基础模型推进医疗健康：挑战、机遇和未来

何宇霆, 黄福香, 姜馨蕊, 聂宇翔, 王明灏, 王吉光, 陈浩*, *Senior Member, IEEE*

摘要—基础模型在大规模的数据上进行预训练进而获得了适应大量潜在任务的能力，正在推动医疗健康的发展。它通过打破了有限的人工智能模型和多种多样的医疗任务之间的矛盾，从而促进医疗人工智能模型的发展。更广泛的医疗场景将从医疗基础模型（Healthcare foundation model, HFM）的发展中受益，促进智能医疗服务在这些场景的应用。尽管 HFM 已经展现出了广泛部署和应用的潜力，但目前人们对其在医疗领域的发展情况、当前的挑战以及未来的发展方向还缺乏清晰的理解。为了回答这些问题，本文对 HFM 的挑战、机遇和未来方向进行了全面深入的调查。我们首先对 HFM 进行了全面概述，包括方法、数据和应用，以便快速掌握当前的进展情况。然后，我们深入探讨了 HFM 在数据、算法和计算基础设施方面的挑战，以便人们了解当下的发展困境，指明了医疗健康领域构建和广泛应用基础模型的困难。本文最后展望了 HFM 在未来的新兴和有前途的发展方向。我们相信，这篇文章将增强社区对 HFM 当前进展的理解，并为该领域未来发展提供有价值的指导。最新的 HFM 论文和相关资源将在我们的网站上进行长期维护。

***Index Terms*—基础模型，人工智能，医疗健康**

I. 介绍

在过去的十年中，随着人工智能（Artificial intelligence, AI）[1] 特别是深度学习（Deep learning, DL）[2] 的发展，医疗技术已经得到了颠覆性的进展 [3]–[5]。受益于对医疗数据的学习，AI 模型能够解耦数据中的相关信息，并进而辅助医疗实践。在一些具有影响力的临床疾病中，包括胰腺癌 [6]、视网膜疾病 [7]、皮肤癌 [8] 等，AI 模型已经获得了近似医疗专家的专业性能，展现出了在该领域光明的应用前景。然而，在此之前，专用于特定医疗任务的专家 AI 模型与多样的医疗场景和需求之间仍存在着巨大的矛盾，阻碍了它们在广泛的医疗实践中的应用 [5]。因此，一个开放问题长期困扰着健康保健领域：“我们能否构建通用的 *AI* 模型来惠及各种医疗任务？”

如图1所示，基础模型的最新研究使得 AI 模型能够学习通用能力并应用于各种医疗场景，为这个问题提供了一个充满希望的答案 [9]–[12]。在包括语言、视觉、生物信息和多模态在内的医疗健康 AI 的相关子领域中，医疗基础模型（Healthcare foundation model, HFM）已经展现出令人印象深刻的成功。**a)** 语言基础模型（Language foundation model, LFM）或称大语言模型（Large language model, LLM）[13], [14] 已经引起了患者和临床医生广泛关注 [13]。它能够学习大规模的医疗语言数据，并在医疗文本处理 [15]、对话 [16] 这类语言任务中展现出了非凡的性能。**b)** 视觉基础模型（Vision foundation model, VFM）在医学图像领域展现出了巨大潜力。针对不同模态 [17], [18]、器官 [19] 和任务 [20], [21] 的 VFM 已经展现出了它们的适应性和对潜在的多种医疗场景的通用性。**c)** 生物信息基础模型（Bioinformatics foundation model, BFM）帮助研究人员揭示生命的秘密，为人们在蛋白质序列、DNA、RNA 等方面的研究提供了巨大的帮助 [22]–[26]。**d)** 多模态基础模型（Multimodal foundation model, MFM）[27]–[29] 为通用的 HFM [10], [30], [31] 提供了一种有效的方法。它同时学习了多个模态的信息，获得了理解多种医学模态的能力，从而能够执行多模态相关任务并获得更好的性能 [11], [31], [32]。因此，这些模型拥有了解决复杂临床问题的能力，为提高医疗实践的效率和效果构建了模型基础，从而推进医疗健康领域的发展 [11]。

HFM 的出现源于医疗数据的持续积累、AI 算法的发展和计算基础设施的提升 [9], [12]。然而目前在数据、算法和计算基础设施方面仍然存在巨大缺陷，然是 HFM 所面对的各种挑战的根源。医疗数据的伦理需求、多样性、异质性和高成本使得难以构建足够大的数据集来训练能够在各种医疗实践中有效应用 HFM [12], [33]。对于 AI 算法，对适应性、模型容量、可靠性和责任性的需求进一步使 HFM 难以应用于真实场景 [34],

*通讯作者：陈浩 (邮箱: *jhc@cse.ust.hk*)

陈浩隶属于香港科技大学计算机科学与工程系、化学与生物工程系和生命科学分部。

何宇霆、黄福香、姜馨蕊和聂宇翔隶属于香港科技大学计算机科学与工程系。

王明灏隶属于香港科技大学化学与生物工程系。

王吉光隶属于香港科技大学化学与生物工程系、生命科学分部，分子神经科学国家重点实验室，SIAT-HKUST 细胞进化与数字健康联合实验室，深圳-香港联合创新研究院，以及香港神经退行性疾病中心。

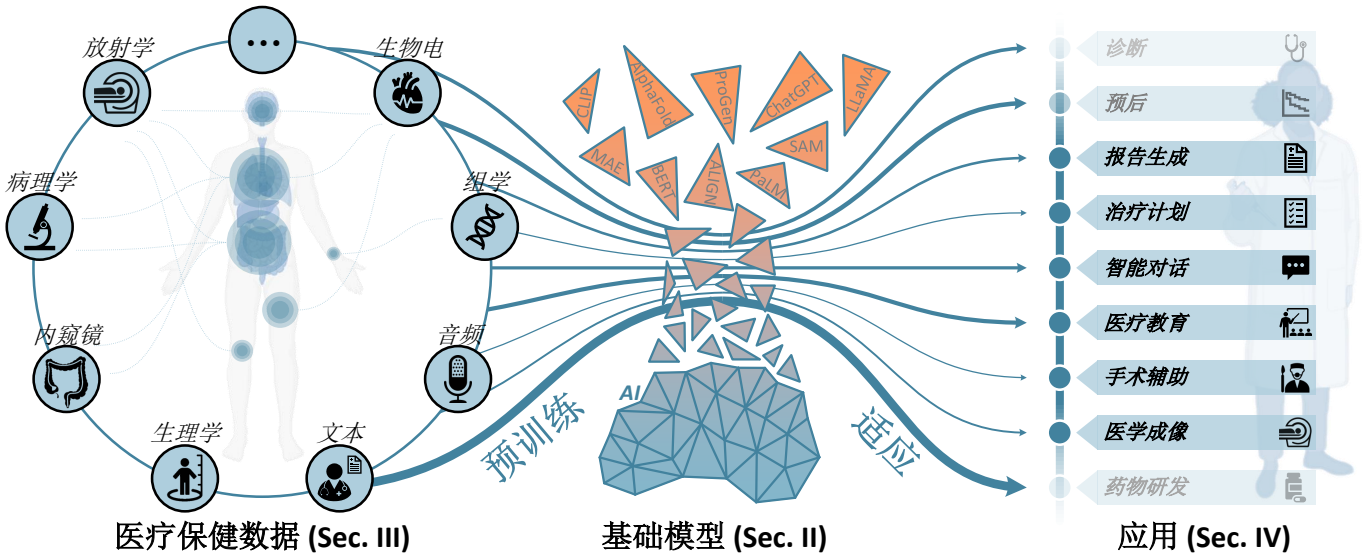

图 1. 医疗健康基础模型（HFM）的流程包括方法（第II节）、数据（第III节）和应用（第IV节）.

