

数理・データサイエンス・AI教育の紹介

東京医科歯科大学における医療系データサイエンス教育の取組み

東京医科歯科大学統合教育機構
 (東京科学大学データサイエンス・AI全学教育機構 2024.10.1より)
 大学院医歯学総合研究科教育メディア開発学分野特任講師

須藤 毅顕



(左から 須藤、木下)

東京医科歯科大学統合教育機構
 (東京科学大学データサイエンス・AI全学教育機構 2024.10.1より)
 大学院医歯学総合研究科教育メディア開発学分野教授

木下 淳博

1. はじめに

2022年11月にOpenAIがChatGPTを発表して以降、AI技術に対する世界中の注目が一気に集まりました。この発表は、データサイエンスに対する関心を大きく引き上げ、ChatGPTの人気を爆発的に高めるとともに、データサイエンスの重要性を一層強調しました。アメリカのグラスゴー社が発表する職業ランキングでは、データサイエンティストが2020年度から2022年度にかけて上位にランクインし続けており、この現象はAIおよびデータサイエンスの需要が急増していることを示しています。技術の進化は加速し続ける見込みで、日本政府もデータサイエンスの人材育成に力を入れ

ています。

学士課程教育では、内閣府の「AI戦略2019」において、基礎に該当するリテラシーレベルを、2025年度までに大学・高専の全学生に相当する50万人の学生が履修するという目標が掲げられました(図1)。一方、2017年度に設立された「数理・AI・データサイエンス教育強化拠点コンソーシアム」では、文理を問わず全国すべての高等教育機関の学生が、数理・データサイエンス・AIを習得できるような教育体制の構築・普及を目指して活動を行っています。各大学の取組みを全国に普及・展開させるため、全国を6つのブロックに分けて、各ブロックに属する加盟校が取組みの情報をお互いに共有しています。

2024年5月時点ではさらに9ブロックに細分化し、加盟校は300校に及んでいます。

本学は、2020年度より、関東首都圏ブロックの特定分野協力校(医学・歯学)として認定を受け、医療系の学部を有する全国の大学にデータサイエンス・AI教育を普及・展開させることをミッションとして活動しています。本稿では2024年5月現在まで約3年間の経緯を振り返りながら、苦勞した点、工夫した点などを紹介したいと思います。

