DL4HC: Deep Learning for Healthcare

Raghavendra Bhat
Intel India Bangalore Karnataka India
Raghavendra.bhat@intel.com

N C Shreyas
Intel India Bangalore Karnataka India
shreyas.nc@intel.com

Sandya Mannarswamy
Intel India Bangalore Karnataka India
sandyasm@gmail.com

ABSTRACT

In Over the last few years, Machine Learning (ML), and particularly Deep Learning (DL), has made great strides and has been successfully deployed in many real-world applications such as healthcare, customer care, finance, autonomous driving etc. One of the recent applications of deep learning has been in healthcare. ML/DL techniques have been applied to healthcare in medical imaging, clinical decision making, electronic healthcare records processing etc. In particular, Federated Learning (FL), an important sub-area of DL has been considered as a critical part of advancing ML driven healthcare systems in a data siloed healthcare ecosystem. FL addresses the critical need to train ML/DL algorithms without needing to move data to a central location. This tutorial is intended to provide a high-level overview of DL4HC at a major data science/ML research venue such as CODS-COMAD to focus research attention in this emerging area. This tutorial provides a detailed overview of DL/FL techniques in healthcare. While there has been considerable progress in this area, several challenges remain, especially in applying probabilistic techniques to healthcare which often requires deterministic correctness, interpretability, and verifiability. We discuss the various challenges involved in applying DL/FL techniques to healthcare and outline some of the future research directions. We believe that having this tutorial at CODS-COMAD would be an important step in facilitating increased research focus from data science community and greater collaboration with mainstream healthcare community on this topic.

CCS CONCEPTS

• **Computing methodologies** → Artificial intelligence; Distributed artificial intelligence.

KEYWORDS

Deep Learning, Federated Learning, Healthcare

ACM Reference Format:
Raghavendra Bhat, N C Shreyas, and Sandya Mannarswamy. 2023. DL4HC: Deep Learning for Healthcare. In *6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD) (CODS-COMAD 2023), January 04–07, 2023, Mumbai, India.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3570991.3571056

CODS-COMAD 2023, January 04–07, 2023, Mumbai, India

ACM ISBN 978-1-4503-9797-1/23/01…$15.00
https://doi.org/10.1145/3570991.3571056

1 GOALS & OBJECTIVES

Rapid progress in Deep Learning (DL) has seen a swift translation to real world commercial deployment in various fields. DL techniques are getting applied to finance, automotive, retail and healthcare. Applications of AI/DL/FL techniques to healthcare is critical as the world battles with pandemic, other population scale challenges, and climate changes [1–3]. While the area of applying DL to healthcare has made considerable progress over the past few years, there are still considerable challenges in taking research from "Bench2Bedside" as real-world deployment of DL/ remains slow. Availability of quality Data is a key barrier towards wider adoption of DL. Data silos, primarily due to regulatory and logistical barriers, exist in Healthcare limiting access to such quality data. Techniques like FL promises to break such Data Silos. It is critical for developing economies like India to have our DL researchers focus on this critical space to enable 'Affordable Healthcare for All'.

FL promises to help overcome the Data challenges thereby unlocking the power of DL/ML. DL/ML algorithms can be developed on a variety of Healthcare data across Image, text, columnar etc., thereby help create novel solutions. NLP techniques, specifically, promise to open new avenues in new emerging use cases like Drug development [4], Digital Twin etc. Given such huge potential for application of DL/ML techniques on Image and text to healthcare, the goal of this tutorial is to focus the attention of the CODS-COMAD community on this emerging topic and foster collaboration with medical research community. As DL techniques mature and newer DL applications in healthcare get deployed in real world, it is essential to develop models for healthcare which uniquely address the challenges of generalization, explainability and verifiability [5, 6]. It is also important to consider fairness and trust as researchers build DL models for healthcare. Our objective is to provide a practical overview of DL4HC, focusing on these critical aspects thereby facilitating a large number of CODS-COMAD attendees to start exploring this topic on their own.