[35]。由于医疗数据的高维性和大尺度（例如 3D CT 图像、全切片图像（Whole slide image, WSI）等），HFM 对计算基础设施的需求远高于其他领域，这导致非常昂贵的计算成本 [10], [12] 和环境成本 [36]。

总的来说，基础模型正在推进医疗健康的发展，展现了一个充满机遇与挑战的新未来。在本综述中，我们提出了以下问题：**1)** 尽管基础模型已经取得了显著的成功，它们在医疗健康方面的当前进展是什么？ **2)** 随着基础模型的发展，它们正面临哪些挑战？ **3)** 为了推动 HFM 的进一步发展，哪些潜在的未来方向值得我们关注和探索？本文通过对上述问题的探索，概述了 HFM 的当前进展情况，并为其未来发展提供清晰的愿景。由于 HFM 的巨大潜力，在近年来已经产生了数百篇论文。因此，在有限的论文空间内审查所有这些论文和各个方面是具有挑战性的。在本文中，我们关注 2018 年（基础模型时代的开始 [9]）至 2024 年间基础模型在医疗健康领域（包括语言、视觉、生物信息和多模态方面）的进展，以及 HFM 所面临的挑战和未来方向。我们希望这项调查能够帮助研究人员快速掌握 HFM 的进展情况，并激发其创造力的火花，进一步推动医疗健康领域的发展。

A. 医疗健康基础模型简史

本综述根据 Bommasani 等人 [9] 的定义，“基础模型”是指任何在大规模数据上进行预训练，并具有适应到各种任务能力的模型。基础模型时代的另一个社会学特征 [9] 是人们广泛接受将某种基础 AI 模型应用于大量不同的任务。基础模型时代的代表性转折点是 2018 年末，自然语言处理（Natural language processing, NLP）领域提出 BERT 模型 [37]，在这之后，预训练模型作为基础模型被广泛应用于 NLP 领域，而后逐渐扩展到其他相关领域。

随着基础模型的快速发展，医疗健康中的人工智能也逐渐从专家模型转向通用模型 [10]。在 2019 年初，BioBERT [38] 基于 BERT [37] 随之问世，实现了医疗健康领域的基础模型。2022 年底，ChatGPT [39] 以其强大的通用性，使更多与医疗健康相关的从业人员受益于基础模型，从而吸引了他们的关注，并进一步引发了 HFM 的研究热潮。仅在 2023 年 8 月，就发表了 200 多篇与 ChatGPT 在医疗健康方面相关的研究 [12]。对于 VFM，许多初步的工作 [40], [41] 都集中在预训练或迁移学习上。由于分割一切模型（Segment anything model, SAM）[20] 的巨大影响，通用视觉模型也在医疗健康中 [42]–[44] 引发了一股研究热潮。在领域，AlphaFold2 [25] 在 2020 年的蛋白质结构预测 CASP14 中获得了第一名，引发了人们对 BFM 的巨大兴趣，并进一步推动了 RNA [45]、DNA [46]、蛋白质 [25] 等方面的研究。在 2021 年初，OpenAI 构建了 CLIP [47] 模型，实现了视觉和语言的大规模学习，取得了显著的性能。由于医疗健康数据天然的多模态特性，这项技术很快就被应用于医疗健康中 [48]，从而融合图像、组学、文本等多模态数据，构建强大的通才模

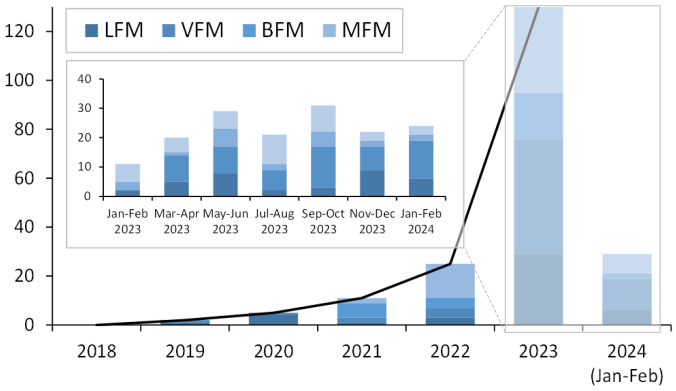

图 2. 2018 年到 2024 年 1-2 月期间医疗健康基础模型的代表性论文数量

型。截至 2024 年 2 月，在四个子领域的代表性论文数量呈指数增长（图2），除了上述典型技术和事件外，一些新兴的范式和技术也在 HFM 中快速发展。

B. 与相关综述的对比以及本文的贡献

通过广泛检索，我们发现了 17 篇与医疗健康基础模型相关的代表性综述，值得注意的是，现有的综述提供了关于 HFM 的不同方面的深刻见解 [10]–[14], [32], [48]–[58]。与这些工作相比，本综述对 HFM 进行了更为全面的概述和分析，包括方法、数据和应用，并对挑战和未来方向进行了深入讨论和展望。具体来说，它具有以下独特优势：**1)** 对 *HFM* “子领域” 的系统分类和研究。本综述涵盖了与 HFM 相关的四个子领域，包括语言、视觉、生物信息和多模态。与现有的综述 [11], [13], [14], [32], [48], [49], [51]–[54] 相比，它提供了更全面的 HFM 领域视角。**2)** 对 *HFM* “方法” 的深入分析。本综述从预训练到适应深入分析了不同子领域中的方法，贯穿了在医疗健康中构建通用 AI 模型的完整过程。与现有的综述 [32], [48], [49], [55], [58] 相比，它提供了 HFM 方法上的系统综述。**3)** 对 *HFM* “属性” 的广泛探讨。本综述并不局限于一些 HFM 的特殊属性，如 “大” [12]，完整介绍了 HFM 的不同特点。与现有的综述 [12], [56] 相比，它提供了对不同属性的 HFM 的广泛视野。**4)** 对 *HFM* 中 “人们关注点” 的全面深入探索。本综述探讨了包括方法、数据、应用、挑战和未来方向在内的全面内容。与现有的综述 [10], [56], [57] 相比，它为 HFM 提供了一个完整的视野，使读者能够更深入地理解。

本综述为当前医疗健康基础模型相关领域提供了深入洞察，因此本文的贡献如下：

1) **方法的系统综述**（第II节）：本综述收录了 2018 年到 2024 年（1-2 月）与 HFM 相关的 200 篇技术论文。我们提出了一种新颖的分类法，并对语言、视觉、生物信息和多模态子领域中的论文进行了预训练和适应的综述，为医疗健康基础模型的潜在技术创新提供了参考。
2) **数据集的全面调查**（第III节）：我们调查了在 HFM 的四个子领域中能够用于模型训练的 114 个大规模数据集/数据库，明确了当前医疗健康数据集的局限性，并为研究人员提供了数据资源指导。
3) **应用的全面概述**（第IV节）：我们概述了当前 HFM 工作中的 16 个潜在医疗应用，展示了 HFM 技术在医疗实践中的当前发展，为未来更多场景中的应用提供了参考。
4) **关键挑战的深入探讨**（第V节）：我们讨论了与数据、算法和计算基础设施相关的关键挑战，指出了 HFM 的当前缺陷，为研究人员进一步探索 HFM 提供了参考。
5) **未来方向的远见探索**（第VI节）：我们展望了 HFM 的未来方向，包括其角色、实施方式、应用和关注点，展示了医疗健康 AI 从传统范式向基础模型时代的转变，指明了潜在的未来发展前景。