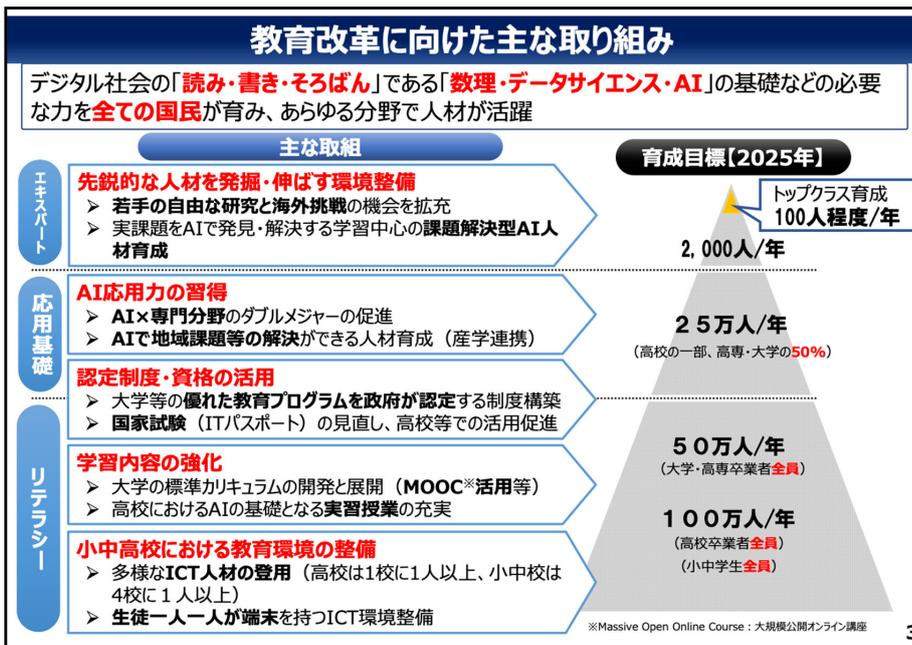


図1 教育改革に向けた主な取組み
AI戦略2019(内閣府)より

2. 2020年度

まず初年度は、学内の既存科目の把握と改変を行いました。コンソーシアムはリテラシーレベルと応用基礎レベルのモデルカリキュラムを公開しています。このモデルカリキュラムには、各レベルを構成する項目とその学修内容、ならびに授業に含めるべきキーワードとなるスキルセットなどが列挙されています。我々コンソーシアム加盟校の教員や、データサイエンス教育をこれから実装しようとしている大学の教員は、このモデルカリキュラムを既存科目と照らし合わせることで、自大学で何が不足しているのかを把握することができます。このモデルカリキュラムは、リテラシーレベル、応用基礎レベルとして、文系・理系問わず全大学・高専が参考とするようなモデルであるため、必ずしも自身の大学に全ての項目を含めて実装する必要はありません。例えば、医療系大学では、医療系の学生が身に付けるべきデータサイエンスを取捨選択して科目を構成していくことを想定しています(医学歯学薬学のモデルコアカリキュラム改訂を受けて、2024年1月には、コンソーシアムから「医歯薬系大学・学部における数理・データサイエンス・AI教育実施に向けた手引き」が公開され、その中でも触れられています)。いずれの医療系大学でも共通していることですが、本学でも、既に数学や統計学、疫学をはじめとするデータサイエンス関連科目が存在していました。そこで、既存の数学・統計学・情報学(バイオインフォマティクスを含む)関連科目の担当教員にヒアリングを行いました。実施している内容を確認し、事業内容およびコンソーシアムの概要を伝えた上で、科目間の重複を極力無くし、既存科目の枠組みの中で、少しでもモデルカリキュラムに沿った内容に変更できないかを検討しました。

例えば、これまでのデータサイエンスの科目では、それぞれの教員が各授業

必修科目と自由科目の2コース

【必修科目】

対象：1年生全学科

授業形式：同期型遠隔授業

履修者数：283人(令和4年度)

【自由科目】

対象：2-6年生(全学科)

授業形式：非同期型授業(LMS活用)

履修者数：本学19人+4 大学連合12人(令和4年度)

前半6コマ講義、後半9コマ演習(令和5年度は30コマに拡大)

最先端のデータサイエンスに触れ、自身のPCで実行することによる動機付け

内でプログラミングの実行環境のインストールから説明していました。変更後は1年前期の必修科目である情報処理/情報科学の授業で、Python導入の演習を全ての学生を対象に行って、自身のPCのOSに応じて、WinPython、またはAnacondaというソフトをインストールさせたため、その後はPythonによるプログラミング演習を学生自身のPCですぐに始められるようになりました。また、数学、統計学、情報学の基礎的な部分はある程度すでに網羅されていることがヒアリングから分かったため、機械学習、深層学習といったいわゆるAIに特化した授業内容とプログラミングの授業のみ追加すれば、モデルカリキュラムを網羅できると判断しました。

3. 2021年度

前年度の検討を受け、2021年度には「医療とAI・ビッグデータ入門」を新規科目として開講しました(図2)。本科目では、データサイエンスへの学生の動機付けを最大の目的とし、コンソーシアムのモデルカリキュラムでのリテラシーレベルに相当する内容で構成しました。開講当初は医学科、歯学科、検査学専攻1年生の必修科目でしたが、現在は全学科・専攻(医学科、歯学科、看護学専攻、口腔保健衛生学専攻、口腔工学専攻)1年生の必修科目としています。科目は前半が講義、後半がPythonによるプログラミング演習となるように構成しました。2020年4月にデータサイエンスの研究・教育に特化したセンターであるM&Dデータ科学センターが本学に設立されたため、前半の講義を同センターの講師陣に依頼して、

科目概要(令和4年度)

回/45分	内容	授業形式
1	大規模データ解析と人工知能によるがんの理解を目指して	講義
2	「未知」を発見する人工知能	講義
3	生物統計学概論 ～医療における意思決定のために～	講義
4	AI・データサイエンスを学んで次世代の医療を切り開こう	講義
5	データ科学アルゴリズム概論	講義
6	データサイエンス・AIの基礎とその活用	講義
7	python基本操作とデータ構造	講義・グループ演習
8	python基本操作とデータ構造	講義・グループ演習
9	①判別理論と実践	講義・グループ演習
10	①判別理論と実践	講義・グループ演習
11	②機械学習の理論と実践	講義・グループ演習
12	②機械学習の理論と実践	講義・グループ演習
13	③深層学習・AIの理論と実践	講義・グループ演習
14	③深層学習・AIの理論と実践	講義・グループ演習
15	③深層学習・AIの理論と実践	講義・グループ演習

