ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

# Multimodal Foundation Models for Precision Medicine

**Emma Richardson**
Asst Professor

## Abstract

Deep learning models pretrained on multimodal Big Data may be used to address many problems in Precision Medicine. Choi et al. (2022) provided background on the core model architectures and training paradigms; on the prevalent types of multimodal medical data; and on the preliminary works aimed at combining model frameworks and training paradigm across existing modalities. Multimodal Foundation Models are introduced on the basis of learned medical representation and multimodal representation transfer and design. The explored areas are centered on data ecosystems that can support the future development of Deep Learning within Precision Medicine using multimodal Big Data.

As subtitled, the section on Data Ecosystems and Governance focuses on data acquisition, curation and quality assurance, privacy, security and ethical considerations. Applications involve the integration of genomics and transcriptomics; the use of medical image and radiomic data; and the building–testing of benchmark datasets to alleviate data bias. The addressed methodological challenges discuss data bias, fairness and generalizability; interpretability and clinician trust; benchmarking protocols; reproducibility; and open science practices.

**Keywords :** Deep Learning; Foundation Models; Multi-task Learning; Genomics; Transcriptomics; Radiology; Clinical Medicine.

## 1. Introduction

Data-driven methods such as Deep Learning have led to significant advances across multiple scientific fields. While Deep Learning models trained on vision datasets such as ImageNet have been shown to achieve state-of-the-art performance in performing other visual tasks, the key factor behind these successes is their datasets, which enable the pre-training of foundation models that serve as a basis for a large family of downstream tasks. To date however, these data-centric techniques have only been applied to a small set of domains, and the scientific impact of such foundation models with their data-centric frameworks, spanning multimodal big data platforms across a diverse spectrum of modalities including image, text, audio, sensor signals, and spatial / temporal data, has yet to be fully explored. This is particularly true for the field of biomedical research and Precision Healthy Ageing where Deep Learning methods are rapidly gaining traction.

Precision Healthy Ageing refers to the understanding of ageing mechanism at multiple levels of biological organization—from genomes to populations—together with the data and tools required to use that understanding to improve and maintain health in older people. Multimodal data are fundamental to Precision Healthy Ageing, and Deep Learning also enables the application of Advanced Artificial Intelligence (AI) to multimodal data. Moreover, recent research has revealed that while individual data modalities yield impressive results, the integration of multiple modalities is critical for achieving superior accuracy and robustness in predictive tasks. Indeed, simply being able to correlate heterogeneous data is not enough in many cases—combining modalities in informative ways is crucial for improving prediction performance.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

**1.1. Background and Significance**

Deep learning (DL)—a field of machine learning (ML) based on artificial neural networks (NN)—has attracted both academic and industrial interests thanks to recent breakthroughs in high-dimensional intelligence-enabled applications, including self-driving cars, Go-playing AI, Large Language Models (LLM) such as ChatGPT, and commercial image and audio generators. Now a vibrant and large-scale area of research, DL has grown beyond its originally understood values and become the research hotspot and buzzword in almost all AI-related topics. Many neglected but important applications of Deep Learning, such as biology and medicine—areas that are often considered different from AI/ML, for example, using human-created features such as Histogram of Oriented Gradients (HOG) features in computer vision or hand-crafted mutations in genomics—are now given new lights due to the availability of large amounts of multimodal multimodal medical data. Further, these are actually bio genomics and transcriptomics image-based data, making them possible to learn features in a "data-driven" DL way.

However, the growing application fields of Deep Learning in medical problems suffer from several issues: Most studies are standalone works focusing on one single task problem under one single type of data, making it difficult to discover the real underlying biological mechanisms, and therefore understanding and clinically applying the models; existing DL models mainly focus on a single data type, while biological developments concentrate on the integrations of different data types such as genomics and transcriptomics imaging integration; many applications aim at data generation. Consequently, no true end–to–end multimodal Deep Learning systems or foundation models have been built in Precision Medicine thus far. To fill this gap in Precision Medicine, A complete analysis of how to integrate, mine, and use all multimodal data together as a Data Ecosystem required for Deep Learning are performed.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

**Fig 1: Foundation Deep Learning Models for Precision Medicine Using Multimodal Big Data**

**1.2. Research design**

The proposed research centers on three crucial elements: the current state and future design of deep learning models for precision medicine, multimodal medical data ecosystems, and applications spanning all clinical areas. Deep learning for biological data relies largely on domain-specific neural network architectures, data sources, and data types. Multimodal Foundation Models offer a unifying approach to their integration and generalization. Attention Models for multimodal Foundation Models generalize across data types to enhance transfer learning, and Data Responsibility provides a framework for ecosystem design and governance. Together, these aspects establish core Foundations for DW-Precision Medicine, which are utilized to accelerate and facilitate predictive tasks across all clinical areas.

Despite substantial advances in precision-medicine tasks—particularly in imaging—deep learning's predictions are often still unreliable and deployed mainly to support clinical decisions rather than replace them. Three solutions underlie the research's methodological impetus: the need for causal inference from observational data rather than randomized control trials; the introduction of ChatGPT and related Large Language Models, which utilize Deep Foundations of Knowledge; and a DELive Data Economy encompassing the creation, deployment, and governance of Data Ecosystems for Specific Applications. Open Foundation Models representing both vision-language understanding and Generation become increasingly popular and pave ways

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

for various research directions beyond standard supervised or partially supervised Precision Medicine tasks, including Visual Question Answering, Visual Dialog, Visual Captioning, and Visual Chatbot. The multilayer cross-testing approach enhances parameter efficiency and substitutes actual paired data requirements.

## 2. Foundations of Deep Learning for Precision Medicine

Deep learning relies on the extraction of hierarchical representations and has been successful across natural language processing, computer vision, and speech recognition. Although these domains vary significantly, the same pattern emerges: feature engineering becomes unnecessary when representation learning effectively extracts features from the data. Precision medicine centers on using individualized genetic profiles and multimodal data to enable personalized drug discovery and disease prevention. Data-rich clinical environments—specially in oncology—provide ample opportunity for deep-learning data representation, and increasing efforts are focusing on powerful deep-learning applications. Hence, these approaches can benefit from the rapidly growing set of foundation models in natural language processing, synthesis, and vision; yet their core concepts—data and model sharing—remain largely unexplored. These models go beyond closed systems that share a data filter or specific prediction task, instead integrating part of the model architecture that accepts training signals from any data source with partial labels.

Crucial to establishing deep-learning foundation models in precision medicine is the exploration of core architecture for multimodal data types—deep networks designed to extract valuable interpretable representations from pictures, videos, and sounds for any downstream task. Partitioning these approaches within precision medicine enhances understanding of their opportunities. Yet successfully adapting the core principles is nontrivial: the usefulness (or ability) of the architecture's filtering-box concept hinges on acquiring salient patterns from numerous AIP types, and the challenging data ecosystems involved must embrace the architectural and algorithmic flexibility essential for success. The result supports the development and collation of data ecosystems and computational tools tailored to the exploration of multimodal foundation models in precision medicine.