II. 方法

图1展示了 HFM 从大规模、多样化的医疗健康数据中学习对信息的表征，然后适应于各种医疗健康应用路线。因此，在本节中，我们从预训练和适应的角度概述了 LFM、VFM、BFM 和 MFM。在本综述中，我们将预训练范式分为以下几类：生成学习（Generative learning, GL）学习对数据的表征，使模型能够从特征中生成有意义的信息；对比学习（Contrastive learning, CL）学习数据的表征，使相似的实例在特征空间中彼此靠近，而不相似的实例远离；混合学习（Hybrid learning, HL）混合不同的学习方法来学习数据的表征；监督学习（Supervised learning, SL）使用带标签的数据训练模型以预测指定的结果。我们将适应范式分为以下几类：微调（Fine-tuning, FT）调整预训练模型内部的参数；适配器微调（Adapter tuning, AT）在预训练模型中添加新的参数（适配器），并仅训练这些额外的参数；提示工程（Prompt engineering, PE）将设计或学习的提示输入到预训练模型中以引导模型执行所需的任务。

A. 用于医疗健康的语言基础模型（*LFM*）

LFM [37], [59] 在医疗健康领域的自然语言处理（NLP）任务中取得了非常显著的进展 [60]–[62]。如表I所示，大多数 LFM 在预训练中采用生成学习方法，并在适应中利用微调和提示工程。

1) 预训练： 在大规模、多样化的医疗文本数据集上的预训练对于 LFM 来说至关重要。它能够使得模型学习对通用特征的表征能力，从而实现对下游任务的迁移。

a) 基于 *GL* 的预训练是 LFM 中最广泛使用的预训练范式，它从大规模的医学语料库中学习生成医学文本，从而获得对语言的表征能力。其中著名的 GL 方法之一是下一 token 预测法（Next token prediction, NTP）[61], [63]–[78]。该方法训练模型通过先前的 token 来预测序列中的下一个 token。其中具有代表性工作是 GatorTronGPT [61]，它混合了医学文本和一般文本，通过 NTP 来预训练一个 GPT（Generative pre-trained transformer）模型，从而在多个医学 NLP 任务上取得了有效的性能。PMC-LLaMA [65] 基于一个已经被预训练过 LLaMA 模型 [79]，构建了一个以数据为中心的知识注入过程，同样通过 NTP 学习并构建了医学领域的语言基础模型。另一个被广泛使用的 GL 方法是掩码语言建模（Masked language modeling, MLM）[38], [62], [80]，它随机掩盖句子中的一部分 token 作为输入数据，并训练模型从这些输入中预测被掩盖的 token。其中，AlphaBERT [81] 和 BEHRT [82] 是基于 MLM 的两种典型的方法，它们同样将医学和一般文本结合起来，在 BERT [37] 架构中使用 MLM 预训练 LFM。

b) 其他预训练范式，类似于 CL 和 HL，研究利用其他方法来学习捕捉医学语言中的结构和关系。MedCPT [60] 是一种代表性的基于 CL 的方法，它利用 PubMed 搜索日志，通过查询-文档对和批内负样本学习对比损失，在六个生物医学任务上取得了 SOTA（State-of-the-art）性能。受 BERT [37] 启发，一些其他方法还融合了下一句子预测法（Next sequence prediction, NSP）方法，即训练网络判断一个句子对是否为相邻的前后关系，与 MLM 一起学习，构造了一种混合的学习方法。一些典型的方法，如 PubMedBERT [62]、BioBERT [38] 和 ClinicalBERT [80]，都构建了类似 BERT 的预训练算法，并通过 MLM 和 NSP [37] 的组合来学习医疗健康 LFM。尽管面临着计算成本和数据质量方面的挑战，但这些方法为医疗健康 LFM 的发展提供了多样化的策略。

2) 适应： 适应方法使用带标签数据或自然语言提示，将通用的 LFM 迁移到具体任务或领域中，从而实现其在医疗健康中的广泛应用。如表I所示，随着基础模型在语言领域的快速发展，医疗健康领域的许多工作都集中在预训练基础语言模型的适应上。大多数 LFM 在适应中使用基于 FT 和 PE 的范式。

a) 基于 *FL* 的适应是一种微调预训练网络内部参数的方法，从而在不增加额外参数的情况下适应下游任务。大量工作 [16], [63]–[65], [67]–[71], [89], [90] 使用全参数 FT 方法，他们使用现有的训练数据集或人类/LLM 生成的指令来调整模型内所有的参数，以提高 LFM 在目标下游任务上的性能。例如，BenTsao [63] 使用来自 CMeKG-8K [103] 中的 8K 个中文指令数据进行微调。HuatuoGPT [64] 使用了 226k 个医学咨询对话和指令数据进行微调。PMC-LLaMA [65] 进一步在医学书籍和论文上使用 LLaMA [79] 模型进行预训练，然后在从医学对话 [71]、医学问答 [104] 和医学知识图谱提示 [105] 中收集的构建指令数据上进行微调。另一种 FT 方法 [72]–[77], [84]–[87] 是参数高效的 FT，该方法只调整模型内部的部分参数，从而在保留预训练模型的一部分表征能力的同时降低了模型适应的成本。一些早期的 LFM [106] 通过微调通用自然语言领域中预训练模型的一部分参数来实现适应。而最近，低秩适应（Low-rank adaptation, LoRA）[107] 作为一种新的参数高效的 FT 方法，在医疗健康 LFM 中取得了成功。它将可训练秩分解矩阵注入到 Transformer 的每一层中，并训练这部分参数，在部署时将这些新参数通过重参数方式融合到模型的原始参数中，大大降低适应时需要训练的参数数量的同时，无需增加额外的参数。包括 Taiyi [73]、GPT-doctor [74] 和 DoctorGLM [75] 在内的许多医疗健康 LFM 都使用了 LoRA 技术，在多个医学语言任务上实现了低成本适应。

b) 基于 *PE* 的适应 [108] 是一种设计高效提示或指令来指导模型预测或调整的方法，由于其强大的任务适应能力，已经在 LFM 中广泛应用。其中一种 PE 方法是手工提示法 [91], [92], [95]–[98]，通过手工构造自然语言提示从而激发通用 LFM 在医疗领域的应用能力。DelD-GPT [95] 使用 ChatGPT 或 GPT-4 作为骨干模型，并采用思维链（Chain of Thought, CoT）[109] 技术生成提示，从而识别医学数据中的敏感信息，如姓名、日期或病灶位置。Dr. Knows [96] 同样使用

表 I

在医疗健康领域的 LFM 研究。这里的缩写是 CL：对比学习（CONTRASTIVE LEARNING），GL：生成学习（GENERATIVE LEARNING），HL：混合学习（HYBRID LEARNING），FT：微调（FINE-TUNING），PE：提示工程（PROMPT ENGINEERING），IR：信息检索（INFORMATION RETRIEVAL），NER：命名实体识别（NAMED ENTITY RECOGNITION），RE：关系提取（RELATION EXTRACTION），QA：问答，VQA：视觉问答（VISUAL QUESTION ANSWERING），DIAL：对话（DIALOGUE），NLI：自然语言推理（NATURAL LANGUAGE INFERENCE），TC：文本分类（TEXT CLASSIFICATION），STS：文本语义相似性（SEMANTIC TEXTUAL SIMILARITY），SUM：摘要提取（SUMMARIZATION），REC：推荐（RECOMMENDATION），CLS：图像分类（IMAGE CLASSIFICATION），RG：报告生成（REPORT GENERATION），SEG：图像分割（IMAGE SEGMENTATION）。