2 TUTORIAL DESCRIPTION

This will be a 3-hour tutorial.

- Part I – DL techniques in Healthcare
- Part II – Federated Learning for Healthcare
- Part III - Datasets, Challenges & Research Directions
- Part IV – Federated Learning Hands on Tutorial

2.1 Part I: DL4HC: History, Applications & Techniques

We start with a brief historical overview of applying ML/DL techniques in healthcare. We then discuss the DL4HC under the categories

- DL for Medical Imaging
- DL/NLP for Clinical Decision Making & EHR Processing [8]

Segment	Data Type	Key Use cases	NLP Task
Payer	Insurance Claim Data	Claim processing Fraudulent claim detection Prior Authorization	Information Extraction, Text Classification
Provider	EHR clinical notes, Discharge summaries, Lab Reports, Online Forum Posts	Clinical Decision Support Disease Prediction Disease Management	Language Modelling, Information Extraction, Question Answering, Topic Modelling,
Pharma & Life Sciences	Adverse event reports, Bio medical literature, Clinical trial documents, Molecules/Protein Sequence data	Adverse Event Detection, Drug Discovery, Clinical Trial Optimization, Protein sequence modelling/structure prediction	Text Classification, Language Models, Information Extraction

DL for medical imaging needs to handle multiple imaging modalities such as Computed Tomography (CT), ultrasound, X-ray radiography, MR Imaging (MRI), and pathology being commonly used for medical imaging diagnostics. While CNN based models earlier dominated the medical imaging analysis, of late, transformer-based approaches [12, 14] have gained predominance in medical imaging. We cover the applications of vision transformer techniques to medical imaging [10, 11].

We also discuss how DL models need to handle medical imaging challenges such as expensive data acquisition, lack of standard image acquisition techniques in terms of tool and scanning settings, modality-specific artefacts, highly imbalanced data, and noisy annotated datasets. We also point out the generalization and transferability issues of DL models in medical imaging. In addition to DL successes in medical imaging, we also cover how DL models have failed in real world applications, the pitfalls involved and how they can be overcome [7]. Given the vast amounts of text data associated with healthcare industry, automated NLP driven approaches [13, 15] are being increasingly used to analyze healthcare text data and extract insights from it for driving better outcomes at lower costs. NLP applications in health care may be grouped into three broad target segments namely Payer (Health Insurance sector), Provider (hospitals/healthcare delivery) and Pharmaceuticals and Life Sciences (PLS). We will cover some of the key NLP use cases for each of these segments as seen in Table below.

2.2 Part II: Federated Learning for Healthcare

In this part, we provide a brief overview of federated learning [16] and then discuss how FL [9] is applied in healthcare. We will cover the following topics:

- Discuss the Data Silo challenges and key contributors to the Data Silo
- Highlight why Federated Learning is critical to adoption of AI in Healthcare in the context of Data Silo challenges [17]
- What it takes to deploy Federated Learning in a Healthcare setup. [18]
- Addressing Security and Privacy requirements. This is especially important considering the regulatory mandates to secure privacy of patients. We will cover how FL and, especially, Secure Federated Learning is a key enabler in addressing such requirements.
- Discuss some of the well-known and used Federated Learning solutions as deployed in healthcare use cases
- Introduce other challenges in deploying FL in healthcare which in turn are an input to the research community.

We first discuss the Data silo challenges in healthcare. We will touch upon the contributors to the Data silo challenge and how they affect advancing DL techniques in Healthcare. We will then focus on how FL promises to address these challenges in healthcare. Next, we will walk through the deployment requirements for FL in healthcare. The section will cover the Data ETL interfacing requirements across the diversity of data that is available, highlight the need for standards to help ease the deployment of standards, talk about the hybrid cloud deployment landscape stretching across on-prem to cloud and across multiple cloud solution providers, before closing with specific insights on FL deployment in a multi-organization and multi-geo scenario. Then we cover the Security and Privacy challenges in FL deployments in healthcare. We will touch upon differential privacy and secure execution as techniques to address the Security and Privacy concerns. We then provide a view of the various FL solutions with a comparative analysis of each of those solutions and finish with a picture of broad research problem statements calling for further attention.

