HEALTHCARE

Solution Brief

Revision 0.1

habana

I.

AI in Healthcare

According to Research and Markets [1], the Global AI in healthcare market size is expected to grow from $6.1 billion in 2021 and reach $64.11 billion by 2027. 76% of the AI use cases in healthcare are related to diagnosis and treatment systems followed by automated customer service agents and digital assistants [2].

As the world learned during the COVID-19 crisis, the fast pace of research in the medical field is utterly critical. AI helped the pharma companies develop a COVID-19 vaccine in record time [3-4]. Analyzing chest X-ray images was the fastest way to diagnose this disease in its early days. However, the number of patients was high, and the healthcare specialists were under a lot of pressure. Computer vision solutions to aid the specialists are extremely helpful in such circumstances.

On the other hand, natural language processing techniques have helped scientists rapidly keep track of new advances in covid treatment. The White House and a coalition of leading research groups have released the COVID-19 Open Research Dataset (CORD-19) of over 1,000,000 scholarly articles about COVID-19 [5]. This freely available dataset is provided to the global research community to apply recent advances in natural language processing to generate new insights and treatments against COVID-19. A number of NLP models have been used to extract insights from this open dataset on infection and mortality rates in different demographics, symptoms of the disease, identifying suitable drugs for repurposing and interactions with other diseases [5].

Habana offers hardware and software solutions to increase the pace of research providing such solutions by shortening the time of AI experiment cycles. We are also proud that our solutions reduce the R&D costs letting the researchers explore more within their budget.

II.

Applications of AI in Healthcare

Medical image analysis is one of the most popular applications of AI in healthcare. Of course, there is always a certain amount of human error involved in this analysis, which can be reduced with the help of AI. Deep learning has recently been used for diagnosis and screening use cases. This includes time-sensitive use cases like the analysis of brain MRI to detect strokes, chest X-rays to detect COVID, histopathology images to detect and monitor cancer tissues, or cardiac ultrasound screening. Thanks to the great performance of deep-learning algorithms in these use cases, several companies have been able to obtain US FDA and European CE approvals for medical imaging AI [6].

Given the vast amounts of text data associated with healthcare, automated NLP-driven approaches are being increasingly used to analyze healthcare text data and extract insights from it for driving better outcomes at lower costs. NLP applications in healthcare may be grouped into three broad target segments namely Payer (Health Insurance sector), Provider (hospitals/healthcare delivery), and Pharmaceuticals and Life Sciences (PLS). Some of the key NLP use cases for each of these segments are listed in the following table.

Applications of AI in Healthcare

CONTINUED

<table>
<tr><th></th><th>Segment</th><th>Data Type</th><th>Key Use cases</th><th>Task</th><th>Popular Model Architectures</th></tr>
<tr><td rowspan="3">NLP</td><td>Payer</td><td>Insurance Claim Data</td><td>Claim processing
Fraud detection
Prior Authorization</td><td>Information Extraction
Text Classification</td><td rowspan="3">BERT
DistilBERT
RoBERTa
AlBERT</td></tr>
<tr><td>Provider</td><td>EHR clinical notes
Discharge summaries
Lab Reports
Online Forum Posts</td><td>Clinical Decision Support
Disease Prediction
Disease Management</td><td>Language Modelling
Information Extraction
Question Answering
Topic Modelling</td></tr>
<tr><td>Pharma & Life Sciences</td><td>Adverse event reports,
Bio medical literature,
Clinical trial documents,
Molecules/Protein Sequence data</td><td>Adverse Event Detection
Drug Discovery
Clinical Trial Optimization
Protein sequence Modeling/structure prediction</td><td>Text Classification
Language Models
Information Extraction</td></tr>
</table>

<table>
<tr><th></th><th>Segment</th><th>Data Type</th><th>Key Use cases</th><th>Task</th><th>Popular Model Architectures</th></tr>
<tr><td rowspan="2">CV</td><td rowspan="2">Provider</td><td rowspan="2">Medical images:
X-rays, CT, PET, MRI, FMRI, etc.</td><td>Diagnosis
Screening
Prognostication of outcomes and treatment response
Pathology segmentation
Disease monitoring</td><td>Image Classification
Image Segmentation
Object Detection
Information Extraction</td><td>DenseNet
U-Net
ResNet
RetinaNet
Mask R-CNN</td></tr>
<tr><td>converting low-dosage, low-resolution CT images into high-resolution images</td><td>Image enhancement
Image reconstruction</td><td>U-Net</td></tr>
</table>

Challenges

Some medical imaging datasets consist of a huge amount of data. For example, the field of radiology has collected about 1 million annotated open-source chest x-ray images [7-9], the closest ImageNet equivalent to date in medical CV [10]. The histopathology datasets can also be large, because different chemical preparations will render different slides for the same piece of tissue, and different digitization devices or settings may produce different images for the same slide [10]. Radiology modalities such as CT and MRI render massive 3D images and larger models are required to extract the relevant information from them.