図2 令和3年度：「医療とAI・ビッグデータ入門」の新設

最先端の研究事例やデータサイエンスの社会実装例を見せることで、医療分野での応用例や未来像を知ってもらえるようにしました。後半のプログラミング演習は、Pythonの基本、機械学習の基本、深層学習の基本の3部構成としました。初めてプログラミングをする学生が多いため、プログラミングのコードはこちらで用意して、プログラミングの内容を理解してもらいながら、自分のPCで機械学習が動くことや、データがどのように変化するかを体験してもらうことに焦点をあてました。基本的な内容にとどめて浅く広く知識を伝える授業とはせず、敢えて深層学習で肺のエックス線画像の分類まで実施させることを、演習の最終目標としました。プログラミングの基礎演習に焦点を当ててしまうとPythonの理解には繋がる内容となりますが、AIに興味を持ってくれるかどうかはわかりません。データサイエンスへの興味を引き出し、その魅力を伝えるためには、学生に「自身のPCで肺のエックス線画像を評価するAIを作ることができる」という驚きを与えるべきと考えた次第です。視覚的に分かりやすいエックス線画像を扱い、COVID-19の医療データを用いることで、医療系学生の興味を惹くとともに、自身が将来取り組む可能性のある課題を最終テーマに据え、そのために必要な最小限の知識を、具体的に分かりやすく伝えることに専念しました。

4. 2022年度

2022年度には、コンソーシアムでの応用基礎レベルを想定した「医療とAI・ビッグデータ応用」を新規科目として開講しました(図3)。リテラシーレベルである「医療とAI・ビッグデータ入

門」は動機付けが目的であったのに対し、この科目は理解をより深めて実践力を身につけることを目的としています。そのため、授業の9割はプログラミング演習としており、扱う題材も8割が深層学習(教師あり機械学習)、2割が自然科学系の基礎研究で頻繁に使用されている教師なし機械学習としました。また、応用基礎レベルはグループ演習の実施が必須とされていますので、本科目でも、グループワークとして深層学習のモデルを作成させ、各グループで作成した最高精度のモデルを提出させるという演習を実施しています。同科目では、急速に普及した生成AIへの対応を意識して、ChatGPTの扱い方から自然言語処理の演習までを授業に含めました。ChatGPTをはじめとする生成AIはあらゆる分野に少なからず影響を与えています。教育分野での影響も大きく、各大学でも生成AIへの対応を迫られていると思います。本学にはデータサイエンス・AI教育を全国の医療系大学・学部にも普及させるミッションもあることから、いち早く全学の学生・教職員に向けた「生成系AI利用ガイドライン」を作成、公表して生成AIの適切な利用を促しています。「ハルシネーション」(生成AIが偽りの情報をあたかも正しいかのようにテキスト生成する現象)など、生成AIの注意点を理解しないまま、生成AIを使いこなせずに、単にエンターテインメントの一環や、簡単なレポート作成ツールとして使用させるのではなく、生成AIの正しい利用方法やデータサイエンスの魅力を伝える方が、学生が得るものも大きいと期待しています。2023年度の演習では、ChatGPTの仕組みや簡単な使用方法を解説し、自然言語処理に関するプログラミング演習を行う上での、演習の題材となるテキスト

科目概要(令和4年度)

必修科目と自由科目の2コース

【必修科目】

対象：2年生(医学科、歯学科)

授業形式：同期型遠隔授業

【自由科目】

対象：2-6年生(全学科)

授業形式：非同期型授業(LMS活用)

履修者数：計171人(令和4年度)

9コマ演習、1コマ講義(令和4年度)

13コマ演習、2コマ講義(令和5年度)

深層学習をメインにデータサイエンス・AIの理解を深める

回/50分	授業題目	授業内容
1	MNISTの読み込み	データセットの読み込みと画像データの理解
2	MNISTの可視化と加工	画像データの可視化と学習の前処理
3	MLP(Multi layer perceptron)①	MLPとは 推論と学習の基本的概念
4	MLP②	MLPの推論と学習に関する基本演習
5	学習の仕組み	学習の仕組みの理解
6	CNN(Convolutional Neural Network)	CNNの基本的概念と基本演習
7	画像分類(グループ演習)	CIFAR10を用いたグループ演習
8	教師なし機械学習①	次元削減の概論と基本演習
9	教師なし機械学習②	クラスタリングの概論と基本演習
10	実際の医療AIの紹介	外部講師(越野沙織先生)による講義

MLP: Multi layer perceptron CNN: Convolutional neural network

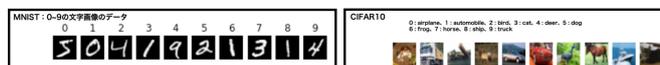


図3 令和4年度：「医療とAI・ビッグデータ応用」の新設

ド文書をChatGPTに作らせたたり、がんを疑う所見かどうかのテキスト分類を行うための疑似データをChatGPTに作らせたたりなど、授業コンテンツの作成にも役立ててChatGPTの活用法に工夫を凝らしています。