2.3 Part III: Datasets, Challenges, Research Directions

In this part, we will highlight,

- Available datasets for DL4HC
- Use of synthetic data for DL4HC
- Challenges & Research Directions in DL4HC

The tutorial will highlight the key Dataset challenges that limit the availability and deployment of DL solutions in healthcare. We will provide a high-level overview of contributors to these challenges and how they limit research in healthcare. The talk will then focus on inorganic techniques like synthetic data creation that promises to advance the research field. However, such techniques also come with limitation on fairness, bias etc., which we will briefly focus on. We will close this section by listing down the challenges and research directions in analytics data management in healthcare.

2.4 Part IV: – Federated Learning Hands on Tutorial

We cover a brief hands-on example of using federated learning for a healthcare application namely medical imaging classification. The tutorial will highlight FL usage using Intel OpenFL solution and a public dataset chosen from among multiple examples available. The tutorial will mainly focus on the salient aspects involved in training medical image classification algorithms in a Federated approach with OpenFL more a vehicle to illustrate these aspects. This will be an interactive session using Jupyter notebook that will help the participants better appreciate the capabilities of FL.

3 ORGANIZERS

- Raghavendra Bhat (Raghavendra.bhat@intel.com), Principal Engineer, Intel.
- Sandya Mannarswamy (sandyasm@gmail.com), Principal Engineer, Intel.
- Shreyas Nc (Shreyas.nc@intel.com), AI Engineer, Intel

Raghavendra Bhat is a Principal Engineer at Intel Technology India Pvt. Ltd. He has a BTech in Computer Science from NIT Warangal and PGDBA (Finance) from Symbiosis Pune. He has over 23 years of industry experience spread across domains like Network Management, Embedded platform development for Mobile, IoT and Biometrics solutions. In his current role at Intel, he leads exploration in healthcare space as part of Vertical Solutions and Services Group. In his role, the primary focus is on design, development and ecosystem adoption of optimized AI solutions on Intel AI portfolio with strong Speech, Language, Image and Video analytics requirements. Currently, a significant focus is on AI Training at the Edge where training methodologies like Federated Learning, Incremental Learning etc are used. He is on the ISO and BIS standardization panel of AI (LITD30). He has several patents and publications to his credit.

Sandya Mannarswamy is currently Principal Engineer at Intel AI compilers group. Her Google Scholar profile is at https://scholar.google.com/citations?user=i27nd3oAAAAJ&hl=th. She was previously a senior research scientist at Xerox Research Centre India in the ML/NLP research group. She holds a Ph.D. in computer science from Indian Institute of Science, Bangalore. Her research interests span AI/machine learning and compilers. She was a senior compiler architect in Hewlett Packard Enterprise working on high level optimizations. Her research career spans over 20 years, at various R&D labs, including Hewlett Packard Ltd, IBM Research etc. She has co-organized several workshops including workshops at International Conference on Data Management, Machine Learning Debates workshop at ICML-2018, CIKM-2021 etc. She has presented tutorials at CODS-COMAD 2020, CODS-COMAD 2021, PAKDD 2021 conferences etc. She holds several publications and patents.

Shreyas is a Software Engineer at Vertical Solutions & Service group, Intel India. Shreyas is currently working on application of Federated learning in healthcare and contributing to Intel OpenFL stack. Prior to this Shreyas has worked on Intel OnBoard fleet Services., Intel NN accelerator and Chrome platforms as Systems engineer contributing Linux kernel and opensource user space libraries.