**Equation 1: Transformer self-attention (core of "attention models" mentioned)**

**Step-by-step derivation**

**Goal:** convert an input sequence into contextual representations by letting each token "attend" to others.

**1. Define input**

Let the input be a sequence of $L$tokens, embedded into vectors of size $d_{\text{model}}$:

$$X \in \mathbb{R}^{L \times d_{\text{model}}}$$

**2. Linear projections to queries, keys, values**

Choose a head dimension $d_k$and define:

$$Q = XW_Q, K = XW_K, V = XW_V$$

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

where

- $W_Q, W_K \in \mathbb{R}^{d_{\text{model}} \times d_k}$
- $W_V \in \mathbb{R}^{d_{\text{model}} \times d_v}$

So:

- $Q, K \in \mathbb{R}^{L \times d_k}$
- $V \in \mathbb{R}^{L \times d_v}$

**3. Similarity scores (dot products)**

For query position $i$and key position $j$, unnormalized compatibility is:

$$s_{ij} = q_i \cdot k_j$$

Matrix form:

$$S = QK^\top \in \mathbb{R}^{L \times L}$$

**4. Why "scaled" dot product?**

If $d_k$is large, dot products grow in magnitude → softmax saturates → gradients vanish.
Assuming elements are roughly mean 0, variance 1, then $\text{Var}(q_i \cdot k_j) \propto d_k$.
So we scale by $\sqrt{d_k}$:

$$\tilde{S} = \frac{QK^\top}{\sqrt{d_k}}$$

**5. Softmax to get attention weights**

Row-wise softmax makes each row sum to 1:

$$A_{ij} = \frac{\exp(\tilde{s}_{ij})}{\sum_{t=1}^{L} \exp(\tilde{s}_{it})}$$

So $A \in \mathbb{R}^{L \times L}$is a distribution over keys per query.

**6. Weighted sum of values**

Output sequence:

$$O = AV \in \mathbb{R}^{L \times d_v}$$

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

Each output token $o_i = \sum_{j=1}^{L} A_{ij}\, v_j$.

✅ **Final self-attention equation**

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V$$

**2.1. Core Architectures and Training Paradigms**

Deep learning enables the automatic extraction of hierarchical representations from large volumes of raw data with minimal manual feature engineering, and has achieved performance levels above human experts in a range of tasks across computer vision, natural language processing, speech recognition, and game playing. In a biomedical context, these core operations are suitable not only for image classification and segmentation tasks on pathology images and radiomics classification, but also for labor-intensive tasks involving time-series data (e.g. de-novo pathology report generation for histopathological images) and sequence prediction or generation tasks not yet tackled in healthy modelling studies. Unsupervised pre-training, supervised fine-tuning, and semi-supervised learning represent the training paradigms that have enabled the latest breakthroughs. Recent large language models are the most prominent example of self-supervised pre-training using vast amounts of unlabeled data scraped from the web, and present the next logical step for multimodal foundation models in medicine: multimodal multilingual representation learning on the tremendous heterogeneous patient records from sources such as MIMIC-III. The latest research has shown that the combination of visual-linguistic pre-training along with joint training of the two modalities results in superior representations for these and downstream tasks across domains, validating HuBERT as well on AudioSet classification.

**2.2. Multimodal Data Types in Precision Medicine**

A diverse range of data types are generated in the process of biomedical research and clinical diagnosis. Multiple scales of biological observations, spanning genes, proteins, and tissues, have been discovered. Datasets characterizing the distribution of diseases and patient prognosis based on textual records, medical imaging, biochemical test results, and clinical laboratory results are collected. Once processed and annotated, these datasets represent multimodal big data in the context of biomedical research. Deep learning has achieved evident success in the analysis of single-modal data. However, integrating complicated and multimodal big data effectively into a single model to improve the ensemble learning scheme remains challenging. Deep learning foundation models are powerful pre-trained models designed to enable transfer learning for a wide range of tasks across different modalities. Recent studies have proposed Foundation Models for precision medicine but have been limited to some specific modalities, including genomics, transcriptomics, medical images, and pathology. Generative Models, capable of understanding and creating multiple forms of data, represent the next generation of Foundation Models in artificial intelligence. To further enhance the Representation Learning, Self-supervised Multimodal Foundation Models enable the integration of representation information and visual knowledge from multimodal big data.

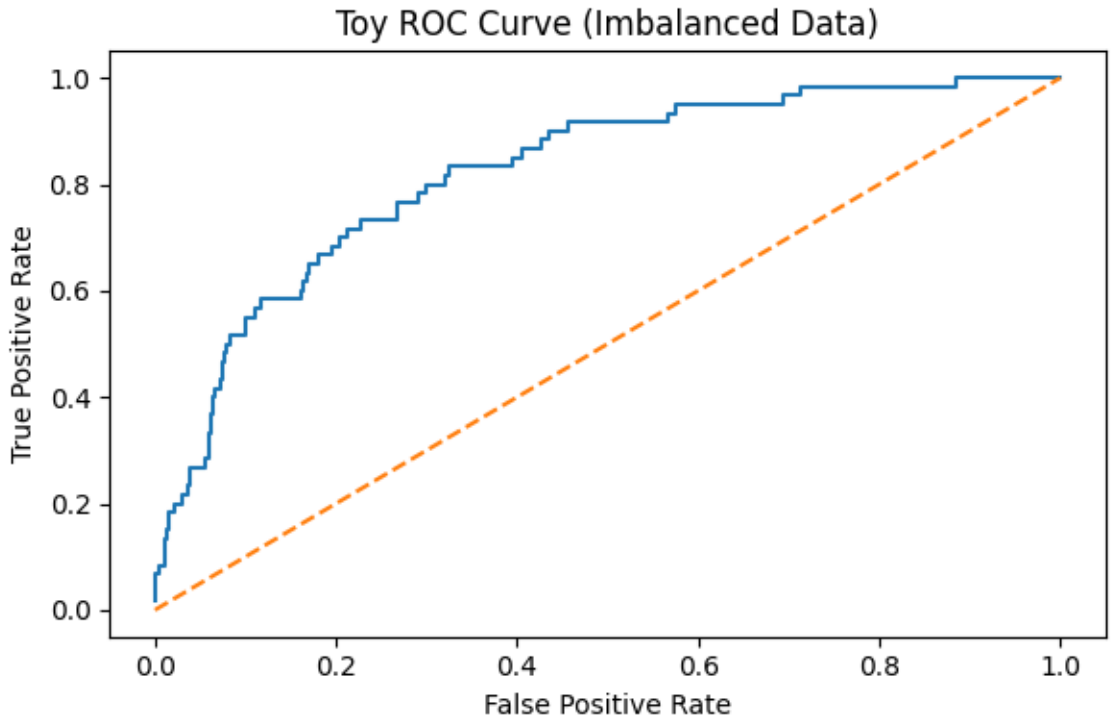
Deep-learning models for text data in precision medicine are well studied and routinely applied in clinical applications, but the expansion of these models for multiple modal data remains a trend in development. The use of multimodal multitask learning models trained with both closed and open data are routine and support supervised or partially supervised self-learning methods.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17




## 3. Multimodal Foundation Models

Foundation Deep Learning Models for Precision Medicine Using Multimodal Big Data: Multimodal Foundation Models. Multimodal formalism enables effective model integration across disparate data modalities, supporting comprehensive views of biomedical processes. Rather than fusing views of an underlying biological ground-truth, typically achieved on the task-specific level (e.g., joint image-genomic disease prediction), the foundation-model approach seeks to jointly harness data across data types into a single model capable of supporting any task requiring those data types. Contrastive foundation models integrate data sampled from untrained or task-free distributions, e.g., captions and images. Biomedical multimodal representation learning addresses the semantics of the data, but instead of modeling word-image relations, it models relations across more general data modalities (e.g., Text2Image) or forms like images and radiomics. In precision medicine, multimodal foundation learning is key for bridge-building between disparate data sources like imaging and genomics.

