Foundation Models for Advancing Healthcare: Challenges, Opportunities and Future Directions

Authors: Yuting He, Fuxiang Huang, Xinrui Jiang, Yuxiang Nie, Minghao Wang, Jiguang Wang, Hao Chen

Year: 2024
Venue: arXiv (survey)

1. Classification

- **Domain Category:**
  - General FM survey / theory
  - This paper surveys healthcare foundation models (HFMs) across language, vision, bioinformatics, and

multimodal domains, and discusses their challenges and future directions.

- **FM Usage Type:**
  - Multimodal FM or cross-modal integration (conceptual survey)
- **Key Modalities:**
  - Text (clinical notes, biomedical literature), medical images (radiology, pathology, ophthalmology, etc.), bioinformatics data (DNA, RNA, protein sequences), and multimodal clinical data (EHR, physiology).

2. Executive Summary

This survey provides a broad overview of how foundation models are transforming healthcare, framing them as healthcare foundation models (HFMs) that can be adapted to many clinical tasks. The authors first define HFMs and trace their roots from early transfer learning to modern large language models (LLMs), vision foundation models (VFMs), bioinformatics foundation models (BFMs), and multimodal foundation models (MFMs). They then review representative models and datasets in each subfield, highlighting successes such as BioBERT, AlphaFold2, SAM-style medical VFMs, and multimodal models integrating images, text, and omics. The paper devotes substantial discussion to the challenges of data, algorithms, and computing infrastructure—ethics, heterogeneity, cost, and environmental impact—as well as open questions around fairness, robustness, and deployment. Finally, it outlines promising research directions, arguing that HFMs can drive the next generation of precision medicine if developed and governed responsibly. For a new grad student, this article is a high-level map of the healthcare FM landscape.

3. Problem Setup and Motivation

- **Scientific / practical problem:**
  - Understand how **foundation models**—pretrained on broad healthcare data and adaptable to many tasks—can advance diagnosis, prognosis, treatment, and research.
  - Identify current capabilities, limitations, and research directions across language, vision, bioinformatics, and multimodal healthcare AI.
- **Why this is hard:**
  - **Data-related challenges:**
    - Healthcare data are sensitive, heterogeneous, and often siloed, complicating large-scale collection and sharing.
    - Ethical, legal, and social concerns (privacy, consent, bias) strongly constrain data use.
  - **Algorithmic challenges:**
    - HFMs must balance scale and adaptability with reliability, interpretability, and safety.
    - Domain shift, spurious correlations, and label noise can degrade performance when models are deployed.
  - **Computing and infrastructure:**
    - Training large FMs on 3D images, WSIs, or multi-omics data is computationally and environmentally expensive.
    - Many healthcare institutions lack the infrastructure to host or fine-tune very large models.

- **Clinical constraints:**
  - HFMs must integrate into existing workflows and meet regulatory and validation requirements before impacting patient care.

---

4. Data and Modalities

The survey organizes healthcare data for HFMs into several key categories:

- **Text data:**
  - Biomedical literature, clinical guidelines, and medical textbooks.
  - Clinical notes (progress notes, discharge summaries, radiology and pathology reports).
  - Used to train medical LLMs and language-centric HFMs.
- **Medical imaging data:**
  - Radiology (CT, MRI, X-ray, ultrasound), pathology (whole-slide images), ophthalmology, endoscopy, etc.
  - Large imaging datasets with segmentation masks, labels, or reports support VFMs and MMVLFMs.
- **Bioinformatics and omics data:**
  - DNA, RNA, protein sequences, 3D protein structures, and other molecular data.
  - Used for BFMs such as protein or genomic FMs (e.g., AlphaFold2-like models).
- **Multimodal clinical data:**
  - EHR tables (labs, vitals, medications), physiological signals (ECG, EEG), and combined image–text–omics datasets.
  - Essential for MFMs that model whole-patient states.

- **Preprocessing / representation themes:**
  - Tokenization and vocabulary design for clinical/ biomedical text and sequences.
  - Patch and voxel-based representations for images and volumes.
  - Graphs, sequences, or 3D coordinate representations for proteins and molecular structures.
  - Harmonization and standardization across institutions for multi-center datasets.

5. Model / Foundation Model

- **Model Type (subfields reviewed):**

  - **Language FMs / LLMs:** medical variants of BERT, GPT, and other transformers (e.g., BioBERT, clinical LLMs).
  - **Vision FMs (VFMs):** models like SAM-style or ViT-based encoders for medical imaging.
  - **Bioinformatics FMs (BFMs):** AlphaFold2-like models, protein language models, DNA/RNA sequence models.
  - **Multimodal FMs (MFMs):** models integrating text, images, omics, and clinical variables.

- **Is it a new FM or an existing one?**

  - This is a **survey**, summarizing many HFMs rather than introducing a single new model.

- **Key conceptual components:**

Aspect	Details
Pretraining	Large-scale self-supervised or weakly supervised objectives on broad healthcare data
Adaptation	Fine-tuning, prompting, and parameter-efficient methods (e.g., adapters, LoRA)
Applications	Diagnosis, prognosis, treatment planning, report generation, drug discovery, and more
Evaluation	Domain-specific benchmarks, robustness tests, and clinical validation studies

- **Training setup themes:**

  - Self-supervised learning (masked language modeling, masked image modeling, contrastive learning).