方法	预训练	适应	骨干网路	下游任务	年份	代码
MedCPT [60]	CL	FT	PubMedBERT	IR	2023	✓
AlphaBERT [81]	GL	FT	BERT	NER, RE, QA	2020	✓
BEHRT [82]	GL	FT	BERT	NER, RE, QA	2020	✓
BioBART [83]	GL	FT	BART	NER, RE, QA	2020	✓
PMC-LLaMA [65]	GL	FT	LLaMA	QA	2023	✓
BioMistral [78]	GL	FT	Mistral	QA	2024	✓
Zhongjing [84]	GL	FT	Ziya-LLaMA	QA, DIAL	2024	✓
Me LLaMA [85]	GL	FT	LLaMA2	NER, RE, QA, NLI, SUM, CLS	2024	✓
OncoGPT [86]	GL	FT	LLaMA	DIAL	2024	✓
JMLR [87]	GL	FT	LLaMA-2	QA	2024	
MEDITRON-70B [67]	GL	FT	LLaMA-2	QA	2023	✓
Qilin-Med [72]	GL	FT	Baichuan	QA	2023	✓
HuatuoGPT-II [66]	GL	FT	Baichuan2	QA	2023	✓
ANTPLM-Med-10B [77]	GL	FT	AntGLM	QA	2023	
GatorTronGPT [61]	GL	FT	Transformer	NER, RE, QA, NLI, STS	2023	✓
BioBERT [38]	HL	FT	BERT	NER, RE, QA	2019	✓
PubMedBERT [62]	HL	FT	BERT	NER, RE, QA, STS	2021	✓
ClinicalBERT [80]	HL	FT	BERT	NLI	2019	✓
GatorTron [15]	HL	FT	Transformer	NER, RE, QA, NLI, STS	2022	✓
BenTsao [63]	-	FT	LLaMA	QA	2023	✓
ChatDoctor [70]	-	FT	LLaMA	QA	2023	✓
MedAlpaca [71]	-	FT	LLaMA	QA	2023	✓
Alpacare [68]	-	FT	LLaMA/LLaMA-2	QA	2023	✓
MedPaLM [88]	-	FT	PaLM	QA	2023	
MedPaLM 2 [89]	-	FT	PaLM-2	QA	2023	
HuatuoGPT [64]	-	FT	Baichuan	QA, DIAL	2023	✓
GPT-Doctor [74]	-	FT	Baichuan2	DIAL	2023	
DoctorGLM [75]	-	FT	ChatGLM	QA	2023	✓
Bianque [69]	-	FT	ChatGLM	QA	2023	✓
Taiyi [73]	-	FT	Qwen	NER, RE, TC, QA	2023	✓
BiMediX [90]	-	FT	Mistral	QA	2024	✓
ClinicalGPT [76]	-	FT	BLOOM	QA, DIAL	2023	
Visual Med-Alpaca [91]	-	FT, PE	LLaMA	VQA	2023	✓
OphGLM [92]	-	FT, PE	ChatGLM	CLS, SEG	2023	✓
ChatCAD [93]	-	PE	ChatGPT	CLS, RG	2023	✓
ChatCAD+ [94]	-	PE	ChatGPT	CLS, RG	2023	✓
DeID-GPT [95]	-	PE	ChatGPT	NER	2023	✓
Dr.Knows [96]	-	PE	ChatGPT	TC, SUM	2023	
Medprompt [97]	-	PE	ChatGPT-4	QA	2023	✓
HealthPrompt [98]	-	PE	ChatGPT	TC	2022	
MedAgents [99]	-	PE	ChatGPT / Flan-PaLM	QA	2023	✓
SPT [100]	-	PE	MedRoBERTa.nl	TC	2023	✓
PBP [101]	-	PE	SciBERT	TC	2022	
NapSS [102]	-	PE	GPT-2	REC	2023	✓

ChatGPT 作为骨干模型，并利用零样本提示技术生成能够回答医学问题并根据患者的症状和情况进行自动诊断的提示。Medprompt [97] 使用 GPT-4 作为骨干模型，结合 CoT 和集成提示技术生成提示，从而执行多

种医疗任务。HealthPrompt [98] 使用六种不同的预训练 LFM 作为骨干模型，并采用手动模板零样本方法生成提示，从而实现医学文本分类的功能。Visual Med-Alpaca [91] 和 OphGLM [92] 将 LFM 与专用的视觉模型集成在一起，从而在不产生视觉-语言基础模型开发成本的前提下，实现了超越语言模态的医学任务。另一种 PE 方法是可学习的提示 [100]，它利用可学习的软提示来学习特定任务的自然语言提示参数，已有一些工作 [59], [101], [102], [110], [111] 使用该方法实现了医学文本分类。PBP [101] 使用 SciBERT [110] 作为骨干模型，并学习可以分类医学文本的自然语言提示。NapSS [102] 使用 GPT-2 [59] 作为骨干模型，利用可学习提示学习可以为临床场景生成个性化推荐的自然语言提示。MedRoBERTa.nl [111] 同样使用软提示来实现医学文本分类。

B. 用于医疗健康的视觉基础模型（VFM）

基础模型的革命性影响也来到了视觉领域，VFM 通过广泛的学习获得了强大的通用能力，从而在多种下游医疗任务中获得了出色的表现 [112]。如表II所示，它们在广泛的带标签或无标签医疗数据集上进行预训练，从而适应众多的下游任务。

1) 预训练: 与语言不同，视觉信息的连续性使得解耦视觉上下文中的语义变得非常困难 [113]。因此，除了自监督学习（Self-supervised learning, SSL）外，VFM [52], [57], [114] 还利用监督学习（SL）进行特定任务的预训练。

a) 基于 *SL* 的预训练范式利用标签来解耦医学图像内部的语义信息，从而学习在特定任务的广泛适用的模型。Med3d [115] 利用 8 个三维医学图像分割数据集来预训练了一个 ResNet，用于下游任务的迁移学习。最近，大多数方法旨在通过大规模监督学习某些特定任务（如分割），从而获得在该任务上具有通用能力的模型。一个典型的工作是 STU-Net [116]，它在 TotalSegmentator 数据集 [117] 上进行大规模预训练，从而实现了人体内 104 个器官的通用分割。由于标注分割标签的高成本性，一些工作通过混合了多个公共带标签数据集来构建大规模监督学习数据集。UniverSeg [118] 在 53 个公开医学图像分割数据集上进行训练，从而实现了通用的医学图像分割能力。最近，SAM [20] 显著推动了交互式医学图像分割基础模型的发展，在医学图像上也展现出了巨大的潜力。它的一些变体，如 SAM-Med3D [119] 和 SAM-Mad2D [120]，进一步将 SAM 的预训练参数迁移到医学图像上，并在由多个公共医学图像分割数据集所混合的大规模数据集上进行微调，实现医学图像的通用交互式分割。虽然这些特定任务的 VFM 已经展示了出色的性能，但高昂的标注成本使得构建大规模预训练数据集变得极其具有挑战性。大多数现有的监督预训练工作仍然只在缺乏任务多样性的医学图像分割任务上进行探索。