Foundation-model multiscale representation learning also applies to heterogeneous biomedical data sources, pairing biomedical text, electronic health records, pathology reports, and X-ray images. As different sets of these multimodal sources become available, each domain has attracted dedicated methods, reflecting a lack of suitable benchmark datasets for joint representation learning. The same applies to pre-trained representation learning for EHRs and other types of multimodal foundation models, where the focus has primarily been on audio-text or image-text relations, without accounting for domain-specific semantics.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

Foundation foundations support precision-medicine tasks at both the multimodal and representation-learning levels, serving as benchmarks for other DM methods.

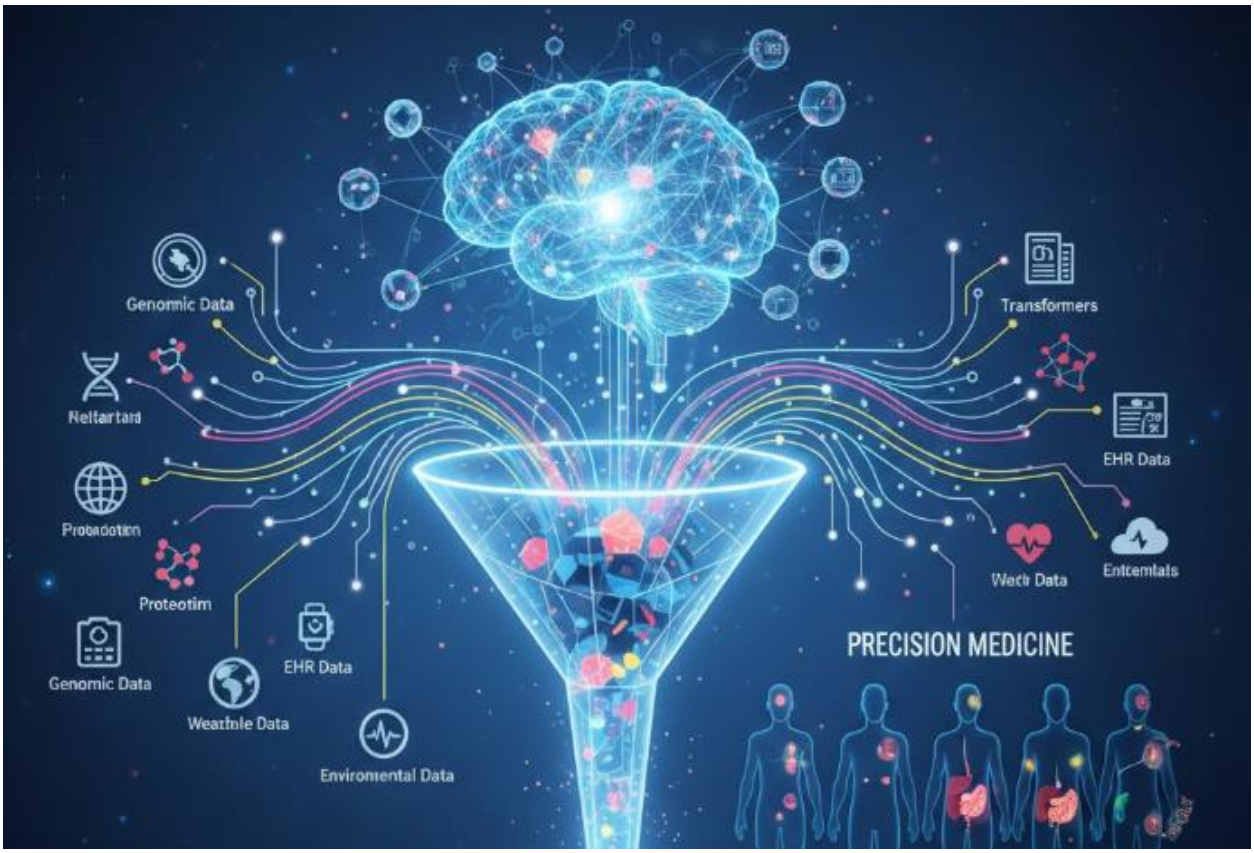



**Fig 2: Multimodal Foundation Models**

### 3.1. Model Integration Across Data Modalities

Heterogeneous models are capable of data-level integration, enabling unified model training that minimizes information loss when combining heterogeneous modalities. Advanced models, which comprehensively map and model data from diverse modalities using a joint multimodal foundation model layer, present an appealing approach for multimodal analysis. Such a model operates on the level of building multimodal representations, integrating learned multimodal representations into a single model after layerwise pretraining on each modality. These representations can then be jointly tuned or used to train multimode-specific specialized head networks.

Generalized language pre-training models represent a pioneering multimodal approach. In such a model, all modalities are converted to make textual stimulus or labels available for a language model. A unified vision-language model (UniVL),

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

employing an encoder-decoder architecture, exploits the correlation between vision and language. Based on an image-text pair dataset, it models the semantic relationship between images and their captions and is trained to generate empirical natural language descriptions of given images. A recent large-scale language-vision model integrates visual concepts into the foundation model of an image-text pair dataset to enhance the vision-language model's zero-shot capability across several vision tasks. The general idea of seeking a task-agnostic foundation model in vision-language applications is promising, as such a model can be adapted to multiple tasks irrespective of the difference between the input–output pairs on different tasks. Multimodal models that learn joint multimodal embeddings from text and images improve the accuracy of image classification and textual–olistic retrieval and captioning tasks.

**Equation 2: Multimodal embedding + fusion (what the paper calls "integration across modalities")**

**Step-by-step**

1. Encode each modality:

$$z_m = f_m(x_m) \in \mathbb{R}^d$$

2. Intermediate fusion (common in multimodal transformers):

- concatenate:

$$z = [z_1; z_2; \dots ; z_M] \in \mathbb{R}^{Md}$$

- then project:

$$h = \phi(Wz + b)$$

3. Task head:

$$\hat{y} = g(h)$$

**3.2. Representation Learning for Heterogeneous Biomedical Data**

Every single day, vast amounts of heterogeneous biomedical data are collected across the world, including genomics, transcriptomics, proteomics, metabolomics, radiomics, and electronic health records, among many others. Deep learning methods are significant in learning feature representation from the data, especially when the feature dimension is extremely large, i.e., images, documents, and genes, etc. Foundation modeling methods learn representations well for one modality and can be used to give reasonably good results when applied to other modalities. Hetero-Encoder is a family of representation learning methods that learn representations for heterogeneous biomedical data using the same input-output structure as BERT. Proxy tasks are designed for medical natural language processing; the proxy tasks used in the original Hetero-Encoder for medical imaging have been adapted to support multi-omics data.

The genomic and transcriptomic data representation learning methods explore applying Hetero-Encoder to learn representations for other modalities of high-dimensional biomedical data. The learned representations are used for primary downstream tasks,

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

such as classification and regression, and for secondary tasks such as few-shot learning and zero-shot learning. The learning strategies of these methods shared an encoder-decoder architecture, supporting transfer learning applications. Block-wise, cross-modality representation alignment maps the representation space of one modality image to the other modality in a block-wise manner in an unsupervised setting. Automating Data-Driven Molecular Imaging employs autoregressive modeling in the design of contrast-enhanced magnetic resonance imaging tasks for each MRI modality.