4 RELATED TUTORIALS

Given the considerable research interest in this area, there are a number of tutorials in applying ML for healthcare, including workshops/conferences such as ML4HC (https://www.mlforhc.org/.. However we are not aware of any recent workshops/tutorials in CODS-COMAD which cover this area, in particular focusing on deep learning and federated learning for healthcare with a hands-on tutorial. Given the emerging challenges for healthcare such as pandemics/climate calamities and in particular the challenges associated with affordable healthcare for all in Indian context, we believe that increasing attention of DL/FL researchers on applying their techniques to real world healthcare problems and that CODS-COMAD is an ideal venue to host the tutorial on this emerging area.

REFERENCES

[1] Recent advances and clinical applications of deep learning in medical image analysis. Xuxin Chen, Ximin Wang, Ke Zhang, Kar-Ming Fung, Theresa C. Thai, Kathleen Moore, Robert S. Mannel, Hong Liu, Bin Zheng, Yuchen Qiu

[2] A Unified Review of Deep Learning for Automated Medical Coding. Shaoxiong Ji, Wei Sun, Hang Dong, Honghan Wu, Pekka Marttinen

[3] Explainable Deep Learning in Healthcare: A Methodological Survey from an Attribution View. Di Jin , Elena Sergeeva , Wei-Hung Weng, Geeticka Chauhan , and Peter Szolovits

[4] Machine Learning and Deep Learning Methods for Building Intelligent Systems in Medicine and Drug Discovery: A Comprehensive Survey. G Jignesh Chowdary, Suganya G, Premalatha M, Asnath Victy Phamila Y, Karunamurthy K

[5] A Review of Challenges and Opportunities in Machine Learning for Health. Marzyeh Ghassemi, PhD1 , Tristan Naumann, PhD2 , Peter Schulam, PhD3 , Andrew L. Beam, PhD4 , Irene Y. Chen, SM5 , Rajesh Ranganath, PhD.

[6] Multimodal Machine Learning in Precision Health. Adrienne Kline, Hanyin Wang, Yikuan Li, Saya Dennis, Meghan Hutch, Zhenxing Xu, Fei Wang, Feixiong Cheng, Yuan Luo

[7] How I failed machine learning in medical imaging – shortcomings and recommendations Gael Varoquaux , Veronika Cheplygina

[8] Scalable and accurate deep learning for electronic health records. Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Peter J. Liu, Xiaobing Liu, Mimi Sun, Patrik Sundberg, Hector Yee

[9] Federated Learning for Smart Healthcare: A Survey. Dinh C. Nguyen, Quoc-Viet Pham, Pubudu N. Pathirana, Ming Ding, Aruna Seneviratne, Zihuai Lin, Octavia A. Dobre, Won-Joo Hwang

[10] Transformers in Medical Imaging: A Survey. Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu

[11] Vision Transformers in Medical Computer Vision – A Contemplative Retrospection. Arshi Parvaiz, Muhammad Anwaar Khalid, Rukhsana Zafar, Huma Ameer, Muhammad Ali, Muhammad Moazam Fraz.

[12] Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Richard J. Chen, Ming Y. Lu, Drew F.K. Williamson, Mane Williams. Bumjin Joo and Faisal Mahmood. Cell. Aug 08, 2022. Volume 40Issue 8p787-894

[13] How to Leverage Multimodal EHR Data for Better Medical Predictions? Yang, Bo; Wu, Lijun. EMNLP 2021.

[14] https://deepai.org/publication/a-multimodal-transformer-fusing-clinical-notes-with-structured-ehr-data-for-interpretable-in-hospital-mortality-prediction

[15] Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I. & Lungren, M. P. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. npj Digit. Med. 3, (2020).

[16] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics, pp. 1273–1282, PMLR, 2017*

[17] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Sci Rep., vol. 10, no. 1, p. 12598, 2020*

[18] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multiinstitutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," *Brainlesion,* vol. 11383, pp. 92–104, 2019.