- Multi-task and continual learning strategies to adapt HFMs to new tasks while retaining prior knowledge.

- Use of transfer learning and model reuse across related tasks and modalities.

6. Multimodal / Integration Aspects (If Applicable)

While much of the survey covers unimodal FMs, **multimodal HFMs** are a central focus for future directions.

- **Modalities integrated:**
  - Images + text (e.g., radiology report generation).

  - Text + omics (e.g., linking genomic variants with phenotypes).

  - Images + omics + clinical variables in comprehensive patient-level models.

- **Integration strategies discussed:**
  - Early fusion of embeddings from different modalities.

  - Cross-attention and joint latent spaces for intermediate fusion.

  - Late fusion via ensemble or decision-level integration for heterogeneous modalities.

  - CLIP-style and other contrastive learning approaches for image–text alignment.

- **New capabilities enabled:**
  - Richer patient representation and more accurate prediction by combining multiple data sources.

  - Cross-modal retrieval (e.g., finding images that match textual descriptions or molecular profiles).

- Enhanced interpretability when multimodal signals agree or disagree in clinically meaningful ways.

7. Experiments and Results

As a survey, the paper does not present new experiments, but **summarizes outcomes** from many HFMs.

- **Tasks reviewed:**
  - Improved diagnosis and prognosis from imaging and text.
  - Accurate protein structure prediction and molecular property forecasting.
  - Automated or assisted report generation and documentation.
  - Drug discovery and treatment recommendation tasks.
- **High-level trends:**
  - HFMs consistently outperform smaller, task-specific models when sufficient pretraining data are available.
  - Zero-shot and few-shot capabilities enable deployment in settings with limited labeled data.
  - However, performance can degrade sharply when task distributions differ from pretraining data, highlighting the importance of robust evaluation.

8. Strengths, Limitations, and Open Questions

Strengths (of HFMs and the survey):

- Presents a **comprehensive taxonomy** of HFMs across language, vision, bioinformatics, and multimodal domains.
- Explicitly links technical developments (e.g., transformers, self-supervision) to concrete healthcare applications.
- Identifies both opportunities (precision medicine, automation, discovery) and systemic challenges.
- Serves as a high-level entry point for researchers from different backgrounds.

Limitations and challenges (field-level):

- Data availability is constrained by **privacy, ethics, and heterogeneity**, limiting the diversity of pretraining corpora.
- Algorithmic issues such as **hallucination, bias, and lack of interpretability** are particularly serious in clinical contexts.
- Computing requirements raise questions about **sustainability, equity of access, and reproducibility**.
- Regulatory frameworks are still catching up with the capabilities and risks of HFMs.

Open Questions and Future Directions:

1. How can we design HFMs that are simultaneously **scalable, reliable, and interpretable** in clinical settings?

2. What governance structures and data-sharing mechanisms are needed to support **responsible, multi-institutional pretraining**?

3. How can HFMs be adapted and validated for **under-served populations and low-resource health systems**?

4. What new architectures and objectives are needed to incorporate **causal reasoning and mechanistic knowledge** into HFMs?

5. How should we evaluate the **real-world clinical impact** of HFMs beyond traditional accuracy metrics?

9. Context and Broader Impact

- **Position in the landscape:**
  - This survey sits at the intersection of AI and healthcare policy, offering a bird's-eye view of HFMs rather than focusing on a single modality or task.
  - It complements more focused reviews on medical MLLMs or MMFMs by connecting them to broader trends in healthcare AI.

- **Relation to well-known ideas:**
  - Builds directly on the concept of **foundation models** articulated by Bommasani et al., extending it to healthcare.
  - Discusses iconic HFMs such as BERT, CLIP, SAM, AlphaFold2, and domain-specific medical LLMs and VFMs.

- **Why this paper is a useful reference:**
  - Helps new researchers understand the **big picture** of how different kinds of HFMs fit together and where

healthcare is headed.

- Encourages critical thinking about the technical, ethical, and infrastructural requirements for deploying HFMs safely.

10. Key Takeaways (Bullet Summary)

- **Problem:**
  - Healthcare needs AI models that can generalize across diverse tasks and modalities, but current systems are often narrow and brittle.

- **Method / model (conceptual):**
  - Healthcare foundation models (HFMs) apply the foundation-model paradigm—large-scale pretraining and broad transfer—to clinical text, images, omics, and multimodal data.
  - The survey categorizes HFMs into language, vision, bioinformatics, and multimodal branches, highlighting representative models and datasets.

- **Results / insights:**
  - HFMs show strong performance and data efficiency, enabling zero-shot and few-shot adaptation across many healthcare tasks.
  - Yet they face significant challenges in data governance, algorithmic robustness, fairness, interpretability, and deployment.

- **Why it matters:**
  - HFMs are poised to become **central infrastructure for future healthcare AI**, but realizing this potential will require coordinated advances in data, algorithms, computing, and regulation.

- This survey offers a strategic overview for anyone aiming to contribute to that evolution.