## 4. Data Ecosystems and Governance

Scientific discovery and technological innovation are increasingly fueled by big data from diverse data-rich domains. Likewise, neural networks and other data-hungry learning methods benefit from scaling up the quantity and quality of labeled training data. However, obtaining large, high-quality datasets remains one of the most challenging aspects in any machine learning project, particularly in the emerging area of precision medicine in which tremendous amounts of multimodal data (genomics, transcriptomics, radiomics, imaging and electronic health records) have been generated in the past decade but few sufficiently large, high-quality, publicly available datasets exist yet.

Creating such a multimodal big data ecosystem for precision medicine at scale requires large-scale data acquisition, integration and curation, as well as tackling the technical challenges associated with privacy, security and ethics. The volume and variety of sensitive data require advanced data governance protocols, including addressing privacy and security concerns while ensuring patient consent for data sharing. The final data ecosystem must facilitate the sharing of multimodal big datasets by the wider research and clinical communities.

**Equation 3: Contrastive multimodal learning (CLIP/MedCLIP-style alignment)**

**Setup**

Given a batch of paired items $\{(x_i, y_i)\}_{i=1}^{N}$(e.g., image + text, scan + report):

$$u_i = f(x_i), v_i = g(y_i)$$

Often normalize:

$$\bar{u}_i = \frac{u_i}{\| u_i \|}, \bar{v}_i = \frac{v_i}{\| v_i \|}$$

Similarity (cosine / dot):

$$s_{ij} = \bar{u}_i^{\top} \bar{v}_j$$

**InfoNCE (image→text direction)**

Define probability the correct match is $j = i$:

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

$$p(j = i \mid i) = \frac{\exp(s_{ii}/\tau)}{\sum_{j=1}^{N} \exp(s_{ij}/\tau)}$$

where $\tau > 0$is temperature.

Loss:

$$\mathcal{L}_{x\to y} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(s_{ii}/\tau)}{\sum_{j=1}^{N} \exp(s_{ij}/\tau)}$$

Similarly $\mathcal{L}_{y\to x}$. Final symmetric loss:

$$\mathcal{L} = \frac{1}{2}\left(\mathcal{L}_{x\to y} + \mathcal{L}_{y\to x}\right)$$

**4.1. Data Acquisition, Curation, and Quality Assurance**

Massive amounts of accessible biomedical multimodal data are becoming available as a result of open-access initiatives, increasing biomedical research's need to retrieve, curate, and guarantee data quality. The foundation of these initiatives is multimodal big data, which has been planned and generated by governmental or non-profit organizations to support biomedical research. Integrated analyses that improve precision medicine necessitate multimodal big data such as genomics, transcriptomics, proteomics, risk factors, medical imaging, electronic health records, clinical test reports, and so on, and these types of data are constantly evolving. For instance, the public repositories of radiology imaging and radiomics data, including The Cancer Imaging Archive and Medical Information Mart for Intensive Care database, are actively supported by many institutions. Recent guidelines for HDT-RNA sequence quality control and the survey of public ribo-seq datasets highlight the current demand for quality control, curation, and annotation of transcriptomic and ribo-seq data. These integrated analyses, such as fresh transcriptome profiling associated with deep-learning prediction of clinical outcomes of patients undergoing esophagectomy, would significantly benefit from these efforts. However, process-oriented curation and quality control of such data, especially for radiomics and transcriptomics, remain deficient.

Development of quality-assured multimodal datasets in precision medicine is also becoming increasingly important. Failure of deep-learning methods in genomics and transcriptomics—areas that have been experiencing an exponential increase in public databases and resources—have sparked a debate about the supposed data hunger of deep-learning methods trained on rRNA sequences. Biases in radiomics could lead to dramatic performance variations of different models. Such problems call for quality assurance processes with perceptual control checks, guidelines for image acquisition, data sharing, and large-scale validation sets with various populations and imaging equipment. Guidelines for multimodal ribo-seq data curation emphasize the need for unified datasets to reduce bias across public databases and focus on how sample origins and chemostat control impact data quality.

**4.2. Privacy, Security, and Ethical Considerations**

When sensitive data are collected, appropriate measures must be undertaken to prevent divulgence of private information. Developers of precision medicine systems should pay particular attention to the establishment of strategic partnerships with

ISSN: 3067-4166

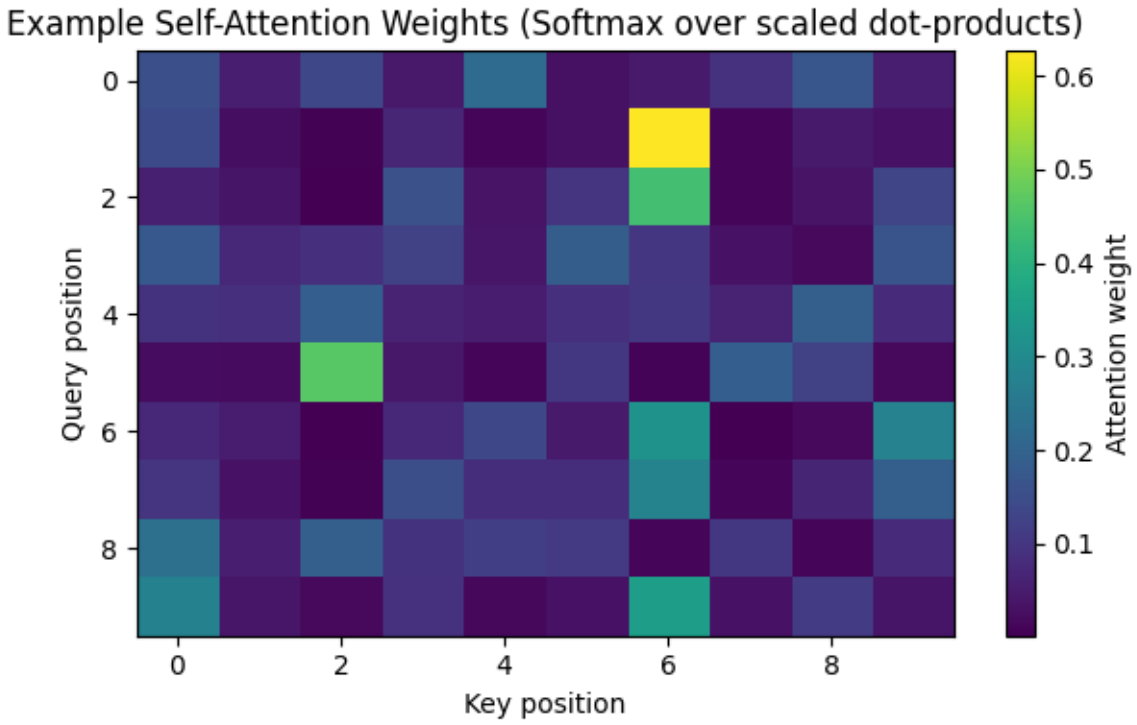
AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

research institutions, hospitals, and other organizations in order to gain access to private health data under secrecy contracts, allowing the development and training of models without the risk of compromising the privacy of individuals.

It is important to recognize the similarity between the data collected by the aforementioned institutions for research purposes and that obtained through commercial endeavors. Specifically, web scraping also enables the collection of clinical data, a fact that emphasizes the necessity for an ethical discussion regarding the commercial exploitation of such data in relation to the wellbeing of society.