由于医学图像标注的高成本，自监督预训练（Self-supervised pre-training, SSP）[40], [121] 已成为在 VFM 中被广泛研究的范式。它通过学习一个代理任务，从而在不需要手工标注的前提下，在大规模数据上进行学习，使模型学习到通用的特征表示。因此，它为构建能够适应到不同下游医学图像任务的 VFM 创造了一种有潜力的范式，有望进一步推动模型多种医疗场景中进行广泛应用。

b) 基于 *GL* 的预训练通过预测或重构原始输入来学习通用视觉表示，已经在医学图像上被广泛研究，包括 RETFound [19]、VisionFM [122]、SegVol [123]、DeblurringMAE [124]、USFM [125] 和 Models Genesis [41], [126]。一种常用的代理任务是掩码图像建模（Masked image modeling, MIM）[127]–[129]。它与语言领域中的 MLM 类似，采用编码器-解码器架构来编码损坏的图像并解码预测被损坏的图像，从而学习图像的上下文依赖。例如，RETFound [19] 和 VisionFM [122] 是基于 MIM 开发的用于视网膜图像和眼科临床任务的模型。SegVol [123] 同样基础 MIM，实现了对三维医学图像的自监督预训练。DeblurringMAE [124] 将去模糊任务引入到预训练中，而 USFM [125] 提出了一种空间频率双遮蔽 MIM 方法。Models Genesis [41], [126] 通过多种方式破坏图像，并将图像恢复作为代理任务，使预训练模型能够有效地捕捉了细粒度的视觉信息。

c) 基于 *CL* 的预训练通过对比图像之间的相似性或差异性来学习具有区分性的视觉表示。随着 CL 在自然图像中的成功，一些工作也被应用到了医学图像上，他们利用 MoCo [113] 或 SimCLR [130] 算法，并且在病理学 [131], [132] 和 X 射线 [133] 图像中取得了成功。C2L [134] 通过构建同质和异质数据对，同学学习对比来自不同的图像表示来学习通用和稳健的视觉表征。Endo-FM [18] 通过空间-时间匹配在多种视频视图上使用教师-学生模型进行预训练。他们通过教师和

学生模型来编码视频序列，并在特征空间进行相互预测。LVM-Med [135] 通过二阶图匹配法，在由 55 个公共数据集混合而成的包含 130 万张图像的大数据集上进行预训练，从而获得了涵盖大量器官和模态表征能力的 VFM。Wu 等人 [136] 提出了一个简单而有效的 VoCo 框架，利用上下文位置先验进行 CL 预训练。此外，MIS-FM [137] 引入了基于伪分割的代理任务，他们通过拼贴来自不同图像的块来生成成对的图像和伪分割标签以预训练 3D 分割模型。Ghesu 等人 [138] 提出了一种基于 CL 和在线特征聚类的自监督学习方法。

d) 基于 *HL* 的预训练将多种预训练方法结合起来，通过联合训练的方式融合多种学习的优点。Virchow [139]、UNI [140] 和 RudolfV [141] 利用了 DINOv2 [142] 训练范式，融合了 MIM 和 CL 实现了强大而有效的预训练。BROW [143] 在自我蒸馏框架中融合了颜色增强、补丁洗牌、MIM 和多尺度输入等学习方式，以预训练基础视觉模型。TransVW [144] 混合了自分类和自恢复学习，从多个信息源中学习表征。GVSL [40] 通过配准学习以训练模型表征医学图像之间的相似性，并通过自恢复学习来获得图像内容的上下文依赖。

2) 适应： 在预训练之后，VFM 进一步构建适应方法从而将基础模型应用到各种下游任务中。除了经典的微调（FT）方法外，最近的一些新方法也被广泛地应用到基础模型中，包括适配器微调（AT）和提示工程（PE）。

a) 基于 *FT* 的适应范式通过优化预训练 VFM 内部的参数，以适应下游任务。一些工作对预训练 VFM 的所有参数进行微调 [40], [42], [148], [151]–[153]，从而在特定任务上表现出明显的性能提升。这些研究更接近于数据驱动的模型参数初始化方法，利用预训练权重作为更好的初始化参数来学习目标任务。然而，这些方法不仅耗时，当数据不足时还容易发生过拟合问题。一些其他的工作通过仅微调模型内部的部分参数 [149], [150], [154]–[157]，从而减少微调成本，提高计算效率，并且能够保持预训练权重中的通用表示。但是，微调的参数位置需要通过手动的设置，由于不同位置参数所表征的信息差异，手动的选择也限制了模型的适应性。因此，最近，受 LFM 中基于 LoRA 的自适应方法 [107] 的启发，VFM 也利用低秩的方法来低成本且有效地适应预训练模型到下游任务。例如，一些 VFM [154], [156], [161] 在保持预训练的 SAM 参数的情况下，采用 LoRA 进行参数高效的微调适应。

b) 基于 *AT* 的适应范式将一些适配器添加到预训练的 VFM 中，并仅优化这些适配器以适应下游任务。与 FT 不同，它不会改变原始参数，从而保留了 VFM 从大规模数据中学习的通用表示。一种早期的称为“线性评估”的方法被广泛用于评估预训练的骨干网路的泛化能力 [40], [41], [126], [133]。它在骨干网路的末端添加一个线性层作为适配器，并在自适应时优化该层，从而评估预训练模型的表示能力。最近，为了获得更好的迁移性能，适配器被进一步添加到网络的内部层中。大量基于 SAM 的实践 [148]–[150], [158]–[160], [163]–[167] 已经证明了 AT 在医学图像分割方面的出色性能。它们通过训练适配器中非常少的参数，在保持 SAM 从大规模自然图像中学习的图像分割能力的同时，有效地将其迁移到医学图像领域。All-in-SAM [174] 进一步构建了一种基于弱监督的适应方法，通过 SAM 构建伪标签，然后采用类似于 SAM-adapter [185] 的方式利用 AT 进行适应 SAM。MA-SAM [147] 进一步将 3D 适配器嵌入到原始的 2D SAM 模型中，构建了一个用于 3D 医学图像的 SAM。

c) 基于 *PE* 的适应范式 [108] 也在 VFM 上取得了强大的适应性能。继 SAM 的成功后，许多基于 SAM 的 VFM 在医疗健康领域 [42], [119], [157], [162], [169], [170], [172] 也利用点、边界框或文本作为医学图像分割的提示，实现交互式分割。Baharoon 等人 [44] 在 DINO v2 上研究了适用于医学图像的提示模板。此外，少样本提示利用少量的图像-标签对作为提示数据，也被用于提示工程。UniverSeg [118] 利用支持集作为提示，从而实现对查询图像上任意目标的分割。Anand 等人 [181] 提出了一个一样本定位和分割框架，利用与模板图像的对应关系来提示 SAM。一些 VFM 研究 [148], [169], [170], [184] 进一步设计了自动提示生成的方法。如 AutoSAM [169] 构造了一个辅助提示编码器，通过输入图像的特征生成一个代理提示，从而驱动 SAM 的分割，消除了手动提示的交互过程。PUNETR [184] 进一步研究了提示微调方法，它将一些可学习的提示嵌入到预训练网络中，通过学习这些提示以适应不同的医学图像任务。

C. 用于医疗健康的生物信息基础模型（*BFM*）

基础模型在生物信息领域也在迅速发展 [53]，正如最近的综述 [53] 所讨论的，随着高通量测序技术的发展 [214]，现有的 BFM 已经在组学数据中取得了显著的成

表 II

在医疗健康领域的 VFM 研究。这里的缩写是 SL：监督学习（SUPERVISED LEARNING），GL：生成学习（GENERATIVE LEARNING），CL：对比学习（CONTRASTIVE LEARNING），HL：混合学习（HYBRID LEARNING），FT：微调（FINE-TUNING），PE：提示工程（PROMPT ENGINEERING），AT：适配器微调（ADAPTER TUNING），CLS：分类（CLASSIFICATION），SEG：分割（SEGMENTATION），DET：检测（DETECTION），PR：预后（PROGNOSIS），RET：检索（RETRIEVAL），以及 IE：图像增强（IMAGE ENHANCEMENT）。

