

# 医療データ科学 I

## Medical Data Science I

「医療データ科学」は、2022年度以降入学者のカリキュラムから卒業要件1単位として開講される必修科目であり、3年次「医療データ科学I」と4年次「医療データ科学II」に分けて開講される。各50点満点で評価され、最終的にI・IIの評価点を総合して「医療データ科学」として4年次末に単位認定される。なお、この授業は「講義及び実習」形式で行われ、出席の扱いは次のとおり。

- ・医療データ科学I（3年次12コマ：システム生物学分野担当）  
実習形式の授業で、最終試験は行わない。ただし、成績評価を受けるためには、全授業時間の半分以上の出席が必要となる。
- ・医療データ科学II（4年次9コマ：生物統計学分野担当 \*予定）  
講義（随時実習あり）形式の授業で、最後に試験を実施予定。試験の受験にはIIの全授業時間の半分以上の出席が必要となる。

なお、4年次末に単位を修得することから、4年次から5年次への進級条件には該当するが、3年次から4年次への進級条件には該当しない科目である。

### 1 授業の概要、目的

本授業は、医学・医療分野で急速に蓄積されつつある多種多様なデータを適切に分析し、その結果を正しく解釈することで、新たな解決策の提案や意思決定を行う能力の育成を目的とする。医療データ科学の基礎知識とデータ解析のためのプログラミングスキルを習得することにより、データ駆動型の医学研究を自立的に遂行するための思考力と実践力を身につける。授業では、統計学、機械学習、深層学習などの知識と適用能力を深めるとともに、データの特性や前処理の方法、各種解析ツールの使用法、結果の解釈や評価方法などについて、具体的な事例を用いてオンザジョブ形式で学ぶ。これにより、理論と実践を効果的に結びつけ、実際の研究や業務で活用できる能力を養成する。最終的には、受講者が自身の研究テーマや関心領域に関連する医学・医療データを適切に選択・活用し、新たな知見や洞察を導き出せるようになることを目指す。本授業で習得した知識とスキルは、医療の質の向上、医療現場での科学的根拠に基づく意思決定の促進、医学データからの新たな仮説の抽出や革新的な治療法の開発など、様々な側面で医学・医療の発展に寄与することが期待される。

This course aims to develop the ability to appropriately analyze various types of rapidly accumulating data in the medical and healthcare fields, correctly interpret the results, and propose new solutions and make decisions based on the findings. By acquiring basic knowledge of medical data science and programming skills for data analysis, students will develop the thinking and practical skills necessary to autonomously conduct data-driven medical research. The course will deepen students' knowledge and application skills in statistics, machine learning, deep learning, and other related areas. Additionally, students will learn about data characteristics, preprocessing methods, the use of various analytical tools, and the interpretation and evaluation of results through hands-on, on-the-job training using concrete examples. This approach effectively combines theory and practice, fostering the ability to utilize the acquired knowledge and skills in actual research and professional work. Ultimately, the course aims to enable students to appropriately select and utilize medical and healthcare data related to their research themes or areas of interest, leading to the discovery of new insights and findings. The knowledge and skills acquired in this course are expected to contribute to the advancement of medicine and healthcare in various aspects, such as improving the quality of medical care, promoting evidence-based decision-making in clinical settings, extracting new hypotheses from medical data, and developing innovative treatment methods.

## 2 到達目標

- ・データ駆動型医学研究の全体像と流れについて理解し、必要な基礎知識と IT スキルを身につける。
- ・医療データ解析に用いられる主要な手法の原理と特徴を理解し、目的に応じて適切な手法を選択・活用できる。
- ・Python/R を用いて実際のデータ解析を行い、結果を適切に解釈・考察することができる。
- ・データ解析による新たな知見を創出し、医療の発展に寄与するための基盤を身につける。

## 3 成績評価の方法と基準

本科目のシラバスの冒頭に示したとおり、医療データ科学 I・II それぞれ 50 点満点で評価され、最終的に I・II の評価点を総合して「医療データ科学」として 4 年次末に単位認定される。成績評価は A+, A, B, C, C-, F で行う。

医療データ科学 I は実習形式の授業で、最終試験は行わない。ただし、成績評価を受けるためには、全授業時間の半分以上の出席が必要となる。

医療データ科学 I の成績評価は、各回の授業において課される課題レポートの提出内容により評価する。なお、レポートは以下の観点から評価を行う。欠席・遅刻等がある場合は、その理由を踏まえて代替措置を講じるなどにより適切に評価する。

- ✓ 授業内容の理解度：授業で扱った医療データ科学の概念、理論、手法等に対する理解の深さを評価する。
- ✓ データ解析力：実際の医療データを用いて適切な解析を行い、結果を正確に解釈できているかを評価する。
- ✓ 考察力：データ解析から得られた知見を元に、医学・医療分野における課題解決や意思決定に向けた考察ができているかを評価する。

## 4 教科書

特に指定しない。授業資料はスライドまたは Jupyter Notebook の形式で配布する。

## 5 参考書

各回の担当教員が必要に応じて参考書を紹介する。

## 6 総括責任者

システム生物学分野 特任教授 島村 徹平 SHIMAMURA Teppei

## 7 授業日程

- ・2024年6月19日（水）
- ・2024年6月26日（水）
- ・2024年7月3日（水）
- ・2024年7月10日（水）
- ・2024年7月17日（水）