## 5. Applications in Precision Medicine

Specific applications of deep learning in precision medicine span many biomedical domains. At a high level, research can be categorized according to the main types of data involved. Genomics — and more broadly, omics — data are predictors in a wide range of conditions, with transcriptomics being particularly important in the COVID-19 pandemic. Integration of transcriptomics with imaging and radiomics represents a promising area of multimodal deep learning for precision medicine. Multi-organ Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) data have also increasingly become the subject of dedicated studies, including radiomics-based approaches.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

Driven by the recent rapid advances in multimodal MRI data acquisition, segmentation, annotation, and deep-learning-based representation learning and generative modelling, the development of novel multi-organ segmentation protocols for non-contrast-enhanced 3D clinical MRI remains an important avenue of research considering the translational gap between bench and clinical practices. Deep learning holds great promise in improving the efficiency and efficacy of intravenous contrast agent use for multi-organ imaging; it can reduce patient exposure to renal-damaging contrast agents by integrating multi-organ MRI-derived radiomics with corresponding CT features. Deep learning also supports non-invasive prediction of the histological parameters of rectal cancer directly from MRI scans — another key step towards integrated, multimodal, and non-invasive centres for precision medicine in oncology.


**Fig 3: Applications in Precision Medicine of Foundation Deep Learning**

**5.1. Genomics and Transcriptomics Integration**

Recent years have witnessed an unprecedented drop in the costs of DNA sequencing that aided rapid growth in the genomic data for multiple organisms. In particular, in GBD's Bacteria and Archaea Data release v7 for the first time 5.2 million assembled genomes are accessible for users. This ongoing phenomenon of growing data motivates the need for new methods to utilize both the vast amounts of genomic data and the rapidly accumulating transcriptomic datasets. Due to their crucial roles in gene expression and regulation, the co-existing information from both genomic and transcriptomic datasets can be effective complements to each other, thereby leading us to explore methods to integrate both types of knowledge so that they can facilitate each other. Accordingly, a Novel Gaining Greatest Information in Transcriptome Modeling (NGGI-TM) technique is proposed for integrating transcriptomic data from different biological conditions within the context of a specific species and mining potential associations between gene expression and genomic features across a wide range of species.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

The foundation is modeled as a hypergraph, with the biological samples serving as nodes in the hypergraph and the various biological conditions that characterize the samples connected by hyperedges in the hypergraph. When two or more biological conditions with a sufficient number of samples co-occur in the same species, transcriptome profiles of that species can be used to build a transcriptome model for that condition. Focused on the bacterial taxonomic family Bacteroidaceae, plants and mammals are respectively tested for potential association between metabolic genes and shikimic acid metabolism/transcriptome expression and between DNA methylation system genes' expression and the locations of tissues/organs hosting correspondingly high methylation levels by evaluating whether gene expression has a predictive effect on the corresponding features. The association results indicate the potential application of using transcriptome data in association with genomic features and ability of presenting new findings or validations that could not be easily explored by simply mining the existing transcriptomic data of a single condition. The proposed method thus lays a foundation for further research by providing an additional way to experimentally analyze these associations.

**Equation 4: Hypergraph modeling (explicitly described for NGGI-TM)**

**Hypergraph basics**

- $V$: nodes (samples), $|V| = n$
- $E$: hyperedges (conditions), $|E| = m$
- Incidence matrix:

$$H \in \mathbb{R}^{n \times m}, H(v,e) = \begin{cases} 1 & \text{if node } v \in e \\ 0 & \text{otherwise} \end{cases}$$

Node degrees:

$$d(v) = \sum_{e} w(e) H(v,e)$$

Hyperedge degrees:

$$\delta(e) = \sum_{v} H(v,e)$$

Let:

- $D_v = \text{diag}(d(v))$
- $D_e = \text{diag}(\delta(e))$
- $W = \text{diag}(w(e))$

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

### 5.2. Imaging and Radiomics

As illustrated earlier, the power of phenomics lies not only in patient stratification but also in prognostic, predictive, and personalized therapeutic modeling. Despite the natural emergence of phenomics from imaging in clinical work, early development has used genomics as the leading edge to democratize efforts across datasets consisting of brain tumor images designated only for image-based experimentation. Nevertheless, MRI represents a powerful imaging modality that can provide surrogate information on multiple kinds of cellular and molecular constituents in human brain tissues, indirectly offering a reasonable idea for treatment design. Large cohorts in the referenced studies enable refined modeling for better prediction and treatment selection.

Multimodal data typed index, capability checklist, and typical data usage pattern provide a foundation for systematic exploration. Text mining of imaging radiomics and genomics papers reveals emerging uses. Radiomics is often used for tumor microenvironment characterization and prognostic modeling.

## 6. Challenges and Limitations

Even with its potential, deep learning in precision medicine is not without challenges and limitations. Data-preparation steps—including collection, acquisition from third-party providers, cleaning, and processing—are time-consuming and, as current studies have shown, can often introduce bias. Furthermore, the quality of clinical datasets determines the generalizability of respective applications, and model performance may drop when models are deployed in different clinical environments. Hence, bias quantification and enhancement of data diversity, fairness, and generalizability in respect to quantity and quality remain important tasks. Other considerations include making the medical and histopathological bases and rationale of data-driven insights explainable and understandable so that clinicians can process the results intelligently and with confidence.

Proposed or created deep learning models and methods have demonstrated that such challenges can be effectively addressed and that a foundation-model approach across deep learning methods enhances model performance and enables novel clinical applications. A robust foundation-model approach can also be seen as supporting the creation of a model zoo for various foundation models applied to special domains in precision medicine. In addition, specific sections of precision-medicine foundational deep learning research can be applied in different disciplines and departments, such as medical imaging, radiomics, genomics, and transcriptomics, thereby broadening and deepening the specialization of respective foundation-model studies. Other directions of future deep-learning studies in precision medicine include benchmark protocols for different clinical scenarios and disciplines, optimization of experimental-settings protocols, creation of annotated datasets, development of open-source codes, and verification of models' generalizability, and accuracy under less representative conditions.

### 6.1. Data Bias, Fairness, and Generalizability

Despite notable advances, ML applications often suffer from bias. Representation of a given race or ethnic group within a training dataset must be sufficiently large for a model to generalize well. Unbalanced or skewed datasets can lead to models that achieve superior performance across the majority class, while failing to properly model underrepresented classes. In practice, ML classifiers trained on unimodal pathogens, for instance, are significantly less accurate for virulent strains of other pathotype subgroups that are present in disproportionately lower numbers. Equally, generative diffusion models pretrained on large collections of natural images can cause concerns in terms of the likelihood of propensity for certain races being over- or underrepresented in the cinema when fine-tuning on small datasets, alongside the debiasing challenges in upsampling datasets or

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

improving classifiers for fairness. Nevertheless, failing to account for such issues may lead to catastrophic real-world outcomes for vulnerable populations. Consequently, framework for data bias assessment and mitigation, embracing techniques from different domains, is becoming paramount.

Bias is not confined to representational under-representation—it can also arise from the content of the data itself. Additionally, recent studies have questioned the generalizability of foundation models pretrained on natural images to other kinds of downstream visual tasks and datasets, and have suggested that incorporating representations from other data sources or modalities during transfer learning or fine-tuning may be beneficial to address potential limitations of pretrained weights and achieve improved performance. Such characteristics should not be ignored in research in precision medicine that explores the transferability of foundation models across different domains of Big Health Data.