方法	与训练	适应	骨干网路	模态	下游任务	年份	代码
Med3D [115]	SL	FT	ResNet	CT, MRI	SEG, CLS	2019	✓
STU-Net [116]	SL	FT, PE	nnU-Net	CT	SEG	2023	✓
UniverSeg [118]	SL	PE	U-Net	Multimodal images	SEG	2023	✓
SAM-Med3D [119]	SL	PE	ViT (SAM)	Multimodal images	SEG	2023	✓
RETFound [19]	GL	FT	ViT (MAE)	CFP, OCT	CLS, PR, DET	2023	
VisionFM [122]	GL	FT	-	Multimodal images	CLS	2023	
SegVol [123]	GL	FT, PE	ViT (SAM)	CT	SEG	2023	✓
Models Genesis [41], [126]	GL	FT, AT	U-Net	CT, X-ray	CLS, SEG	2019	✓
DeblurringMAE [124]	GL	FT, AT	ViT (MAE)	US	CLS	2023	✓
USFM [125]	GL	AT	-	US	SEG, CLS, IE	2024	
C2L [134]	CL	FT	ResNet/DenseNet	X-ray	CLS	2020	✓
Endo-FM [18]	CL	FT	ViT	Endoscopy	SEG, CLS, DET	2023	✓
Ciga *et al.* [132]	CL	FT	ResNet (SimCLR)	Pathology	CLS, SEG	2022	✓
CTransPath [131]	CL	FT	ViT (MoCo v3)	Pathology	RET, CLS	2022	✓
LVM-Med [135]	CL	FT	ResNet, ViT	Multimodal images	SEG, CLS, DET	2024	✓
MIS-FM [137]	CL	FT	Swin	CT	SEG	2023	✓
VoCo [136]	CL	FT	Swin	CT	SEG, CLS	2024	✓
MoCo-CXR [133]	CL	FT, AT	ResNet, DenseNet	X-ray	CLS	2021	
TransVW [144]	HL	FT	U-Net	CT, X-ray	CLS, SEG	2021	
Ghesu *et al.* [138]	HL	FT	ResNet	X-ray, CT, MRI, US	DET, SEG	2022	
UNI [140]	HL	FT	ViT (DINOv2)	Pathology	CLS, SEG	2024	✓
BROW [143]	HL	FT	ViT	Pathology	CLS, SEG	2023	
Campanella *et al.* [145]	HL	FT	ViT (MAE, DINO)	Pathology	CLS	2023	
RudolfV [141]	HL	FT	ViT (DINOv2)	Pathology	CLS	2024	
Swin UNETR [146]	HL	FT	Swin	CT	SEG	2022	✓
GVSL [40]	HL	FT, AT	U-Net	CT, MRI	SEG, CLS	2023	✓
Virchow [139]	HL	AT	ViT (DINOv2)	Pathology	CLS	2023	
MA-SAM [147]	-	FT, AT, PE	ViT (SAM)	CT, MRI, Endoscopy	SEG	2023	✓
Pancy *et al.* [148]	-	FT, AT, PE	YOLOv8, ViT (SAM)	Multimodal images	SEG	2023	
3DSAM-adapter [149]	-	FT, AT, PE	ViT (SAM)	CT	SEG	2023	✓
SP-SAM [150]	-	FT, AT, PE	ViT (SAM)	Endoscopy	SEG	2023	✓
Baharoon *et al.* [44]	-	FT, AT, PE	ViT (DINOv2)	X-ray, CT, MRI	SEG, CLS	2023	✓
MedSAM [42]	-	FT, PE	ViT (SAM)	Multimodal images	SEG	2023	✓
Skinsam [151]	-	FT, PE	ViT (SAM)	Dermoscopy	SEG	2023	
Polyp-SAM [152]	-	FT, PE	ViT (SAM)	Endoscopy	SEG	2023	✓
SAM-OCTA [153]	-	FT, PE	ViT (SAM)	OCT	SEG	2023	✓
SAMed [154]	-	FT, PE	ViT (SAM)	CT	SEG	2023	✓
SAM-LST [155]	-	FT, PE	ViT (SAM)	CT	SEG	2023	✓
Feng *et al.* [156]	-	FT, PE	ViT (SAM)	CT, MRI	SEG	2023	
SemiSAM [157]	-	FT, PE	ViT (SAM)	MRI	SEG	2023	
AFTer-SAM [158]	-	AT, PE	ViT (SAM)	CT	SEG	2024	
Mammo-SAM [159]	-	AT, PE	ViT (SAM)	CT	SEG	2023	
ProMISe [160]	-	AT, PE	ViT (SAM)	CT	SEG	2023	✓
Med-SA [161]	-	AT, PE	ViT (SAM)	Multimodal images	SEG	2023	✓
SAM-Med2D [162]	-	AT, PE	ViT (SAM)	Multimodal images	SEG	2023	✓
Adaptivesam [163]	-	AT, PE	ViT (SAM)	Multimodal images	SEG	2024	✓
MediViSTA-SAM [164]	-	AT, PE	ViT (SAM)	US	SEG	2023	✓
SAMUS [165]	-	AT, PE	ViT (SAM)	US	SEG	2023	
SegmentAnyBone [166]	-	AT, PE	ViT (SAM)	MRI	SEG	2024	✓
Swinsam [167]	-	AT, PE	ViT (SAM)	Endoscopy	SEG	2024	
SAMAug [168]	-	PE	ViT (SAM)	Multimodal images	SEG	2023	✓

下一页继续。

接上一页。

AutoSAM [169]	-	PE	ViT (SAM)	Multimodal images	SEG	2023	
DeSAM [170]	-	PE	ViT (SAM)	Multimodal images	SEG	2023	✓
CellSAM [171]	-	PE	ViT (SAM)	Multimodal images	SEG	2023	✓
Sam-u [172]	-	PE	ViT (SAM)	Fundus	SEG	2023	
Sam-path [173]	-	PE	ViT (SAM)	Pathology	SEG	2023	
All-in-sam [174]	-	PE	ViT (SAM)	Pathology	SEG	2023	
SurgicalSAM [175]	-	PE	ViT (SAM)	Endoscopy	SEG	2024	✓
Polyp-SAM++ [176]	-	PE	ViT (SAM)	Endoscopy	SEG	2023	✓
UR-SAM [177]	-	PE	ViT (SAM)	CT	SEG	2023	
MedLSAM [178]	-	PE	ViT (SAM)	CT	SEG	2023	✓
nnSAM [179]	-	PE	ViT (SAM)	CT	SEG	2023	✓
EviPrompt [180]	-	PE	ViT (SAM)	CT, MRI	SEG	2023	
Anand *et al.* [181]	-	PE	ViT (SAM)	CT, MRI, US	SEG	2023	
SAMM [182]	-	PE	ViT (SAM)	CT, MRI, US	SEG	2023	✓
SAMPOT [183]	-	PE	ViT (SAM)	X-ray	SEG	2023	
PUNETR [184]	-	PE	-	CT	SEG	2024	✓

果，包括单细胞 RNA 测序（single-cell RNA sequencing, scRNA-seq）[188]、DNA [191]、RNA [193] 和蛋白质数据 [25]。如表II-C所示，BFM 中的方法受到 LFM 的大量启发，大多数方法都按照 LFM 中的基本架构构建的，例如 BERT 和 Transformer 解码器。它们同样采用了在 VFM 和 LFM 中被广泛使用的预训练和自适应范式，以获得生物信息任务的通用能力。

1) 预训练: 受 LFM 的启发，BFM 中大量的预训练策略也基于 GL 范式（表II-C），拥有捕捉特征上下文依赖的强大能力，这对于理解生物系统至关重要。

a) 基于 *GL* 的预训练范式训练 BFM 学习对特征的上下文依赖的表征，从而使得模型能够自我发现组学特征之间的潜在关系。

与 LFM 和 VFM 中的 GL 方法类似，掩码组学建模（Masked omics modeling, MOM）[188] 和下一 token 预测 (NTP) 是 BFM 中最受欢迎的 GL 预训练任务。MOM 随机遮盖生物数据中的表达值或序列，并训练模型重构被遮盖的信息。对于单细胞 RNA 测序数据，一些 BFM 工作 [188], [200], [201], [203] 利用 MOM 对表达值和基因名称进行编码，从而从高维稀疏数据中提取代表性信息。例如，scBERT [188] 将 BERT 中的 MLM 从 LFM 应用到 BFM 作为 MOM，并在 110 万个人类单细胞 RNA 测序数据上训练其模型以有效表示表达值。此外，对于 DNA 和 RNA 数据，MOM 还学习了核苷酸内部序列之间的依赖关系，从而建模基因之间的关系 [190]–[194]。例如，基于 BERT [37]，GENA-LM [191] 和 SpliceBERT [193] 分别通过 MOM 预训练了对人类 DNA 序列和 RNA 序列的表征，从而在它们的目标下游任务上实现了强大的迁移能力。在一些蛋白质预训练工作中 [196], [197]，MOM 也通过学习重构蛋白质序列或结构而取得了成功。BFM 中的 NTP 学习基于先前标记或序列预测下一个标记或序列，在序列数据（如 DNA、RNA 和蛋白质序列）上取得了巨大成功 [46], [186], [187], [199]。HyenaDNA [199] 利用单个核苷酸作为标记，并在每个层引入全局上下文以预测下一个核苷酸。DNABERT [46] 使用 3-mer 到 6-mer token 共计 2.75 亿个核苷酸碱基上训练了四个模型。然而，由于该方法对数据的序列特性要求，在 seRNA-seq 数据中仍未得到研究。

除了最初在 LFM 中设计的 MLM 和 NTP 之外，一些工作 [25], [205], [206] 基于生物数据的特性构建了新的基于 GL 的预训练任务。UTR-LM [206] 提出了一个二级结构和最小自由能预测预训练任务，实现了高效的 RNA 数据预训练。Alphafold [25] 还将自蒸馏学习与掩码多序列对齐学习相结合，利用无标签的蛋白质序列实现下游高精度蛋白质结构预测任务。

b) 其他预训练范式也已经被研究用于表征生物信息 [204], [207]–[210]。受到 GPT [215] 的启发，scGPT [203] 将 MOM 与 NTP 相结合，构建了一个单细胞多组学基础模型。RNABERT [204] 设计了一个结构对齐学习，学习两个 RNA 序列之间的关系，以获得同一列中的碱基的更紧密的嵌入。DNAGPT [202] 利用三个预训练任务，包括 NTP、鸟嘌呤-胞嘧啶含量预测和序列顺序预测，预训练了 DNA 序列的表示。GeneBERT [208] 遵循 BERT 中的基本学习范式，将两个任务（MOM 和 NSP）结合在一起，构建了一个 DNA 基础模型。CellLM [207] 将 MOM 与细胞类型鉴别和 CL 任务相结合以实施模型的预训练。CodonBERT [209] 构建了一种同源序

表 III
在医疗健康领域的 BFM 研究。这里的缩写是 CL：对比学习（CONTRASTIVE LEARNING），GL：生成学习（GENERATIVE LEARNING），HL：混合学习（HYBRID LEARNING），FT：微调（FINE-TUNING），AT：适配器微调（ADAPTER TUNING），PE：提示工程（PROMPT ENGINEERING），SA：序列分析（SEQUENCE ANALYSIS），IA：交互分析（INTERACTION ANALYSIS），SFA：结构和功能分析（STRUCTURE AND FUNCTION ANALYSIS），以及 DR：疾病研究和药物反应（DISEASE RESEARCH AND DRUG RESPONSE）

方法	预训练	适应	骨干网路	模态	下游任务	年份	代码
ProGen [186]	GL	FT	Transformer Decoder	Protein	SA	2023	✓
ProGen2 [187]	GL	FT	Transformer Decoder	Protein	SA, SFA	2023	✓
scBERT [188]	GL	FT	BERT	scRNA-seq	SA	2022	✓
Geneformer [189]	GL	FT	BERT	scRNA-seq	IA, DR	2023	✓
DNABERT [46]	GL	FT	BERT	DNA	SA, SFA	2021	✓
DNABERT-2 [190]	GL	FT	BERT	DNA	SA, DR	2023	✓
Nucleotide Transformer [23]	GL	FT	BERT	DNA	SA, SFA	2023	✓
Gena-LM [191]	GL	FT	BERT	DNA	SA	2023	✓
RNA-FM [26]	GL	FT	BERT	RNA	SA, IA, SFA	2022	✓
RNA-MSM [192]	GL	FT	BERT	RNA	SFA	2024	✓
SpliceBERT [193]	GL	FT	BERT	RNA	SA, SFA	2023	✓
3UTRBERT [194]	GL	FT	BERT	RNA	SA	2023	✓
ESM-2 [195]	GL	FT	BERT	Protein	SFA	2023	✓
ProtTrans [196]	GL	FT	BERT	Protein	SFA	2021	✓
MSA Transformer [197]	GL	FT	BERT	Protein	SFA	2021	✓
ESM-1b [198]	GL	FT	BERT	Protein	SFA	2021	✓
AlphaFold [25]	GL	FT	Evoformer	Protein	SFA	2021	✓
HyenaDNA [199]	GL	FT, PE	Transformer Decoder	DNA	SA	2023	✓
scFoundation [200]	GL	PE, AT	Asymmetric Encoder-decoder	scRNA-seq	DR	2023	✓
UCE [201]	GL	AT	BERT	scRNA-seq	SFA	2023	✓
DNAGPT [202]	HL	FT	Transformer Decoder	DNA	SA	2023	✓
scGPT [203]	HL	FT	Transformer Decoder	scRNA-seq	IA, SFA	2023	✓
RNABERT [204]	HL	FT	BERT	RNA	SA, SFA	2022	✓
AminoBERT (RGN2) [205]	HL	FT	BERT	Protein	SFA	2022	✓
UTR-LM [206]	HL	FT	BERT	RNA	SA, SFA	2023	✓
CellLM [207]	HL	FT	BERT	scRNA-seq	SFA, DR	2023	✓
GeneBERT [208]	HL	FT	BERT	DNA	SA, DR	2021	✓
CodonBERT [209]	HL	FT	BERT	RNA	SFA	2023	✓
xTrimoPGLM [210]	HL	FT, AT	GLM [211]	Protein	SFA	2023	
GenePT [212]	-	PE	GPT	DNA	IA, DR	2023	✓
scELMo [213]	-	PE	GPT	scRNA-seq	SFA, DR	2023	✓

列预测方法，直接建模序列表示，并理解 mRNA 序列之间的进化关系。xTrimoPGLM [210] 通过使用 MOM 和通用语言模型（general language model, GLM）[211] 预训练任务，在 1 万亿个标记中训练了一个具有超过 1000 亿参数的模型，成为当前最大的蛋白质基础模型。