第1時限 08:40～09:40    第2時限 09:50～10:50  
第3時限 11:00～12:00

授業は第2講義室またはZoomを利用して行う。  
受講者は各自のノートPC（タブレット不可）を必ず持参すること。

月	日	時限	講座等名	担当教員名	職名	講義概要
6	19	1	システム生物学分野	島村 徹平	特任教授	1 データ解析環境
		2	東京医科歯科大学	阿部 興	特任講師	2 データ可視化
		3	東京医科歯科大学	阿部 興	特任講師	3 統計モデルと仮説検定
	26	1	東京医科歯科大学	阿部 興	特任講師	4 次元削減、クラスタリング
		2	システム生物学分野	島村 徹平	特任教授	5 回帰、教師あり学習
		3	東京医科歯科大学	林 周斗	准教授	6 深層学習
7	3	1	東京医科歯科大学	廣瀬 遥香	助教	7 オミクス解析①
		2	東京医科歯科大学	廣瀬 遥香	助教	8 オミクス解析②
	10	1	広島大学	本田 直樹	教授	9 数理モデル
		2	広島大学	本田 直樹	教授	10 画像解析
	17	1	産業技術総合研究所	小関 準	主任研究員	11 MDシミュレーション
		2	東京医科歯科大学	林 周斗	准教授	12 分子設計

## 8 授業内容

### 「データ解析環境」

内容：医療データ解析を行うためのプログラミング環境の構築と基本的な操作方法について解説する。

キーワード: Google Colaboratory、R、Python、Jupyter Notebook、パッケージ管理、バージョン管理

### 「データ可視化」

内容：医療データの可視化手法とその解釈について解説する。データの種類に応じた適切な可視化手法の選択を学ぶ。

キーワード：散布図、ヒストグラム、boxplot、バイオリンプロット、ヒートマップ、次元削減、インタラクティブ可視化

### 「統計モデルと仮説検定」

内容：医療データ解析における統計的推測の基礎となる統計モデルと仮説検定の考え方について解説する。

キーワード：確率分布、尤度、最尤推定、ベイズ推定、検定統計量、有意水準、p値、多重検定

## 「次元削減、クラスタリング」

内容：高次元医療データの低次元表現を得るための次元削減手法とデータの構造を探索するクラスタリング手法について解説する。

キーワード：主成分分析、多次元尺度構成法、t-SNE、UMAP、階層的クラスタリング、k-means 法、エルボー法、シルエット分析

## 「回帰モデル、教師あり学習」

内容：医療データを用いた予測モデルの構築に用いられる回帰分析と教師あり機械学習の手法について解説する。

キーワード：線形回帰、ロジスティック回帰、一般化線形モデル、決定木、ランダムフォレスト、勾配ブースティング、交差検証

## 「深層学習」

内容：医療画像や時系列データなど、複雑な構造を持つ医療データの解析に有用な深層学習の基礎と応用例について解説する。

キーワード：ニューラルネットワーク、畳み込みニューラルネットワーク、再帰型ニューラルネットワーク、深層生成モデル、転移学習、データ拡張

## 「オミクス解析①」

内容：シングルセル RNA-seq データを用いて細胞の多様性や動態を解析するための手法について解説する。

キーワード：発現量正規化、バッチ効果除去、細胞型アノテーション、クラスタリング、RNA Velocity、マーカー遺伝子同定

## 「オミクス解析②」

内容：空間的な位置情報を含むトランスクリプトームデータを解析するための手法について解説する。

キーワード：空間的発現パターンの解析、一細胞分解、細胞間コミュニケーション

## 「数理モデル」

内容：生命現象の数理モデリングや計算機シミュレーションの基礎について解説する。

キーワード：微分方程式、確率過程、ヌルクライン解析、線形安定性

## 「画像解析」

内容：生体イメージデータを解析するための画像処理の基礎と応用例について解析する。

キーワード：畳み込み、Savitzky-Golay フィルタ、フーリエ変換、ピーク検知

## 「MD シミュレーション」

内容：分子動力学シミュレーションの基礎と創薬への応用例について解説する。

キーワード：力場、運動方程式、thermostats、拘束条件、フリーエネルギー計算、分子ドッキング、アロステリック効果

## 「分子設計」

内容：深層学習を用いた分子設計の基礎と応用例について解説する。

キーワード：記述子、深層生成モデル、de novo 設計、ベイズ最適化

## 9 授業時間外学習の指示

以下のプログラミング言語 R、Python に関する動画や資料を参照し、基礎知識を各自習得しておくことを強く推奨する。

- 東京大学 Python プログラミング入門  
<https://utokyo-ipp.github.io/>  
PDF 版  
[https://utokyo-ipp.github.io/IPP\\_textbook.pdf](https://utokyo-ipp.github.io/IPP_textbook.pdf)  
Colab 版  
<https://colab.research.google.com/github/utokyo-ipp/utokyo-ipp.github.io/blob/master/colab/index.ipynb>
- 名古屋大学 CIBoG メディカルデータサイエンスのための R 入門  
20230703 CIBoG メディカルデータサイエンスのための R 入門 (1 日目午前) 動画  
<https://youtu.be/qCec-FRU1XQ>  
20230703 CIBoG メディカルデータサイエンスのための R 入門 (1 日目午後) 動画  
<https://youtu.be/dOIvGaerz0g>  
20230710 CIBoG メディカルデータサイエンスのための R 入門 (2 日目午前) 動画  
<https://youtu.be/-E0MHu34ILY>  
20230710 CIBoG メディカルデータサイエンスのための R 入門 (2 日目午後) 動画  
<https://youtu.be/biUaNIxggVg>

## 10 質問への対応方法

課題に関する質問は、授業用の Slack 上で随時受け付ける。

全体に共通するフィードバックは、授業内で適宜行う。