**Equation 5: Multi-task learning objective (foundation model training paradigm)**

Foundation Deep Learning Models…

If tasks are $t = 1, \dots, T$, with task losses $\mathcal{L}_t$:

$$\mathcal{L}_{\text{total}} = \sum_{t=1}^{T} \lambda_t \mathcal{L}_t$$

Common examples:

- classification (cross-entropy):

$$\mathcal{L}_{CE} = -\sum_{c} y_c \log \hat{p}_c$$

- regression (MSE):

$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i} (y_i - \hat{y}_i)^2$$

**6.2. Interpretability and Clinician Trust**

Interpretability and auditability for foundation models have emerged as critical themes in research on large language models (LLMs), yet the underlying concepts differ from traditional machine learning interpretability definitions. When examining biomedical foundation models, two layers of trust are critical. Trust in the output requires a greater understanding of decision-making patterns, leading to the ability to explain predictions. In turn, clinician trust in foundation models relies on correct predictions as well as the interpretability of guidance and tools developed around them.

Clinician trust is paramount to the clinical introduction of any new method or application. A recommend model's predictions are of secondary concern to providing a near-certain four-tiered diagnosis: that the patient does indeed have or not have the condition and is or is not being misdiagnosed. The model's role is guidance on the decision, performed by an experienced clinician trained

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

for such assessments but augmented with a model capable of checking mistakes against prior data. When multiple sources of clinician error are examined, the output can be trusted in different ways. Foundation model guidance requires the utmost scrutiny, of both results and development. Building confidence that the model is indeed supporting the final decision is paramount, often more important than the decision quality.

## 7. Methodological Advances and Experimental Design

Overcoming the inherent limitations of training and validating foundation models under precisely controlled experimental conditions is a fundamental challenge in the field, setting a high bar for interpretable and generalisable results. With a view to advancing practical applications of foundation models in precision medicine, state-of-the-art benchmarking resources are therefore indispensable. Recent efforts to establish such scaffolding focus on three concepts: first, delineating a robust set of properties—rather than aspects alone—associated with quality foundation models, which encompasses representation power, incident-task performance, transferability, computational cost, and process interpretability; second, providing benchmarking protocols and resources that minimise—though cannot entirely eliminate—the influence of uncontrolled covariates on method comparisons; and third, ensuring that tested models are freely accessible along with their full training or fine-tuning datasets and protocols.

Multiple open-source databanks that facilitate and enhance model interpretation and applicability within the precision-medicine domain are also emerging. Such platforms clarify what is being modelled, highlight any potential shortcomings of the foundation model, and explain how the model can support subsequent precision-medicine investigations, benchmarks, and applications for which it was not specifically developed. Integrating quality-assurance measures into data governance pathways lays the groundwork for reliable heterogeneous- or cross-modal AI inprecision medicine, while implementing thorough documentation, wider accessibility, and improved model-interrogation protocols promotes user understanding of when and why models may falter.

Multiple open-source databanks are increasingly shaping the landscape of precision medicine by providing transparent, well-curated resources that strengthen both model interpretability and downstream applicability. These platforms make explicit what biological or clinical phenomena are being modeled, expose potential biases or blind spots in foundation models, and contextualize how such models can be responsibly repurposed for secondary analyses, benchmarking, and real-world clinical decision support. Embedding robust quality-assurance procedures into data governance frameworks—such as standardized validation, provenance tracking, and continuous auditing—creates a foundation for dependable heterogeneous and cross-modal AI systems. At the same time, comprehensive documentation, broad and equitable access, and advanced model-interrogation tools empower researchers and clinicians to understand model behavior, anticipate failure modes, and judge suitability for specific clinical contexts. Collectively, these practices foster trust, reproducibility, and more effective translation of AI innovations into precision-medicine research and care.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

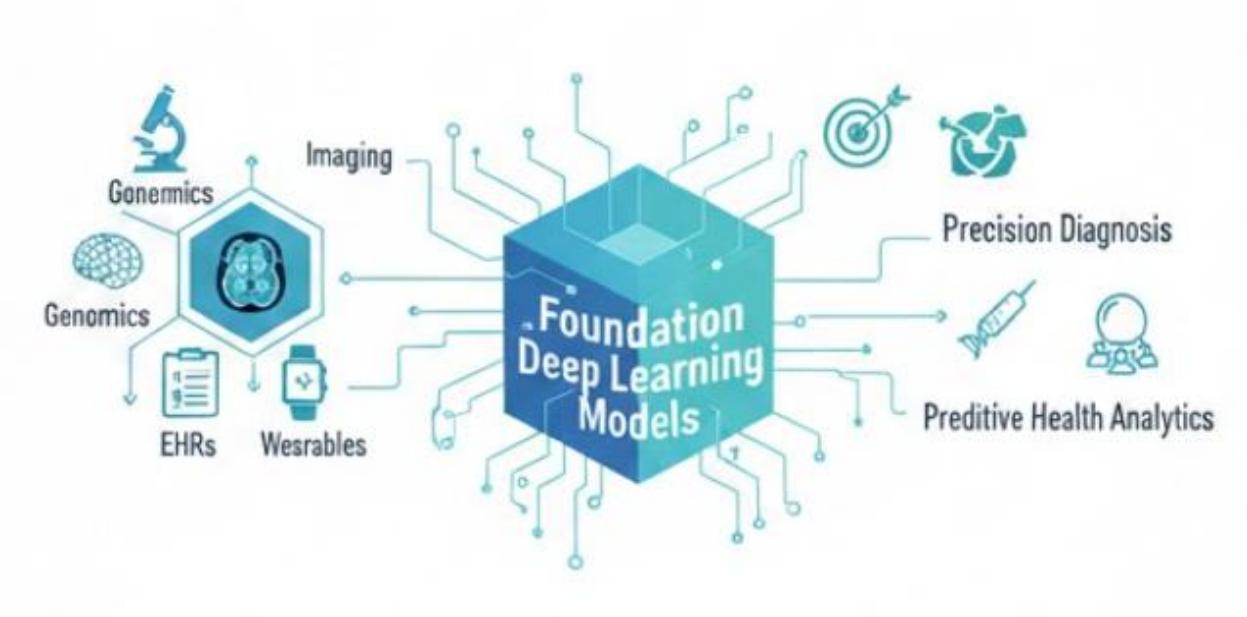



**Fig 4: Methodological Advances and Experimental Design**

**7.1. Benchmarking Protocols and Datasets**

A standardized benchmarking protocol for evaluating the performance of multimodal foundation models remains open. Such a protocol should encompass a common set of tasks and a collection of specialized datasets from multiple heterogeneous sources. Designing reusable datasets with clear task definitions for training, testing, and validating precision medicine foundation models will play a pivotal role in stimulating future methodological advances.

Generative pretraining using contrastive learning presents a novel approach to multimodal foundation model training for clinical genetics. The pretraining task combines codes from disease-risk loci with publicly available whole-genome sequencing data, Interaction, and UK Biobank demographics, Environment, Sequence expression, and Socioeconomic status information. The unique population context of Cardiovascular Disease (CVD) in Taiwan provides a distinctive advantage in the prediction of CVD risk.