Healthcare text data is also huge. In 2015, HIMSS estimated that approximately 1.2 billion clinical documents were produced by the healthcare industry [11]. Since then, the amount of text data in healthcare has grown tremendously with the increased digitization in healthcare. In addition to the clinical text documents and Electronic Health Records (EHR), there is an ever-growing amount of text data from biomedical literature, health insurance documents, clinical trial data, online patient forums, social media posts, etc. These data sources have their own diverse vocabulary even within the same domain of healthcare. Hence fast and accurate processing of these unstructured text data using state-of-the-art transformer-based NLP models at an affordable cost to extract insights is critical.

Training large models on such large datasets takes a long time, and slow down the pace of research. Time is of great essence for most healthcare applications. If a loved one is suffering from a disease or the world is facing a pandemic, we want the solutions to be in doctors' hands as soon as possible. The faster the training cycles of AI models, the sooner we can achieve the desired results.

III.

Why is Gaudi a good fit for healthcare use cases?

Deep neural network-based medical image analysis or genomics includes a large amount of processing that can be parallelized and thus accelerated. Healthcare use cases benefit specifically from accelerators that can handle data parallelism when the training dataset is huge and model parallelism when the models are large.

The two primary considerations that come into play in employing AI processing—whether for computer vision or NLP applications-- are time-to-train models to the desired level of accuracy and cost-to-train. Habana's Gaudi Training Processors are expressly designed—in both hardware and software—to deliver high-efficiency cost- and time-to-train, making AI training more accessible to more organizations and for more applications. This helps to reduce development and validation costs and to enable rapid innovation and faster time to market.

Training with Gaudi clusters is available both in the cloud with AWS EC2 DL1 instances consisting of 8 Gaudis and on-premises with the Supermicro X12 Gaudi Training Server, also featuring 8 Gaudis.

The ideal equation for end users is to achieve desired AI price-performance, meaning that the cost and time to train each image or language sequence meets cost and time investment criteria. In other words, enabling more training at a low cost is the objective for data scientists and IT infrastructure management.

Why is Gaudi a good fit for heathcare use cases?

CONTINUED

First-generation Gaudi, in fact, has proven delivery of up to 40% better price-performance than with comparable GPU-based solutions—for both the AWS EC2 DL1 instance as well as for on-premise systems. And, there are customer cases that have proven even greater cost savings, which will be shared in the next section.

IIn addition, Gaudi2, which launched in May, offers substantial performance advances that enable significantly faster training of models, while preserving cost-efficiency. Gaudi2 systems will be available in 2H 2022 for on-premises implementation.

IV.

News and customer testimonials

leidos

"Given Leidos and its customers' need for quick, easy, and cost-effective training for deep learning models, we are excited to have begun this journey with Intel and AWS to use Amazon EC2 DL1 instances based on Habana Gaudi AI processors."

Chetan Paul
CTO Health and Human Services
Leidos

Leidos, built a rich portfolio of healthcare applications for various federal agencies. These include NLP applications for regulatory review, subject matter queries, fraud, waste and abuse detection, and medical literature search and correlation for various government health agencies. They successfully demonstrated 60% cost savings with using DL1 vs. GPU instances training deep learning solution to facilitate medical benefit application processing in Veterans Health agency on DL1 Habana Gaudi instances.

V.

References

[1] Global AI in Healthcare Market (2021-2027) by Sections, Diagnosis, End user and Geography. IGR Competitive Analysis, Impact of Covid-19, Ansoff Analysis, available at https://www.researchandmarkets.com/reports/5451294

[2] “IDC's Worldwide Artificial Intelligence Spending Guide Taxonomy”, IDC, Release V1, 2022

[3] “AI and the COVID-19 Vaccine: Moderna’s Dave Johnson”, available at https://sloanreview.mit.edu/audio/ai-and-the-covid-19-vaccine-modernas-dave-johnson/

[4] “How a Novel ‘Incubation Sandbox’ Helped Speed Up Data Analysis in Pfizer’s COVID-19 Vaccine Trial”, available at https://www.pfizer.com/news/articles/how_a_novel_incubation_sandbox_helped_speed_up_data_analysis_in_pfizer_s_covid_19_vaccine_trial

[5] Wang, Lucy Lu, et al., “CORD-19: The COVID-19 Open Research Dataset”, Proceedings of the 1st Workshop on NLP for COVID-19, Association for Computational Linguistics (ACL), 2020

[6] Benjamens, Stan, Pranavsingh Dhunnoo, and Bertalan Meskó. "The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database." *NPJ digital medicine* 3.1 (2020): 1-8.

[7] Johnson, Alistair EW, et al. "MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports." Scientific data 6.1 (2019): 1-8.

[8]. Irvin, Jeremy, et al. "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. No. 01. 2019.

[9]. Wang, Xiaosong, et al. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[10] Esteva, Andre, et al. "Deep learning-enabled medical computer vision." *NPJ digital medicine* 4.1 (2021): 1-9.

[11] “AI in Healthcare: How It’s Changing the Industry”, available at https://www.himss.org/resources/ai-healthcare-how-its-changing-industry