実行環境の変更



Zoom画面+実行画面で見やすさを重視

Google Colaboratoryの特徴

- Googleが提供するPythonの実行環境無料で利用可能、導入が容易、GPUも使用できる
WindowsとMacのOSの違いがない
- Googleのアカウントが必要
一度の使用時間(ランタイム)が連続最長12時間まで

図4 「医療とAI・ビッグデータ入門」の改変(令和5年度から)

5. 途中での改変や工夫した点

「医療とAI・ビッグデータ入門」や「医療とAI・ビッグデータ応用」の開講当初は、前述のようにWinPythonやAnacondaというソフトウェアをインストール後、Spyderという実行環境でプログラミング演習を実施していましたが、2023年度からは、全てのプログラミング実行環境をGoogle Colaboratoryへ移行しました。Google ColaboratoryはGoogleが提供するPythonの実行環境であり、Googleのアカウントさえ持っていればWebブラウザ上で誰でも無料で気軽にPythonのプログラミングを始めることができます。Spyderとの違いは画面の見やすさとGPUを使用できる点です。授業を開始して分かったことですが、学生は持ち運ぶことを想

複雑な深層学習を実施するため、高い計算処理能力が求められます。以前の演習では、長い処理時間で授業が止まるのを防ぐため、実行ボタンを押して機械学習が行われている間にプログラムの説明を行うなど、工夫を強いられていました。それでも、説明が終わって次のデモに進みたいがまだ機械学習の処理が終わっていないということもしばしばありました。Google Colaboratoryでは一定期間内の決められた時間の範囲内などという制約はあるものの、無料で計算処理能力の高いGPUを使用できるため、授業を円滑に実施する上でのメリットも大きいと感じています。

また、質問やオンデマンド受講学生への対応も工夫が必要でした。授業は基本的にZoomを活用した遠隔同期型で実施しましたが、両科目とも非同期のオンデマンドでも受講可能な科目として開設しています。Zoomの録画動画はその日のうちに編集し、本学のLMSであるWebClassを活用して、オンデマンドでの受講や復習のために公開しています。さらに授業実施日の翌週金曜日の17時から質問コーナーを設置し、遠隔同期型(Zoom)でも、遠隔非同期型(オンデマンド型)でも、極力公平に授業を受けられる工夫をしました(図5)。