**7.2. Reproducibility and Open Science Practices**

An emphasis on reproducibility in the deep-learning literature can mitigate some of the challenges outlined in the preceding section. Transparent benchmarking protocols describing implementation details are therefore essential. Furthermore, these protocols should include standardized datasets used for experimentation so that performance claims can be quantified against common baselines under the same conditions. Other reproducibility tools, including open-source code repositories and collaborative platforms such as Papers with Code, can facilitate protocol implementation and public scrutiny of experimental results.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

To support independent validation, publicly available implementation codes and model weights with instruction files are essential. Adhering to ethical guidelines established by domains affected by ongoing research, these assets will allow the research community to leverage pretraining in pursuit of new advances. Moreover, Investors-committee recommendations for open science remain poorly practiced despite their foundational importance. The work will thus conform to their full spectrum, maximally expanding the availability and utility of all data and method assets.

## 8. Conclusion

This research serves as a foundation for future applications of foundation deep learning models using multimodal biomedical data for precision medicine. The scope of these models aligns closely with ongoing large-scale efforts to build foundation models for natural language processing and computer vision and is indeed inspired by considerable success and flourishing interest in domains other than biomedicine. The models, methods, and areas of application provide an initial basis for such research. Deep learning for precision medicine is in a key transitional phase in which considerations of data quantity and quality are rapidly shifting due to the deployment of large-scale, high-quality efforts.

Yet imminent large-scale healthcare datasets come with challenges. These include data bias affecting the accessibility and generalizability of models, a lack of interpretability of complex models limiting clinician trust, and a call to closely explore methods outside supervised deep learning for data-rich applications. Addressing these issues is essential for deep learning to be used as a trusted and efficient assistant for clinicians and radiologists, rather than as a black-box decision-making system. Outlining future data-rich applications of deep learning in precision medicine and benchmarking efforts to drive progress is thus crucial.

### 8.1. Future Directions

An emerging framework for exploiting multimodal biomedical data in precision medicine has been presented, underscored by the establishment of foundation deep learning models and multimodal foundation models. Benchmarked core architectures serve as a reference for task-specific models operating on audio, imaging, radiomics, and genomics data. Ongoing experimental efforts are extending these predictions across transcriptomic and electrocardiogram data, bolstered by an ecosystem involving data acquisition, curation, privacy--preserving sharing, and governance. Capturing the complementary information within the genomics and transcriptomics layers, and integrating imaging, radiomics, and other modalities encoded in foundation models, such as Electrocardiogram Assistant or MriAssistant and Dr-Kim-Multi-Modal-Language-Model, will address the multimodality nature of precision medicine.

All these models will serve as the basis for future research in specific applications, enabling breakthroughs for precision medicine. However, the data bias classifications underlying training data distributions, the depth and dimensionality of data modality, and the resulting representation of the deep learning models must be considered. Achieving trustworthy clinical application requires satisfying the four criteria: (1) the model can accurately predict unknown data; (2) the prediction is suitable for the clinical setting; (3) the prediction is interpretable for clinical consultation; and (4) the prediction is trustworthy for clinical decision-making.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

## 9. References

1. Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., & Stegle, O. (2020). MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biology, 21, 111.
2. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234–1240.
3. Chen, R. J., Lu, M. Y., Wang, J., Williamson, D. F. K., Rodig, S. J., Lindeman, N. I., & Mahmood, F. (2020). Pathomic Fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Transactions on Medical Imaging, 41(4), 757–770.
4. Mashetty, S., Malempati, M., Paleti, S., Adusupalli, B., & Singireddy, J. (2025). A Multidisciplinary Framework for AI and Data-Driven Transformation in Taxation, Insurance, Mortgage Financing, and Financial Advisory: Integrating Cloud Computing, Deep Learning, and Agentic AI for Community-Centric Economic Development. Insurance, Mortgage Financing, and Financial Advisory: Integrating Cloud Computing, Deep Learning, and Agentic AI for Community-Centric Economic Development.
5. Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., & Kather, J. N. (2021). Deep learning in cancer pathology: A new generation of clinical biomarkers. British Journal of Cancer, 124, 686–696.
6. Wang, T., Shao, W., Huang, Z., Tang, H., Zhang, J., Ding, Z., & Huang, K. (2021). MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nature Communications, 12, 3445.
7. Burugulla, J. K. R., Kannan, S., Malempati, M., Challa, H. A. H. S. R., Pandugula, C., & Goma, T. (2025, April). Federated Learning Based Cloud Computing Solutions with Privacy Preservation for Intelligent Utilities. In 2025 4th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0 (pp. 1-6). IEEE.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

8. Gayoso, A., Steier, Z., Lopez, R., Regier, J., Nazor, K. L., Streets, A., & Yosef, N. (2021). Joint probabilistic modeling of single-cell multi-omic data with totalVI. Nature Methods, 18, 272–282.
9. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning, 139, 8748–8763.
10. Adusupalli, B., Malempati, M., Paleti, S., Mashetty, S., & Singireddy, J. (2025). Integrated financial ecosystems: AI-driven innovations in taxation, insurance, mortgage analytics, and community investment through cloud, big data, and advanced data engineering. Journal of Information Systems Engineering and Management, 10, 1103-1117.
11. Thornton, J. M., Laskowski, R. A., & Borkakoti, N. (2021). AlphaFold heralds a data-driven revolution in biology and medicine. Nature Medicine, 27, 1666–1669.
12. Gu, Y., Tinn, R., Cheng, H., Lucas, M., Usuyama, N., Liu, X., Naumann, T., Gao, J., & Poon, H. (2021). Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare, 3(1), 1–23.
13. Sivanand, R., Kumar, D. P., Nagabhyru, K. C., Natarajan, E. P., Pamisetty, V., & Kapila, D. (2025, September). IoT and AI for Real-Time Monitoring in Substation Automation. In 2025 International Conference on Computing and Communications (COMPUTINGCON) (pp. 1-5). IEEE.
14. Lotfollahi, M., Naghipourfar, M., Luecken, M. D., Khajavi, M., Büttner, M., Wagenstetter, M., Avila Cobos, F., Wessels, L. F. A., Buettner, F., & Theis, F. J. (2021). Mapping single-cell data to reference atlases by transfer learning. Nature Biotechnology, 39, 370–380.
15. Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., & Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review. npj Digital Medicine, 5, 171.
16. Suura, S. R., Chava, K., Chakilam, C., Nuka, S. T., Maguluri, K. K., & Goma, T. (2025, April). Blockchain-Based Secure and Scalable Models for Healthcare Network Traffic Monitoring and Optimization. In 2025 4th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0 (pp. 1-6). IEEE.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

17. Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T.-Y. (2022). BioGPT: Generative pre-trained transformer for biomedical text generation and mining. Briefings in Bioinformatics, 23(6), bbac409.
18. Yang, X., Chen, A., PourNejatian, N., Shin, H. C., Smith, K. E., Parisien, C., Compas, C., Martin, C., Costa, A. B., Flores, M. G., Zhang, Y., Magoumou, G., Bui, T. D., Semturs, C., Rodriques, S. G., Ranganath, R., & others. (2022). A large language model for electronic health records. npj Digital Medicine, 5, 194.
19. Chakilam, C., Kannan, S., Recharla, M., Suura, S. R., & Nuka, S. T. (2025). The impact of big data and cloud computing on genetic testing and reproductive health management. American Journal of Psychiatric Rehabilitation, 28(1), 62-72.
20. Alayrac, J.-B., Donahue, J., Luc, P., Miech, A., Barr, I., Hasson, Y., Lenc, K., Mensch, A., Millican, K., Reynolds, M., Ring, R., Rutherford, E., Cabi, S., Han, T., Gong, Z., Samangooei, S., Monteiro, M., Menick, J. L., Borgeaud, S., Brock, A., Nematzadeh, A., Sharifzadeh, S., Binkowski, M., Barreira, R., Vinyals, O., Zisserman, A., & Simonyan, K. (2022). Flamingo: A visual language model for few-shot learning. Advances in Neural Information Processing Systems, 35, 23716–23736.
21. Li, J., Li, D., Xiong, C., & Hoi, S. (2022). BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. Proceedings of the 39th International Conference on Machine Learning, 162, 12888–12900.
22. Kummari, D. N., Singireddy, J., Sheelam, G. K., Nandan, B. P., Pandiri, L., Lakkarasu, P., & Dwaraka. (2025, August). Generative AI Models for Process Optimization in Semiconductor Wafer Design and Yield Prediction. In International Conference on Artificial Intelligence: Theory and Applications (pp. 220-233). Cham: Springer Nature Switzerland.
23. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., & Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science, 379(6637), 1123–1130.
24. Ashuach, T., Gabitto, M. I., Koodli, R. V., Saldi, G.-A., Jordan, M. I., & Yosef, N. (2023). MultiVI: Deep generative model for the integration of multimodal data. Nature Methods, 20, 1222–1231.
25. Pamisetty, V. (2019). Machine Learning Models for Real-Time Tax Fraud Detection and Risk Assessment in Digital Government Systems. Global Research Development (GRD) ISSN, 2455-5703.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

26. Wang, X., Wu, Z., Liu, G., Li, J., Zhang, Y., Wang, Y., & others. (2023). Multimodal data fusion for cancer biomarker discovery with deep learning. Nature Machine Intelligence, 5, 351–362.
27. Li, J., Li, D., Savarese, S., & Hoi, S. (2023). BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. Proceedings of the 40th International Conference on Machine Learning, 202, 19730–19742.
28. Theodoris, C. V., Xiao, L., Chopra, A., Chaffin, M. D., Al Sayed, Z. R., Hill, M. C., Mantineo, H., Brydon, E. M., Zeng, Z., Liu, X., & others. (2023). Transfer learning enables predictions in network biology. Nature, 618, 616–624.
29. Kalisetty, S., Lakkarasu, P., Singireddy, S., Burugulla, J. K. R., Challa, K., & Gadi, A. L. (2025, August). Next-Gen Payment Gateways: Leveraging Federated Learning for Fraud Detection in Cross-Border Transactions. In International Conference on Artificial Intelligence: Theory and Applications (pp. 200-211). Cham: Springer Nature Switzerland.
30. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., Agüera y Arcas, B., Webster, D., Corrado, G. S., Matias, Y., Chou, K., Gottweis, J., Tomasev, N., Liu, Y., Rajkomar, A., Barral, J., Semturs, C., Karthikesalingam, A., & Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620, 172–180.
31. Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., & Wang, B. (2024). scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. Nature Methods, 21, 1470–1480.
32. Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X., Wang, T., Ma, J., Zhang, X., & Song, L. (2024). Large-scale foundation model on single-cell transcriptomics. Nature Methods, 21, 1481–1491.
33. Paleti, S., Baliyan, M., Aitha, A. R., Reddy, B. A., Bhadauria, G. S., & Sing, S. A. (2025, August). Graph—LSTM Hybrid Model for Improving Fraud Detection Accuracy in E-Commerce Financial Services. In 2025 2nd International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS) (pp. 1-6). IEEE.
34. Warner, E., Lee, J., Hsu, W., Syeda-Mahmood, T., Kahn, C. E., Gevaert, O., & Rao, A. (2024). Multimodal machine learning in image-based and clinical biomedicine: Survey and prospects. International Journal of Computer Vision, 132, 3753–3769.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

35. Steyaert, S., Pizurica, M., Nagaraj, D., Khandelwal, P., Hernandez-Boussard, T., Gentles, A. J., & Gevaert, O. (2023). Multimodal data fusion for cancer biomarker discovery with deep learning. Nature Machine Intelligence, 5, 351–362.
36. Mak, R. H., Birkbak, N. J., & Aerts, H. J. W. L. (2024). Foundation model for cancer imaging biomarkers. Nature Machine Intelligence, 6, 354–367.
37. Warner, E., Lee, J., Hsu, W., Syeda-Mahmood, T., Kahn, C. E., Gevaert, O., & Rao, A. (2024). Multimodal machine learning in image-based and clinical biomedicine: Survey and prospects. International Journal of Computer Vision, 132, 3753–3769.
38. Wang, P., Zhang, H., & Yuan, Y. (2024). MCPL: Multi-modal collaborative prompt learning for medical vision-language model. IEEE Transactions on Medical Imaging, 43(12), 4224–4235.
39. Mashetty, S. (2025). Technology-driven analytics in mortgage-backed securities for single-family mortgage financing. Available at SSRN 5236573.
40. Ferber, D., Wölflein, G., Wiest, I. C., Ligero, M., Sainath, S., Ghaffari Laleh, N., El Nahhas, O. S. M., Müller-Franzes, G., Jäger, D., Truhn, D., & Kather, J. N. (2024). In-context learning enables multimodal large language models to classify cancer pathology images. Nature Communications, 15, 10104.
41. Pandya, P., Kather, J. N., & others. (2024). A systematic pan-cancer study on deep learning-based prediction of multi-omic biomarkers from routine pathology images. Communications Medicine, 4, 48.
42. Steyaert, S., Pizurica, M., Nagaraj, D., Khandelwal, P., Hernandez-Boussard, T., Gentles, A. J., & Gevaert, O. (2024). Harnessing artificial intelligence in multimodal omics data integration: Paving the path for the next frontier in precision medicine. Annual Review of Biomedical Data Science, 7, 195–219.
43. Shi, D., Zhang, W., Yang, J., Huang, S., Chen, X., Xu, P., Jin, K., Lin, S., Wei, J., Yusufu, M., Liu, S., Zhang, Q., Ge, Z., Xu, X., & He, M. (2025). A multimodal visual-language foundation model for computational ophthalmology. npj Digital Medicine, 8, 302.
44. Xiang, J., Wang, X., Zhang, X., Xi, Y., Eweje, F., Chen, Y., Li, Y., Bergstrom, C., Gopaulchan, M., Kim, T., Yu, K.-H., Willens, S., Olguin, F. M., Nirschl, J. J., Neal, J., Diehn, M., Yang, S., & Li, R. (2025). A vision-language foundation model for precision oncology. Nature, 638, 769–778.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 04 ISSUE: 01

RECEIVED: JANUARY 02 REVISED: FEBRUARY 19 ACCEPTED: FEBRUARY 27 PUBLISHED: MARCH 17

45. Ryu, J. S., Kang, H., Chu, Y., & Yang, S. (2025). Vision-language foundation models for medical imaging: A review of current practices and innovations. Biomedical Engineering Letters, 15, 809–830.
46. Krones, F., Marikkar, U., Parsons, G., Szmul, A., & Mahdi, A. (2025). Review of multimodal machine learning approaches in healthcare. Information Fusion, 114, 102690.