clinically intuitive understanding that an increasing number of high impact clinical instances does signal a change in clinical course and a patient’s prognosis. Simply documenting at such an instance took place at some point in time does not properly reflect the progressive nature of a patient’s system as a function of the disease and treatment trajectory. There are other features which can be incorporated in future iterations of this work, including tally of palliative care processes. Specifically, factors such as the use of specific symptom-control medications i.e. pain, nausea, constipation, the number of different pain medications, changes in doses or formulations, and interventional pain procedures each provide an added nuanced view of a cancer patient’s clinical course. Importantly, the extent to which such features may improve a classifiers performance is necessarily dependent on meticulous preprocessing of data to allow proper algorithmic integration with the above established
features in discrete time periods. Again, the understanding of the fact that most terminally ill cancer patients who are in the EOL phase of care will require medications for symptoms control; however, the tailored adjustment in use in the weeks to months preceding death can contribute significantly to a classifiers performance during this EOL phase, which in turn can help to guide other treatment decisions and discussions with a patient.

With regard to the concern over class imbalance given that death constitutes a relatively smaller portion of the data, classification algorithms applied to rare outcomes may normally not work well given their aim to minimize the overall error rate. However, this potential limitation of class imbalance was averted in this study given the homogeneity of the patient populations being those with specific cancers, all of whom had advanced stage disease. Another contributing factor is that the time-relative occurrence of clinically relevant events were more evident in the 3 month period before death, thereby providing for more accurate and consistent prediction of death.

In considering the use of supervised machine learning it must be emphasized that the determination of salient data features for extraction, and the characterization of different data to provide metadata should not be an arbitrary process. Domain expertise necessarily plays an important role in this study, especially from the perspective of interpretability and the goal of employing such predictive algorithms in the clinical setting. The alternative approach of unsupervised learning using more formalized feature selection approaches may provide for other feature identification and selection. This would be a worthwhile endeavor to compare and contrast the performance of such developed algorithms, and its application within clinical settings.

**CONCLUSION:**

The incorporation of an individual patient’s augmented and time-stratified feature sets properly curated from the electronic medical record allowed for the construction of better performing classification models that can accurately predict cancer mortality within three months. Given that most relevant biomedical and clinical variables are heterogeneous, multidimensional, and non-linearly related their integration within a disease- and patient-specific clinical context through modern machine learning and informatics techniques provides opportunity for real-time point of care knowledge representation of an oncologist’s clinical intuition and complement to their clinical inference. Next steps include further integration of curated data based on palliative care expertise, such as changes in pain meds over time, interventional procedures, etc. Larger implications for this work include guiding end-of-life process improvements, policy, and resource utilization.
REFERENCES:


