Lessons from the Development and Deployment of an Interactive Oncological Risk Estimator

Nafiul Nipu*
University of Illinois Chicago

L. V. van Dijk†
University Medical Center Groningen

Guadalupe Canahuate‡
University of Iowa

C. David Fuller§
University of Texas MD Anderson Cancer Center

G. Elisabeta Marai¶
University of Illinois Chicago

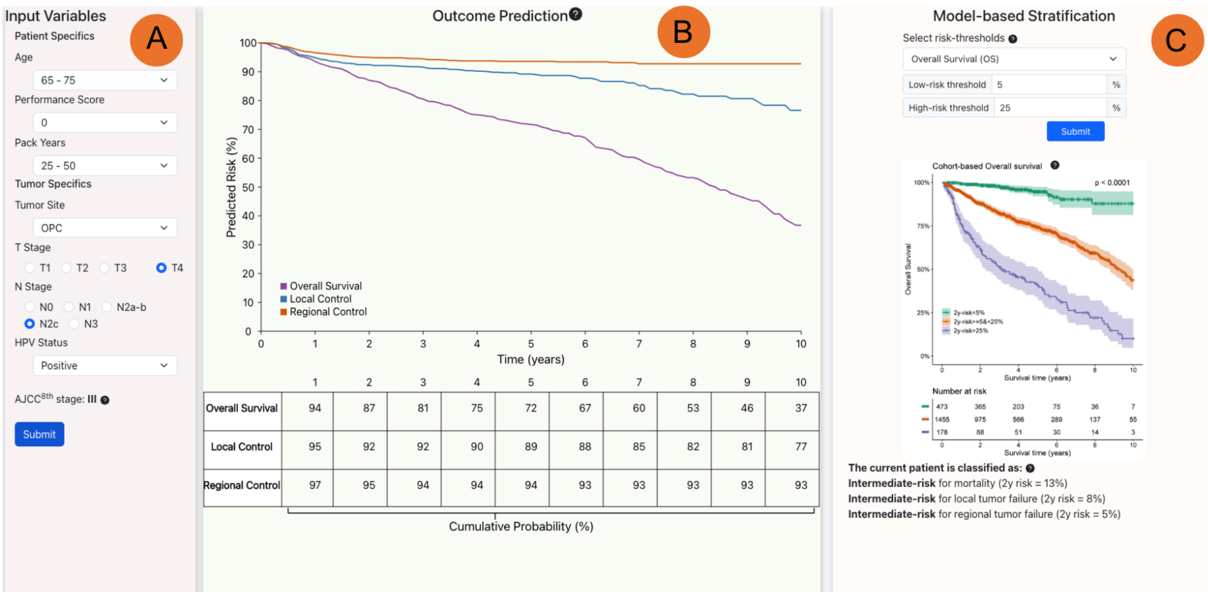

Figure 1: Interactive Oncological Risk Estimator. (A) Panel for inputting the current case features and for calculating the corresponding disease stage. (B) 10-year prediction of the probability for overall survival (OS), local control (LC), and regional control (RC), and outcome table showing cumulative probabilities per year over those 10 years. (C) Stratified risk probability for the corresponding cohort over 10 years, based on client-selected thresholds and the current patient risk. Question markers link to the published AI models behind the calculation.

Abstract

In the precision medicine paradigm, oncological treatment leverages complex ensemble datasets of similar patients to estimate the outcomes for a current patient. A key challenge is developing and deploying easy-to-understand AI predictive models for the outcomes of a specific patient, based on patient data from multiple institutions. We describe the lessons learned from the development and deployment of an interactive dashboard to support the analysis of individual head and neck cancer patient outcomes based on cohort data. As required by the project, the dashboard design aims to handle a large client base. The dashboard combines an AI solution with a multi-view interface featuring domain-specific plots to facilitate the visual analysis of patient outcomes and to quickly stratify new patients into risk groups. A year after the successful public deployment of the dashboard, we evaluate it with clinician domain experts. We report the feedback and we reflect on the lessons learned through this experience.

Index Terms: VA-machine intelligence for healthcare data visualization, Human-centered AI for health decision-making, Dashboard, Risk Stratification, Precision Medicine.

1 Introduction

Under the precision medicine paradigm in oncology [28], clinicians seek to tailor treatment for a current patient based on data collected from cohorts of similar patients. In head and neck cancer (HNC) treatment, such patient cohorts are an instance of heterogeneous multivariate spatio-temporal data, where clinical and demographic information is provided by electronic health records, and outcome predictions are temporal data over several months to several years.

While several visual analysis systems support cohort analysis [14, 15, 50], relatively little work has been done in supporting the outcome analysis of a single-patient based on cohort data [42]. This type of analysis also requires a mix of human expertise and machine learning models for predicting treatment outcomes. These models would be using large data with both clinical and demographic attributes. Last but not least, there is a paucity of interactive tools that can effectively handle a large client base, spanning multiple countries and multiple centers, who are focused on understanding outcomes for a specific patient in the context of a similar

*e-mail: mnipu2@uic.edu
†e-mail: l.v.van.dijk@umcg.nl
‡e-mail: guadalupe-canahuate@uiowa.edu
§e-mail: cdfuller@mdanderson.org
¶e-mail: gmarai@uic.edu

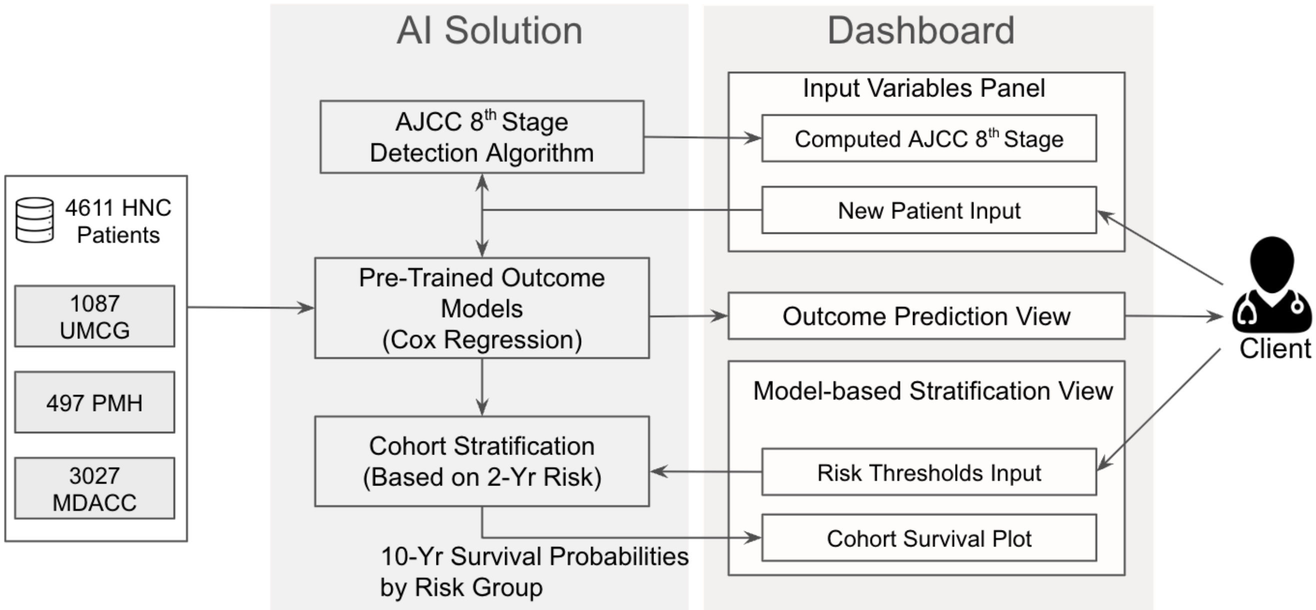

Figure 2: System architecture. The system integrates a pre-trained outcome model based on 4611 HNC patients, alongside a 2-year cohort mortality risk based on 3027 MDACC patients, and an AJCC-8^{th} edition stage detection algorithm. Clinicians interact with the dashboard by inputting new patient data, which is used to compute AJCC stage, overall 10-year survival outcomes, and visualize the results. The input for risk thresholds is then used to retrieve 2-year mortality outcomes and visualize risk-based cohort stratification.

cohort, and who generally have low data visualization literacy.

This work describes the design, development, and deployment of an interactive dashboard to support the analysis of individual patient outcomes in head and neck cancer (HNC) patient data. A data dashboard is a type of graphical user interface that provides at-a-glance views of key performance indicators relevant to a particular objective (in our case, mortality risk calculators). The dashboard leverages machine learning models and large HNC patient datasets from multiple treatment centers. It predicts the outcome and stratification of a new individual patient into high-risk or low-risk. A multi-view interface leverages domain-specific encodings to support visual analysis of patient outcomes. The contributions of this work are: 1) a description of the application-domain data and tasks, with an emphasis on analyzing individual new patient survival outcomes and stratification risks; 2) a description of the design and implementation of the cohort-based dashboard; 3) an evaluation with domain experts a year after public deployment; and (4) a discussion of the lessons learned.

2 Related Work

EMR Cohort Visualization. In cancer [32, 37] and HNC [49], patient longitudinal electronic medical records (EMR) are visualized using time-series visualization techniques such as line graphs [18], parallel coordinate plots [21] or stacked bar charts [1]. Many visual computing frameworks have been introduced to show outcome trajectories of cohort patients [34] and analyze outliers and trends for dense clinical data [23, 44, 43, 45, 46, 47, 26]. In contrast to our focus, these approaches do not support analyzing temporally individual new patients.

Several other cohort visual analysis systems have been proposed to identify disease evolution [19, 48], survival risk analysis [28, 3, 10], cohort history [11, 7], and attribute comparison [30]. These works also leverage conventional visualization techniques such as time-series plots [16], bar charts [25], histograms [5], radial plots [17], scatterplots [14], and matrices [13]. We also visualize patient cohort data, however, with an emphasis on individual new patients, their survival outcomes, and risk thresholds.

Dashboards in Medical Visualization. Dashboards are commonly used in various business organizations [24] and personal applications [33] due to their easy-to-use visual representations. In healthcare, hospitals use dashboard design tools to monitor quality improvements across hospitals, improve patient care outcomes [12, 6], enable surveillance of potential drug events [41], and monitor general health management [31]. Healthcare is particularly interested in designing effective dashboards to achieve clinical relevance, efficiency, and an optimal end-user experience. Various evaluation criteria, such as user customization, knowledge discovery, information delivery, and user interaction, are crucial for dashboard design. A single-screen, minimalist visual design with no scrolling is also necessary to meet these criteria [22]. Our dashboard design was informed by these factors.

3 Methods

3.1 Project Setting

The project was developed over the course of a year through a remote, interdisciplinary collaboration. The core group included two radiation oncologists, one data mining expert, and two visual computing researchers. The larger group included clinicians and data scientists at three medical centers who partnered to pool data and create the AI models.

3.2 Activity Analysis

Our visual computing framework uses an Activity-Centered Design (ACD) approach, an extension of the Human-Centered Design Paradigm, focusing on user activities and workflows [29]. Using this approach, the core team met multiple times to identify functional requirements, prototype, evaluate visual encodings, and provide feedback on necessary changes. We summarize the resulting workflow and functional requirements for the project as tasks below, where local failure denotes the recurrence of cancer at the original site of the primary tumor or within the surrounding area treated with local therapy, such as surgery or radiation, and regional failure

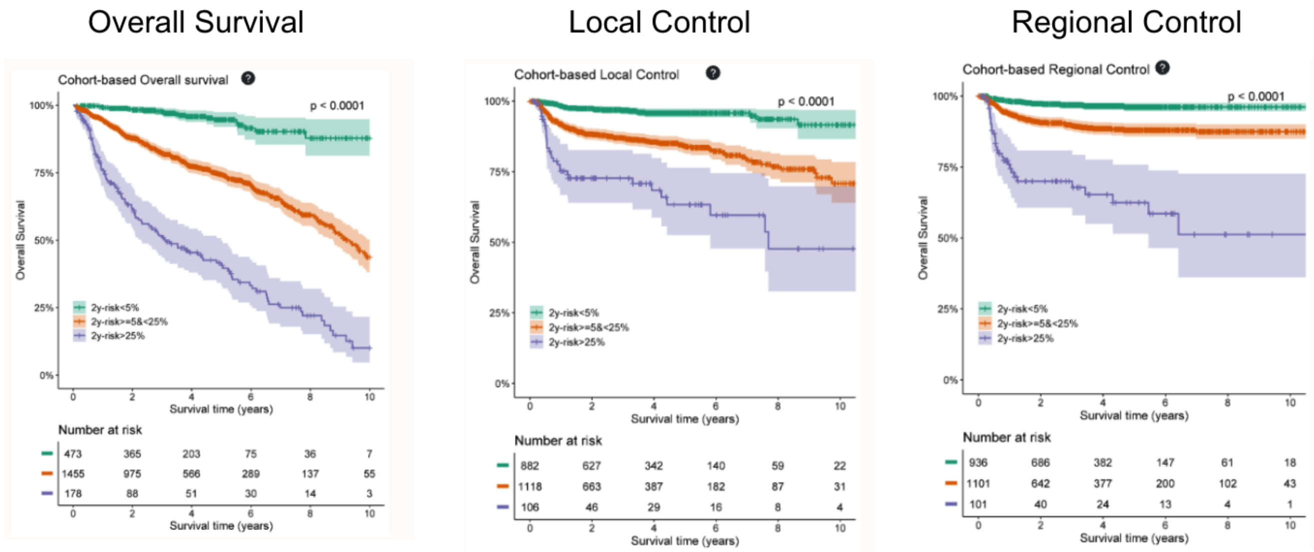

Figure 3: Kaplan-Meier plots showing survival probabilities of overall survival, local control, and regional control of patient subcohorts, with a low-risk threshold of 5% and a high-risk threshold of 25%. The tables at the bottom indicate the number of patients at risk over ten years.

refers to the recurrence or persistence of cancer within the region surrounding the primary tumor site, specifically involving the regional lymph nodes or nearby tissues, signifying that the cancer has spread or returned beyond the initial tumor location, but not to distant parts of the body:

- T1. Input the data for a new individual patient, calculate and show their HNC stage.
- T2. For the new individual patient, use an AI solution to calculate and show the following predicted outcomes over the next ten years: survival probability, local tumor failure probability, and regional tumor failure probability.
- T3. Given a user-specified set of risk thresholds for each outcome, calculate and show the risk group for the individual patient, along with the patient subcohorts used for the calculation.

Non-functional requirements included visual scaffolding that would include visual encodings familiar to the oncologists [27], scalable design that can display patient outcomes along with a large cohort information, and 24/7 online availability for clinicians at multiple sites.

3.3 Data Abstraction

In this study, we aim to investigate individual new patient outcomes based on an AI solution leveraging machine learning models. These models are trained using a total of 4611 HNC patient data collected at the MD Anderson Cancer Center (MDACC) (3027 patients) in the U.S., University Medical Center Groningen (UMCG) (1087 patients), and Princess Margret Hospital (PMH) (497 patients) in the Netherlands. Patients with proven squamous cell carcinoma of the head and neck, who received radiotherapy with or without chemotherapy as definitive or adjuvant treatment, and had no prior head and neck radiation, were included in these cohorts. The patients were treated at MDACC, UMCG, and PMH between 2001 and 2020, with prescribed tumor radiotherapy doses of 60-72 Gy.

The clinician team extracted from the Electronic Health Records structured data several features for each patient. Each patient in the datasets had several categorical clinical attributes such as gender, smoking status, performance score, T-stage, N-stage, tumor site, HPV status, tumor stage according to the AJCC-8th (American Joint Committee on Cancer 8th edition) standard, and numerical attribute age. For a new patient, the models used these attributes as input, and generated temporal data of 10 years of outcome prediction based on overall survival, local control, and regional control, and categorical attributes of the patient's 2-year-risk threshold.

3.4 AI Solution

The AI solution we leverage implements three outcome models based on Cox regression (Fig. 2). These models are used to predict the following outcomes: overall survival (OS) probability, local tumor failure probability (LC, or local control) [35], and regional tumor failure probability (RC, or regional control) (T1).

To generate these models, the MDACC dataset was split into training and validation groups with a 60:40 ratio. External validation was carried out on the UMGC and PMH datasets. Patients with missing attributes were used for the training set, and only complete cases were used for validation. The step-wise forward variable selection method (i.e., the process of choosing the most relevant attributes to include in a regression model) was used to select attributes for the Cox regression models of OS, LC, and RC. This process was carried out on 10 imputed datasets using multivariate imputation (i.e., the process of replacing missing data) with predictive mean matching [38]. After analyzing the results of variable selection and inter-variable correlation, potential models were chosen and tested in the validation cohorts. OS with AJCC-8th stage, performance score, pack years, and age, LC with T-stage, performance score, HPV status, and pack years, and RC with AJCC stage, performance, and tumor site were selected as the final Cox regression models for individual patient analysis and patient cohort stratification. These models were selected based on the c-index, a metric used to evaluate risk models in patient survival analysis. For a selected patient, the regression models predict the survival probability of OS, LC, and RC over ten years.

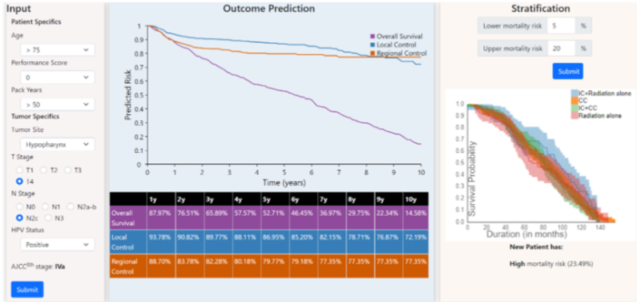

Figure 4: Intermediate design iteration of the dashboard. Note the toned-down color scheme and the additional annotations with careful font formatting in the final deployed version.

3.5 Cohort Stratification

Clinicians wished to understand how different HNC subcohorts were stratified into high-risk or low-risk, to better contextualize the predicted outcomes for their current patient (T3). To support this goal, we selected the MDACC patient cohort and calculated the 2-year mortality risk using the Cox regression models for OS, LC, and RC. Then, we stratified the patients into three groups - high-risk, intermediate-risk, and low-risk based on their risk levels. Using the risk threshold determined by the clinicians for low-risk and high-risk groups, we calculated the OS, LC, and RC survival probabilities along with the number of patients at risk per group over ten years for each risk group (Fig. 2).

3.6 AJCC-8th Stage Detection

When evaluating a cancer patient, it is important to determine the disease stage, which plays a significant role in deciding the appropriate treatment plan for them. The American Joint Committee on Cancer (AJCC) 8th edition provides a standard-of-care tumor, node, and metastasis (TNM) staging system, which can be used to describe the extent of cancer in a patient's body. The stage ranges from one (I) to four (IV), with IV being the most advanced stage. It is determined based on the size and location of the tumor, its spread to regional lymph nodes, and whether it has metastasized (spread) to distant organs or tissues [2]. In order to help clinicians identify the AJCC-8th stage of a new patient, we use a procedure based on the T-stage, N-stage, tumor site, and HPV status. For a patient whose HPV status was unknown, we considered the unknown status as negative. For a selected patient, the procedure first checks the tumor site and the HPV status. It then determines the AJCC-8th stage based on the T-stage and N-stage of the patient (T1).

4 Visual Front-End

We designed the front-end based on the specific instructions and suggestions from the core group, with many iterations using a range of colors and formats (Fig. 4). While we had access to a larger group of radiologists, we worked closely with one of them (LVD) as our primary contact point for the project. Based on guidance from LVD, the front end final design consists of three main views: 1) Input Variables (Fig. 1.A) Panel assists in adding the clinical attributes of a new patient; 2) Outcome Prediction View (Fig. 1.B) supports the examination of patient outcomes over time; 3) Model-based Stratification View (Fig. 1.C) allows exploring cohort-based risks over time. In our prior visual design experience, we have followed the Tufte principle of maximizing information density [39], by populating and displaying each panel of a visual analysis system as soon as the interface loads. In this case, however, as explicitly requested by the clinicians, only the input variables panel is displayed when the dashboard loads. The remaining panels are only shown after the data is inputted or selected in the input panel, to emphasize the workflow aspect of the dashboard, and the connection between the input data and the outcomes.

4.1 Input Panel

This panel allows inputting the attributes of a new patient (T1). The attributes are grouped into two categories: patient specifics such as age, performance score, and smoking pack per year, and tumor specifics such as tumor site, T-stage, N-stage, and HPV status (Fig. 1.A). Instead of using medical image data or patient chart notes, our focus is on presenting structured features extracted from the Electronic Health Record (EHR) data by our domain scientists. These features were chosen by domain experts based on historical data related to HNC risk and for the purpose of identifying robust treatment outcomes based on patient and tumor-specific clinical variables. Providing patient and tumor-specific information will update the outcome prediction view and model-based stratification view. In addition, this panel interactively calculates and displays the patient's AJCC-8th stage in real-time (T1).

4.2 Outcome Prediction View

This view shows the predicted risk probability of overall survival (OS), local control (LC), and regional control (RC) over the course of ten years using line curves based on the clinical attributes of the patient (T2) (Fig. 1.B top). The lines are color-coded according to their respective outcomes. Hovering over the lines or the legend will highlight the outcome in order to enhance readability. In addition, this view also includes an outcome table that presents the cumulative probabilities per year over ten years (T2) (Fig. 1.B bottom). This makes it easier for clinicians to identify the survival probability of the current patient. Initially, each row in the table was color-coded based on its respective outcome prediction curve (Fig. 4). However, this approach was found to impede rather than aid the interpretation of outcome probabilities, so we decided to remove the color of the table for better readability. Instead, we added the functionality to highlight the survival curve in the plot when hovering over an outcome in the table.

4.3 Model-based Stratification View

This view helps to analyze the survival chances of a group of patients over a period of ten years (T3). It shows the overall survival probabilities for a selected low-risk and high-risk threshold, using Kaplan-Meier survival curves [28] (Fig. 1.C). These static survival curves are commonly used in the application domain to explain therapy outcomes across different patient cohorts by showing survival probability and standard deviation over time. The view shows three plots that correspond to the survival outcomes of patients with low, intermediate, and high risks. Users can select the risk probability of OS, LC, or RC survival (Fig. 3). Each risk group is color-coded in the graph, with a lighter ribbon of the same hue surrounding each plot representing the 95% confidence interval of the prediction. Additionally, the view provides a table of the number of patients at risk (i.e., the total number of survivors at the beginning of each year) in each group over ten years (Fig. 3 bottom). Clinicians can utilize this view to identify the risk level of the current patient. Selecting a low-risk and high-risk threshold will automatically update the risks of the patient accordingly (T3) (Fig. 1.C bottom). LVD estimated this view would have the highest learning curve for clinicians. As a result, a single set of survival plots is shown at any time.

4.4 Implementation and Release

Our computational back-end was built using R and Python with Flask and Numpy, and the front end was developed using JavaScript with D3.js [8], and React libraries (Fig. 2). The dashboard is publicly accessible via https://risk-calculator.evl.uic.edu/, and it's designed to handle large user traffic on multiple devices, including laptops, desktops, and tablets, although not phones.

5 Evaluation and Results

We conducted several demonstrations of our system and received valuable qualitative feedback from our main contact point throughout the development of the dashboard. We demonstrate below the functionality of the dashboard. Furthermore, at this point the dashboard has been deployed and used by clinicians across multiple sites and countries. We report feedback collected using the system usability scale (SUS) [9].

5.1 Use Case Demonstration

This use case investigates the outcomes of a new patient, in terms of their AJCC-8th HNC stage, overall risks, and patient cohort context. A radiation oncologist input information about the new patient (T1), who was in the age group of 65-75, had a performance score of 0, and had been smoking for 25-50 years. The patient's tumor was located in the oropharynx (OPC) and was categorized as T4 T-stage, N2c N stage, and HPV positive. The AJCC-8th stage calculation (T1) (Fig. 1.A) indicated stage III. The oncologist explained that this was accurate because the tumor was large (T4 T-stage) and had spread to the lymph nodes (N2c). Additionally, she noted the cancer was located in the oropharynx, and that this type of cancer can develop on the rear third of the tongue, the tonsils, the soft palate, as well as the side and back walls of the throat, with a spread consistent with stage III.

Next, she focused on the outcome prediction view and noticed that the overall survival decreased significantly over time (T2) (Fig. 1.B). The patient had a 94% chance of survival in the first year, and above 80% up to the third year, but it decreased to 37% in the tenth year. She noted that as the patient was in an advanced stage, this made sense. While examining the local control and regional control, she found that the patient had over 90% local control for up to 4 years and regional control for up to 10 years.

Finally, the expert explored the patient cohort based on the risk thresholds, so she focused on the model-based stratification panel (T3) (Fig. 1.C). She selected overall survival for a risk threshold of 5% low and 25% high. The Kaplan-Meier plot confirmed that low-risk patients had better survivability than high-risk patients. She also noted that a higher number of patients had survivability in the first few years, which was correct, as patient survivability would decrease over time. Based on the low-risk threshold, the current patient had an intermediate risk for mortality (2yr risk 13%), local tumor failure (2yr risk 8%), and regional tumor failure (2yr risk 5%) (T3). Due to the advanced stage of cancer and intermediate mortality risk, the expert recommended that the patient start the treatment right away.

5.2 Deployment Questionnaire Evaluation

A year after deployment, we collected feedback from eight clinicians at the three medical centers, who responded to our survey. We presented thirteen questions based on the System Usability Score guidelines (SUS) [9], where the first 12 questions used a Likert scale from one to five and the last question (likelihood of recommending the dashboard to a colleague) used a Likert scale from one to ten. Additionally, we included optional questions about the most liked features and points for improvement.

The dashboard obtained a usability score of 87.81, which indicates clinicians' high satisfaction (SUS scores between 80-90 are equivalent to an adjective rating of excellent [4]). In particular, evaluators agreed that they would use the dashboard frequently (M=4±.93), that the dashboard was easy to use (M=4.75±.46), that the functions were well integrated (M=4.63±.52), and that most people would learn to use it very quickly (M=4.63±.52). Evaluators were also confident in using the dashboard (M=4.13±.99). Furthermore, evaluators disagreed on the following items: the dashboard was unnecessarily complex (M=1.13±.35), they would need support from a developer to use the dashboard (M=1.38±0.52), the dashboard had too much inconsistency (M=1.38±.74), the dashboard was cumbersome to use (M=1.38±.52), and they need to learn many things before starting to use it (M=1.75±1.16). In addition, evaluators agreed that they would recommend our dashboard to their colleagues (7.88/10), that the dashboard met their expectations (M=4.38±.52), and that they got a very positive impression of the dashboard (M=4.38±.75).

The dashboard has received positive feedback for its effective layout, ease of use, and clear visualizations. Clinicians appreciated the dashboard's ability to provide individualized risk stratification and found the predicted outcomes at various time points to be valuable. The simplicity of the design, along with appealing colors and easy-to-understand metrics, was noted as a significant strength. Some minor points for improvement were also identified. One responder mentioned browser compatibility issues, noting that the right panel did not update properly in certain browsers like Edge. Additionally, it was suggested that the dashboard include further clarification on variable definitions, such as performance scores. One clinician also recommended the inclusion of treatment input options to explore potential changes in predictions. Overall, the dashboard was well-received.

In general, we observed that the clinician onboarding process tends to be very fast, as most clinicians are familiar with Kaplan-Meier and survival curves and thus found the interface intuitive. Only the stratification model required at most a 5-minute learning curve. Tutorials, workshops, or other guides were not requested or provided, although LVD provided an interactive demo when presenting the system at various venues. In terms of challenges, the most common challenge was related to some clinicians unrealistically expecting the AI solution, which had been trained on retrospective data, to be able to adapt to additional input decisions (e.g., *I would like to input chemo*). In terms of trust, there were no documented instances where the clinicians encountered a counter-intuitive prediction, as the prediction models are relatively intuitive (i.e., increasing age, smoking, or performance scores result in worse patient outcomes). From the three prediction tasks, the survival model was most trusted, because the data in this case is binary (dead or alive). In contrast, local and regional control are harder to assess, because the data (whether the tumor has recurred or not) is murkier.

6 Discussion and Lessons Learned

Our dashboard is designed to help HNC clinicians assess potential outcomes for a specific patient, in particular survival outcomes and the chances of local and regional treatment failure. The dashboard also provides the cohort information used to make these determinations. In terms of clinical impact, the current HNC radiotherapy practice relies on a "one-dose-fits-all" strategy, where all HNC patients receive the same tumor radiation doses based on clinical trials that pre-date the recent dramatic increase in HPV positive HNC patients [40]. This strategy does not account for the patient disease specifics, in particular the fact that HPV-positive HNC responds well to radiation therapy. Thus, the current approach can lead to suboptimal treatment in terms of balancing survival and radiation-induced toxicity. By calibrating the treatment of individual patients based on the risk predictions based on cohorts of similar patients, clinicians can make better decisions that improve patient outcomes.

Several important lessons emerged from this development and deployment experience:

L1. Designing for Lower Information Density, with Gradual Reveal. Most visual analytics solutions build on visual design recommendations made by Tufte [39]. We found that some of the

Tufte recommendations, such as maximizing information density, are sound in the context of statistical visual analysis, given that Tufte himself is a statistician, but are not necessarily sound in the clinical context of an AI predictive solution. According to visual analysis design principles, our initial design was rich and showed, by default, the cohort data. In contrast, the clinicians asked that the dashboard be initially unpopulated except for the input panel. Specifically, our collaborators indicated that presenting visual outcomes when the data has not been provided, or has been auto-filled, would mislead the viewer.

L2. Designing for Constrained Interaction. The dashboard also does not feature extensive interaction or exploration of the patient characteristics; these characteristics are provided through the input panel only. Even though earlier iterations allowed hovering over the plots and displaying details-on-demand, the clinicians asked for the actual values or numbers to be shown below the charts. Furthermore, the dashboard is characterized by a fixed flow of information, from input to outcome prediction, as opposed to repeated linked observations across multiple views. This type of extremely constrained interaction was also the result of clinician input. The resulting interface resembles more closely a banking application than a traditional visual analytics interface, with the explicit goal of minimizing error possibilities.

L3. Transparency in Data Sources and Published Model. One of the strengths of our AI solution is that it was built using multi-institutional data. Many AI healthcare solutions have been built in the past using single-race, single-institution datasets that did not generalize well to other populations. Due to increasing skepticism in AI solutions [42, 20], our dashboard provides explicit links to the peer-reviewed manuscript describing the datasets and the prediction model.

L4. Open Access Implications on Patient and Patient Family Access. Building a risk prediction tool that is completely open-access means it could also be accessed by patients and their families. Since the deployment of the dashboard, which was designed for use by clinicians, our group has been contacted by patients and family members of patients who had accessed and used the tool. In response to ethical concerns, e.g., the fact that, as a result, some patients might feel discouraged or experience increased anxiety, our clinician collaborators countered firmly that patients already come to appointments armed with incorrect information provided by influencers on social media. Our collaborators consider open public access to real institutional data and validated prediction models as a clear informational benefit to patients and their families.

L5. Building for Reliable Access. Once our solution was deployed, we learned that clinicians across the world were happy to use it, but would only contact us when the system was down. We designed our visual interface to be stable across different display sizes and browsers. We also encountered multiple challenges in achieving reliable access. Initially, we hosted the dashboard on GitHub Pages and used a virtual server for the AI solution. However, we faced cross-origin resource sharing (CORS) issues that made it difficult for the frontend and AI solution to communicate. To solve this problem, we moved both our frontend and backend to a single virtual server. Although this solved the CORS issue, some of our collaborators were still unable to access the dashboard. Moreover, the dashboard would occasionally go down, requiring us to restart the front and backend manually. To resolve these issues, we used Docker containers to host our frontend and backend with a proxy setting to serve the dashboard to the internet using the standard ports. This approach has enabled us to provide clinicians with a reliable and secure platform that they can access anytime and from anywhere.

L6. Tracking Access and Adoption. As in our other projects that have been adopted at more than 40 institutions across the world [36], tracking access and adoption is difficult. We know the community is using the dashboard because they occasionally contact us with requests. Since the dashboard is being used in clinical practice, we do not expect a large number of citations to our work. We did not require client registration on purpose, to make the dashboard as easy to access as possible. We also did not instrument the dashboard with cookies, which would have degraded performance. The cost of these decisions is an impaired ability to track access and adoption.

6.1 Assumptions and Limitations

The case study and the domain expert feedback demonstrate the dashboard's ability to help clinicians identify the cancer stage, survival outcomes over time, mortality rate, local and regional failure risk of a patient, and overall risk-based outcomes of their respective patient cohorts. We leveraged visual encodings familiar to our domain experts [27], facilitating wider adoption of the dashboard within the head and neck oncology community.

In terms of scalability and generalizability, the outcome prediction view can be generalized to show and compare several temporal datasets, including for other cancer patient data. There are several assumptions and limitations to the current design of our dashboard. First, the dashboard assumes users have an understanding of Kaplan-Meier plots, which may limit its accessibility to the general public. Anecdotally, the family members of patients who reached out to us were college-educated. Second, patient-specific attributes such as age and pack years are limited to a few options. Future work includes the ability to update the models and to allow clinicians to add more patient attributes to observe outcomes. Adding gradually more functionality could further enhance the system without sacrificing usability.

7 Conclusion

In this work, we presented the design, implementation, and lessons resulting from the deployment of an interactive dashboard for the analysis of individual HNC patient outcomes. We described the application domain data and workflow related to assessing the survival outcomes and stratification risks of a specific patient, within the context of precision medicine. Our dashboard leverages an AI solution to generate individual outcomes based on large HNC patient cohorts. An evaluation with domain experts demonstrates the functionality and usefulness of the resulting dashboard. Last, we believe the lessons we extracted from this experience, 1.5+ years after deployment, benefit the wider visualization-in-healthcare community.

Acknowledgments

Our work is supported by NIH NCI R01CA258827, NIH UG3 TR004501, NSF CNS-2320261, and the UIC Institute for Health Data Science Research. We thank the UIC Electronic Visualization Laboratory.

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Springer, 2011. 2

[2] M. Amin, F. Greene, S. Edge, et al. The eighth edition AJCC cancer staging manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA: A Cancer J for Clinicians*, 67(2), 2017. 4

[3] E. A. Anyimadu, Y. Wang, C. Floricel, S. Kamel, et al. PRO-Based Stratification Improves Model Prediction for Toxicity and Survival of Head and Neck Cancer Patients. *IEEE J. Biomed. and Health Infor.*, 2024. 2

[4] A. Bangor, P. Kortum, and J. Miller. Determining what individual sus scores mean: adding an adjective rating scale. *J. Usability Studies*, 4(3), 2009. 5

[5] T. Baumgartl, M. Petzold, M. Wunderlich, M. Hohn, et al. In Search of Patient Zero: Visual Analytics of Pathogen Transmission Pathways in Hospitals. *IEEE Trans. Vis. Comp. Graph.*, 27(2), 2021. 2

[6] J. Bernard, D. Sessler, J. Kohlhammer, and R. A. Ruddle. Using dashboard networks to visualize multiple patient histories: A design study on post-operative prostate cancer. *IEEE Trans. Vis. Comp. Graph.*, 25(3), 2019. 2

[7] J. Bernard, D. Sessler, T. May, T. Schlomm, et al. A visual-interactive system for prostate cancer cohort analysis. *IEEE Comp. Graph. and App.*, 35(3), 2015. 2

[8] M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE Trans. Vis. Comp. Graph.*, 17(12), 2011. 5

[9] J. Brooke. SUS: A "quick and dirty" usability scale. In *Usability Evaluation in Industry*. Taylor & Francis, 1996. 5

[10] G. Canahuate, A. Wentzel, A. S. Mohamed, L. V. van Dijk, et al. Spatially-aware clustering improves AJCC-8 risk stratification performance in oropharyngeal carcinomas. *Oral oncology*, 144, 2023. 2

[11] K. Chui, J. B. Wenger, S. A. Cohen, and others. Visual Analytics for Epidemiologists: Understanding the Interactions Between Age, Time, and Disease with Multi-Panel Graphs. *PLoS one*, 6(2), 2011. 2

[12] D. Dowding, R. Randell, P. Gardner, G. Fitzpatrick, et al. Dashboards for improving patient care: Review of the literature. *Int. Journal Med. Infor.*, 84(2), Feb 2015. 2

[13] F. Du, C. Plaisant, N. Spring, and B. Shneiderman. EventAction: Visual analytics for temporal event sequence recommendation. In *IEEE Conf. Vis. Anal. Sci. and Tech. (VAST)*, 2016. 2

[14] C. Floricel, N. Nipu, M. Biggs, A. Wentzel, et al. THALIS: Human-Machine Analysis of Longitudinal Symptoms in Cancer Therapy. *IEEE Trans. Vis. Comp. Graph.*, 28(1), 2022. 1, 2

[15] C. Floricel, A. Wentzel, A. Mohamed, C. Fuller, et al. Roses have thorns: Understanding the downside of oncological care delivery through visual analytics and sequential rule mining. *IEEE Trans. Vis. Comp. Graph.*, 30(1), 2024. 1

[16] D. Gotz and H. Stavropoulos. DecisionFlow: Visual Analytics for High-Dimensional Temporal Event Sequence Data. *IEEE Trans. Vis. Comp. Graph.*, 20(12), 2014. 2

[17] S. Guo, F. Du, S. Malik, E. Koh, et al. Visualizing uncertainty and alternatives in event sequence predictions. In *Proc. CHI Conf. Human Factors in Comput. Sys.*, 2019. 2

[18] R. L. Harris. *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, Inc., USA, 1999. 2

[19] C. W. Huang, R. Lu, U. Iqbal, S. H. Lin, et al. A richly interactive exploratory data analysis and visualization tool using electronic medical records. *BMC Med. Infor. and Dec. Making*, 15, 2015. 2

[20] L. Humbert-Vidan, S. Kamel, A. Wentzel, et al. Externally validated digital decision support tool for time-to-osteoradionecrosis risk-stratification using right-censored multi-institutional observational cohorts. *Radiotherapy and oncology*, 207, 2025. 6

[21] A. Inselberg. Multidimensional detective. In *Proc. VIZ: Vis. Conf., Info. Vis. Symp. and Parallel Rendering Symp.*, 1997. 2

[22] M. Karami, M. Langarizadeh, and M. Fatehi. Evaluation of effective dashboards: Key concepts and criteria. *The Open Medical Infor. Journal*, 11, 2017. 2

[23] M. Karpefors and J. Weatherall. The Tendril Plot—a novel visual summary of the incidence, significance and temporal aspects of adverse events in clinical trials. *J. Am. Med. Info. Assoc.*, 2018. 2

[24] B. Lea and F. Nah. Usability of performance dashboards, usefulness of operational and tactical support, and quality of strategic support: A research framework. In *Human Interface and the Management of Information. Information and Interaction for Health, Safety, Mobility and Complex Environments*. Springer, 2013. 2

[25] M. H. Loorak, C. Perin, N. Kamal, M. Hill, et al. Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients. *IEEE Trans. Vis. Comp. Graph.*, 22(1), 2015. 2

[26] T. Luciani, A. Wentzel, B. Elgohari, et al. A spatial neighborhood methodology for computing and analyzing lymph node carcinoma similarity in precision medicine. *J. Biomed. Informatics*, 112, 2020. 2

[27] G. Marai. Visual Scaffolding in Integrated Spatial and Nonspatial Analysis. In *EuroVis Workshop Vis. Analyt. (EuroVA)*, 2015. 3, 6

[28] G. Marai, C. Ma, A. Burks, F. Pellolio, et al. Precision risk analysis of cancer therapy with interactive nomograms and survival plots. *IEEE Trans. Vis. Comp. Graph.*, 25(4), 2019. 1, 2, 4

[29] G. E. Marai. Activity-centered domain characterization for problem-driven scientific visualization. *Trans. Vis. Comp. Graph.*, 24(1), 2017. 2

[30] A. Maries, N. Mays, M. Hunt, K. F. Wong, et al. GRACE: A Visual Comparison Framework for Integrated Spatial and Non-Spatial Geriatric Data. *IEEE Trans. Vis. Comp. Graph.*, 19(12), 2013. 2

[31] J. McMenamin, R. Nicholson, and K. Leech. Patient dashboard: The use of a colour-coded computerised clinical reminder in whanganui regional general practices. *J. Primary Health Care*, 3(4), 2011. 2

[32] J. Müller, V. Zebralla, S. Wiegand, and S. Oeltze-Jafra. Interactive visual analysis of patient-reported outcomes for improved cancer aftercare. In *EuroVA@ EuroVis*, 2019. 2

[33] M. Riphagen, M. van Hout, D. Krijnen, and G. Gootjes. Supporting reflective learning for daily activities using an interactive dashboard. In *ICERI2013 Proceedings*, Int. Conf. Edu., Res. Innovation. IATED, 18-20 November, 2013 2013. 2

[34] J. Rogers, N. Spina, A. Neese, et al. Composer—visual cohort analysis of patient outcomes. *App. Clin. Info.*, 10(2), 2019. 2

[35] T. Sheu, D. Vock, A. Mohamed, et al. Conditional survival analysis of patients with locally advanced laryngeal cancer: construction of a dynamic risk model and clinical nomogram. *Scientific reports*, 7(1), 2017. 3

[36] A. M. Smith, W. Xu, Y. Sun, J. R. Faeder, et al. Rulebender: integrated modeling, simulation and visualization for rule-based intracellular biochemistry. *BMC bioinformatics*, 13, 2012. 6

[37] N. Steinhauer, M. Hörbrugger, A. D. Braun, T. Tüting, S. Oeltze-Jafra, et al. Comprehensive visualization of longitudinal patient data for the dermatological oncological tumor board. In *EuroVis 2020 Short Papers*. The Eurographics Association, 2020. 2

[38] H. Toutenburg. Rubin, D.B.: multiple imputation for nonresponse in surveys. *Statistical Papers*, 31(1), 1990. 3

[39] E. Tufte and P. Graves-Morris. *The visual display of quantitative information*. Graphics press Cheshire, CT, 1983. 4, 5

[40] L. van Dijk, A. Mohamed, S. Ahmed, N. Nipu, et al. Head and neck cancer predictive risk estimator to determine control and therapeutic outcomes of radiotherapy (HNC-PREDICTOR): development, international multi-institutional validation, and web implementation of clinic-ready model-based risk stratification for head and neck cancer. *European Journal of Cancer*, 178, 2023. 5

[41] L. Waitman, I. Phillips, A. McCoy, et al. Adopting real-time surveillance dashboards as a component of an enterprisewide medication safety strategy. *J. Quality Patient Safety*, 37(7), 2011. 2

[42] A. Wentzel, S. Attia, X. Zhang, et al. Ditto: A visual digital twin for interventions and temporal treatment outcomes in head and neck cancer. *Trans. Vis. Comp. Graph.*, 2024. 1, 6

[43] A. Wentzel, C. Floricel, G. Canahuate, et al. DASS Good: Explainable Data Mining of Spatial Cohort Data. *Comp. Graph. For.*, 24(3), 2023. 2

[44] A. Wentzel, P. Hanula, et al. Cohort-based T-SSIM visual computing for radiation therapy prediction and exploration. *Trans. Vis. Comp. Graph.*, 26(1), 2019. 2

[45] A. Wentzel, P. Hanula, L. V. van Dijk, et al. Precision toxicity correlates of tumor spatial proximity to organs at risk in cancer patients receiving intensity-modulated radiotherapy. *Radiotherapy and Onc.*, 148, 2020. 2

[46] A. Wentzel, T. Luciani, L. V. van Dijk, et al. Precision association of lymphatic disease spread with radiation-associated toxicity in oropharyngeal squamous carcinomas. *Radiotherapy & Onco.*, 161, 2021. 2

[47] A. Wentzel, A. S. R. Mohamed, M. A. Naser, et al. Multi-organ spatial stratification of 3-d dose distributions improves risk prediction of long-term self-reported severe symptoms in oropharyngeal cancer patients receiving radiotherapy: development of a pre-treatment decision support tool. *Front. in onc.*, 13, 2023. 2

[48] K. Wongsuphasawat and D. Gotz. Outflow: Visualizing patient flow by symptoms and outcome. In *IEEE VisWeek Workshop Vis. Anal. in Health.* American Medical Informatics Association, 2011. 2

[49] V. Zebralla, J. Müller, T. Wald, Boehm, et al. Obtaining patient-reported outcomes electronically with "oncofunction" in head and neck cancer patients during aftercare. *Front. in Onco.*, 10, 2020. 2

[50] Z. Zhang, D. Gotz, and A. Perer. Iterative cohort analysis and exploration. *Info. Vis.*, 14(4), 2015. 1