

Projektleitung: Dr. med. Jan-Philipp Cieslik, MRes Cancer Informatics

Cancer Informatics: Studierende für die Klinik und Wissenschaft von Morgen ausbilden

Ausgangssituation

Bioinformatik findet zunehmend Einzug in unseren klinischen und wissenschaftlichen Alltag. Im Zeitalter von „Big Data“ und Hochdurchsatzverfahren wie Next Generation Sequencing (NGS) muss eine Vielzahl von Informationen zu einer klinischen Entscheidung zusammengefasst werden.

Zielgruppe

Unsere Lehrinhalte sind bewusst auf Studierende ohne (bio-)informatische Vorkenntnisse ausgelegt. Wir haben die Inhalte bereits erfolgreich mit Studierenden der Medizin sowie der Naturwissenschaften erprobt.

Auto Grader

Moderne E-Learning Plattformen (wie z.B. Ilias an der HHU) unterstützen Methoden zur Überprüfung von Wissen (z.B. Multiple-Choice-Fragen, Lückentexte, ...). In der Domäne der Computerwissenschaften werden seit längerer Zeit sogenannte „Auto Grader“ eingesetzt, welche Quellcode als Antwortmöglichkeit akzeptieren und diesen auf Korrektheit überprüfen. Auto Grader werden aktuell nicht von Ilias unterstützt, sodass wir auf etablierte Methoden wie gradeR zurückgreifen.

Open Access / Open Source

Alle erstellten Inhalte werden von uns quelloffen publiziert. Darunter fallen nicht nur Texte und „Code Alongs“, sondern auch Auto Grader Aufgaben sowie der Quellcode zum Simulieren der Datensätze (s. Abschnitt „Story Telling“). Dies erlaubt jedem, unseren Kurs nachzubauen und nach Belieben zu erweitern.

Story Telling: (Bio-)Informatik ist spannend

Unsere Zielgruppe sind u.a. alle Studierenden der Medizin. Wir möchten explizit vermeiden, nur eine kleine Gruppe von „Bioinformatikern“ anzusprechen. Um diesem Anspruch gerecht zu werden, müssen unsere Lehrinhalte klinisch relevant, verständlich und interessant sein.

Im Internet gibt es bereits viele verfügbare Übungsdatensätze auf Plattformen wie z. B. Kaggle, welche allerdings oft ohne größeren klinischen Kontext bearbeitet werden, sodass schnell Fragen wie „Warum sollte ich Methode X auf diesen Datensatz anwenden?“ entstehen.

Moderne, internationale Konsortien wie z.B. *The Cancer Genome Atlas* veröffentlichen zudem ebenfalls riesige Datensätze mit konkretem Bezug (zum Teil NGS-/Bildgebungs-/Überlebensdaten), welche allerdings ohne Anleitung überwältigend sein können. Zusätzlich sind große Teile der öffentlichen Daten nicht ohne weitere Aufarbeitung nutzbar. So muss z.B. eine (oftmals nicht triviale) Normalisierung der Daten erfolgen, um Schlussfolgerungen zu erlauben.

Einflussreiche klinische Studien zu neuen therapeutischen Ansätzen publizieren zudem selten Rohdaten, die eine bioinformatische Auswertung zulassen würden. Um Studierenden die dahinterliegenden Auswertungsprozesse zugänglich zu machen, simulieren wir dahinterliegende Datensätze und erklären daran praxisnah, anschaulich und zugänglich, wie die Verarbeitung der Daten abläuft und wie daraus der Plot entstehen kann, den sie in der Publikation finden.

Ein Großteil unseres Projektes bestand also darin, die Szenarien aus relevanten aktuellen Publikationen in unseren Datensätzen nachzustellen, z.B. durch Simulation von Überlebensdaten, um Szenarien aus bekannten wissenschaftlichen Artikeln zu imitieren. Dadurch können wir den Studierenden aktuelle Behandlungsleitlinien erläutern und gleichzeitig gemeinsam die Datengrundlage erarbeiten und Therapiekonzepte wissenschaftlich untermauern.

Formate der Veranstaltung

Zur Vermittlung von „Cancer Informatics Basiswissen“ haben wir in den vergangenen Semestern diverse Lehrformate erprobt: Eine 5-tägige Blockveranstaltung in Präsenz, semesterbegleitende virtuelle Seminare sowie hybride Modelle.

Zukünftig sollen unsere E-Learning Module auch unabhängig von unseren Lehrveranstaltungen nutzbar sein.

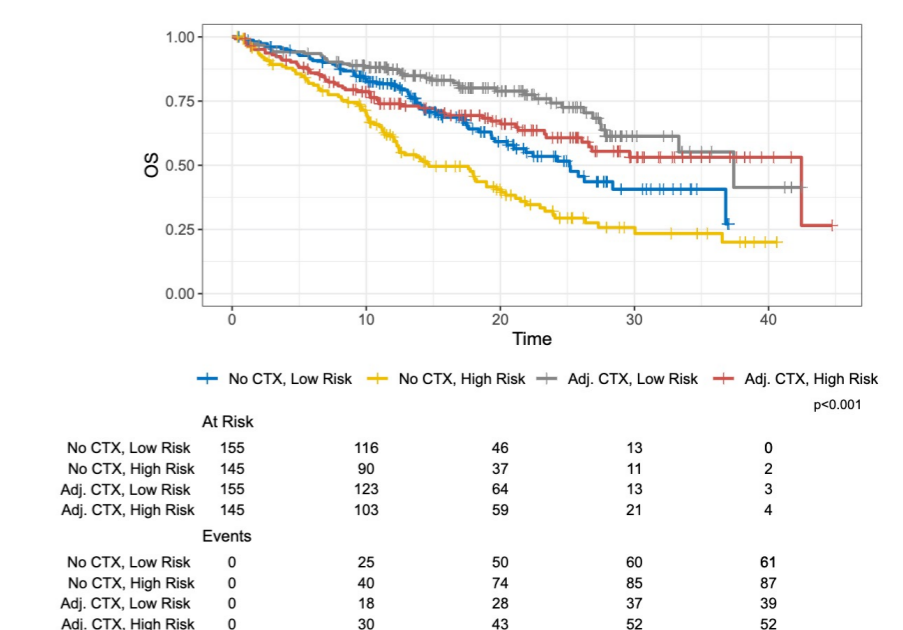
Evaluation

Unsere letzte „Cancer Informatics Summer School“ wurde von insgesamt 15 Studierenden besucht. Die Studierenden bewerteten das Wahlfach durchweg positiv (Gesamtbewertung in Schulnoten: 1,4). Aus Dozentensicht fällt vor allem eine große Diskrepanz in der digitalen Kompetenz der Studierenden auf: Einige Studierende waren unterfordert, während wenige andere grundlegende Schwierigkeiten bei der Bedienung der Computer hatten. Im Freitextfeld der Evaluation kam der Wunsch nach Lehrmaterialien im Vorhinein des Kurses auf; diesen werden wir perspektivisch mit einem zusätzlichen E-Learning Kurs beantworten.

Cancer Informatics

```
# Surv() creates a survival object with the time and status variables
# survfit2() requires a formula (dependent variable ~ independent variable(s))
# Here we specify the survival as the dependent variable and Treatment-Arm
# and RiskGroup as independent variables
surv_obj <- survfit2(Surv(Time, Status) ~ Treatment+RiskGroup, data = data_filtered)

# We utilize ggsurvfit() to generate a Kaplan-Meier plot
# the scale_color_jco() function is optional and provides a color palette from the ggsci package
ggsurvfit(surv_obj, linewidth = 1) + add_pvalue() + add_censor_mark() + labs(
  y = "OS",
  x = "Time"
) + ylim(0,1) + add_risktable() + scale_color_jco()
```



Beispiel: Simulierter Datensatz der ICON1 Studie

Ein Beispiel für einen simulierten Datensatz, welcher in unserem E-Learning Modul genutzt wird, ist die ICON1 Studie, welche sich mit der Frage nach adjuvanter Chemotherapie bei frühen Stadien des Ovarkarzinoms auseinandersetzt. Studierende können mit unserem Datensatz relevante statistischen Tests durchführen und Diagramme erstellen.

Cancer Informatics: Klinik trifft Bioinformatik

Kommende Termine für unsere Summer/Winter School:
18.03. - 22.03.2024
23.09. - 27.09.2024

Weitere spannende Wahlfachangebote und mehr Informationen zu unserem Projekt finden Sie auf unserer Homepage

Cancer-Informatics.org
oder über den QR-Code