データサイエンスの演習を実施す

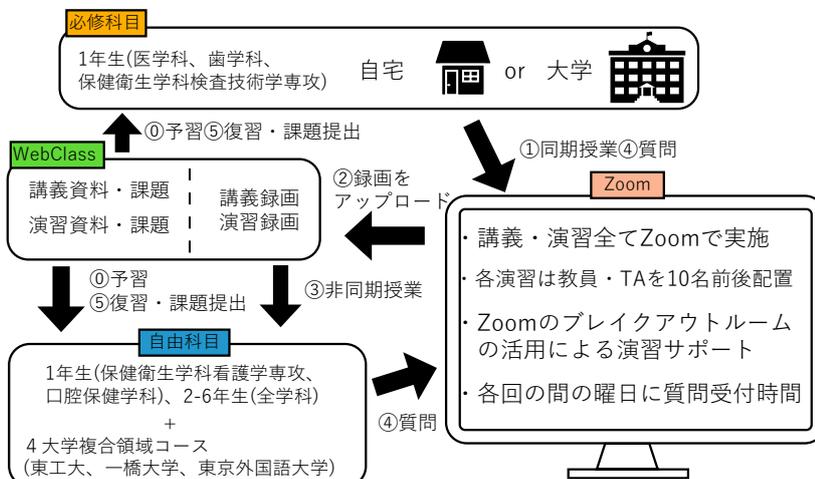


図5 ICT、LMSを活用した授業形態

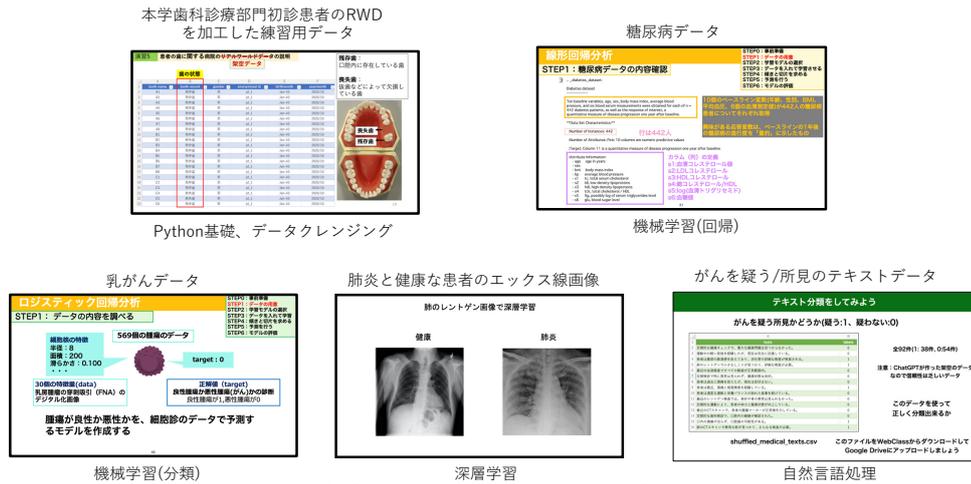


図6 現在演習で使用している医療系データセット

るためには、演習で扱うデータセットが必要不可欠です。特に、医療系学生向けのデータサイエンス演習では、医療系のデータセットが有用であると考えられます。そこで本学では可能な限り医療系のデータセットで演習を行えるように、メンバーが絶えずアンテナを張り、演習で使えるデータセットを探して少しずつ改変してきました。「医療とAI・ビッグデータ入門」では、初年度の深層学習における医療系データは肺のエックス線画像のみでしたが、2023年度はPythonのライブラリが提供している公共データである、糖尿病や乳がんのデータを用いた機械学習の演習を追加し、「医療とAI・ビッグデータ応用」での自然言語処理では、がんの所見のテキストデータをChatGPTに作らせて、擬似データとして演習に利用しました(図6)。データセットの充実とは他大学が実装する際にも役立つと思いますので、今後も拡充していく予定です。

6. 普及展開への取組み

こうした学内の取組みを、学外にも発信してきました。2020年度には数理・データサイエンス・AI教育強化拠点コンソーシアムとの共催でワークショップを、また2021年度には本学主催で医療系大学の教員向けワークショップを開催し、これまでの取組みの紹介とともに「医療とAI・ビッグデータ入門」の教材を配布しました。2022年度、2023年度は「医療系数理DSワークショップ」と題して、複数回に分けて「医療とAI・ビッグデータ入門」および「医療とAI・ビッグデータ応用」の授業のダイジェスト版を、他大学の医療系教員に体験してもらう取組みを実施しました。さらに日本医学教育学会大会や医療系eラーニング全国交流会などでもワークショップや発表を行い、積極的に本学の取組みの普及・展開を促進し

ています。また、ワークショップや授業で使用したスライドはGitHub上に公開しており、誰でも閲覧・使用できるようにしています。使用する際には、最初のスライドのQRコードから、授業対象者や対象人数等を回答していただく簡単なアンケートをお願いしています。医療系AI教育をすでに実施している大学にも、これから実施する大学にも参考にしていただくとともに、各大学からもフィードバックをいただいで今後の改善につなげてまいりたいと思います。

7. 最後に

医療系の学生に、どこまでデータサイエンスやAIを学ばせるかという問題は、ワーキンググループ内でも絶えず議論になっています。私見ではありますが、全ての医療系の学生に、一からAIを実装できるまでのデータサイエンティストレベルの知識や技術を求める必要はないと思っています。重要なことは、今後さらに加速するであろうAIの進化に対して、柔軟に対応できる基礎的な知識と姿勢を身につけることと、AIがどのようなことをしているのかという概念をしっかりと理解してもらうことだと思っています。そして何よりも重要なことは、医療の専門家として、AIについても興味を持ち続けてもらうことだと考えています。医療系の学生がAIを学ぶ際に、「AIが本当に医療に役立つのか」という疑問を抱くこともあるようですが、現在でも既に画像処理や画像診断などの領域で、医療系アプリケーションソフトにAIが組み込まれています。また、医療などの専門分野に限らず、日常生活のあらゆる場面でAIが活用される時代はすでに到来しています。今後、我々の医療従事者の生産性を上げるためには、AIは必須のものとなるでしょう。頭が柔軟な学生のうちからAIを積極的に活用する意識を持つことで、どの場面でAIが活用できるのか、またAIと既存の何を組み合わせれば良いのかについて、自由な発想が生まれることを期待しています。本取組みが少しでも多くの医療系教育関係者に認知され、医療系データサイエンス教育の普及に役立つことを願っています。