2) 适应： 如表II-C所示，BFM 进一步在如功能分析、序列分析等下游任务中适应预训练模型，从而将其应用于特定的生物信息应用。

a) 基于 *FT* 的适应是 BFM 中最广泛使用的范式。与 LFM 和 VFM 一样，在 BFM 中使用的 FT 方法调整预训练模型的内部参数以适应不同的特定下游任务。这些工作 [23], [25], [26], [186]–[189], [191]–[198], [202]–[209] 直接调整网络的所有参数以适应特定的下游任务，从而评估预训练模型的泛化能力和其在生物信息中的应用潜力。例如，scBERT [188] 利用预训练模型未见过的和用户特定的单细胞 RNA 测序数据，将模型微调到的 9 个细胞类型注释任务，在多种基准测试中超过了现有的先进方法。BFM 领域中也研究了参数高效的 FT，试图设计更大的模型并利用 LoRA 来调整部分参数 [46], [210]，从而实现高效的适应性。DNABERT-2 [46] 就是这样一种模型，它引入了 LoRA，并在几乎不损失性能的情况下显著降低了计算和内存成本。

b) 其他适应范式，包括 AT 和 PE，也已经在 BFM 中得到了应用。由于 BFM 在预训练期间学习了大规模的生物信息，因此基于 AT 的方法通过在特定位置添加并训练少量参数来有效降低适应的计算成本 [25], [200],

[201]。其中具有代表性的工作是 xTrimoPGLM [25] 和 UCE [201]，它们在预训练的骨干网络后添加和训练额外的 MLP 层，从而使模型适应下游任务。该方法提供了一种新的编码生物数据的方式，因此利用预训练模型所提取的特征，只需训练一个简单的分类器，就可以在多种任务上获得良好的性能。在 BFM 中，基于 PE 的适应范式仍然研究较少，只有少数几个工作 [199], [200], [212], [213] 尝试利用 PE 方法实现对下游任务的适应。例如，GenePT [212] 探索了一种 PE 的简单的方法，它利用基于文献从 ChatGPT 中获得对基因的提示嵌入，并利用了零样本的方法来预测潜在的基因功能。

D. 用于医疗健康的多模态基础模型（MFM）

医疗健康数据本身天然是多模态的（图1），因此在医疗实践中将语言、视觉、生物信息等多种模态数据集成起来，构建多模态基础模型（MFM）具有很大的潜力。与单模态模型不同，MFM 能够理解每种模态内在的特征以及它们之间的相互关系，从而增强基础模型处理医疗复杂场景的能力。由于模态和它们之间的组合类型的多样性，MFM 的预训练和适应具有其独特的设计。

1) 预训练: MFM 涉及多种模态的学习，因此在 LFM、VFM 和 BFM 中对于不同模态的典型学习范式也被广泛应用与 MFM。然而，多模态预训练需要模型理解和融合来自多种模态的信息，因此也存在着一些新的独有挑战。

a) 基于 *GL* 的预训练范式通过引导网络预测或重构图像、文本或其他类型的数据，使模型获得更强的生成能力。类似于 MLM，MIM，MOM 的掩码表征建模被单独使用或在不同模态之间进行组合，进行模型预训练。例如，MMBERT [216] 将图像特征融合到一个 BERT 架构中，利用 MLM 训练模型对医学图像和文本的理解能力。MRM [217] 通过结合 MLM 和 MIM 进一步提高了视觉表示。这些生成式预训练提供了一种直接的方法来促进跨模态交互，使得模型可以基于更通用的多模态表示来恢复被遮蔽的多模态数据。随着 MFM 的发展，一大趋势是研发更具有通用性的 AI 模型，这些模型在更大、更多样化的多模态数据集上进行训练，从而以在单一架构内处理多个任务。在这些方法中，RadFM [27] 针对放射学数据训练了一个视觉条件自回归语言生成模型，实现了大量的医学任务。BiomedGPT [29] 采用单模态表示建模和任务特定的多模态学习，预训练了一个统一的序列到序列模型。

b) 基于 *CL* 的预训练范式利用对比损失来学习多模态数据，从而增强 MFM 的跨模态理解能力。例如 CLIP [47] 用于图像和文本之间的对比学习，实现不同模态之间的特征对齐。在 CLIP 之前，ConVIRT [218] 在胸部 X 光和肌肉骨骼图像中开创了视觉-语言的对比学习，而 ETP [224]、MI-Zero [236]、BiomedCLIP [230] 和 MoleculeSTM [233] 进一步将这种策略扩展到心电图信号、病理图像、生物医学图像和分子结构信息中，展示了 CL 范式在医学领域的有效性。此外，一些后续研究尝试通过改进训练策略来扩展视觉-语言对齐预训练。考虑到关键的语义信息可能集中在医学数据的特定区域，GLoRIA [225]、LoVT [219]、MGCA [222] 和 IMITATE [226] 专注于探索不同图像子区域和文本标记嵌入之间的细粒度语义对齐，展示了细粒度对齐在捕捉微妙的语义信息方面的有效性。为了增强预训练数据的使用效率，MedCLIP [227] 将预训练扩展到包括大型不成对的图像和文本，以组合方式扩展训练数据的数量。CXR-CLIP [229] 和 UCML [231] 采用提示模板从图像标签数据集中生成图像-文本对。考虑到医学语言的专业性质，MedKLIP [223]、KAD [232]、CLIP-Lung [234] 和 UniBrain [220] 进一步融合了医学数据集中的领域特定知识。BioBRIDGE [271] 利用知识图谱学习一个单模态基础模型到另一个单模态基础模型的迁移，而不需要微调任何单模态基础模型即可实现多模态功能。总的来说，CL 方法使模型能够更好地理解复杂的关系，而不需要特定任务的微调。

c) 基于 *HL* 的预训练范式也被用于融合不同的学习范式的优点并激发模型的学习能力。Clinical-BERT [237]、Li 等人 [272]、M^3AE [238]、MedViLL [239] 和 ARL [241] 利用了掩码表征建模和图像-文本匹配学习的组合进行模型预训练。PMC-CLIP [240]、PIROR [247]、MaCo [242] 和 BioViL [246] 利用了掩码表征建模和对比学习。MUMC [243] 同时融合了上述的三种学习方法。T3D [244] 在两个以文本为驱动的预设任务上进行了预训练，即文本驱动的图像恢复和文本驱动的对比学习。CONCH [248] 联合了图像-文本对比损失和字幕预测损失 [273] 进行学习。GIMP [245] 设计了一个 MIM 和基因诱导的三元组学习。ProteinDT [249] 结合了对比学习、自回归和扩散生成范式。这些方法涉及将模态间和内模态的生成或对比任务相结合，以相互增强彼此的有效性。

表 IV

在医疗健康领域的 VFM 研究。这里的缩写是 GL：生成学习（generative learning），CL：对比学习（contrastive learning），HL：混合学习（hybrid learning），FT：微调（fine-tuning），AT：适配器微调（adapter tuning），PE：提示工程（prompt engineering），CLS：分类（classification），DET：检测（detection），SEG：分割（segmentation），RG：报告生成（reports generation），VQA：视觉问答（visual question answering），CMR：跨模态检索（cross-modal retrieval），CMG：跨模态生成（cross-modal generation），PG：短语定位（phrase-grounding），NLI：自然语言推理（Natural language inference），PPP：蛋白质属性预测（protein property prediction），TS：文本摘要（text summarization），GVC：基因组变异检测（genomic variant calling），GMG：凝视图生成（gaze map generation），以及 MPP：分子属性预测（molecular property prediction）

方法	预训练	适应	骨干网路	模态	下游任务	年份	代码
MMBERT [216]	GL	FT	ResNet+BERT	Multimodal images, Text	VQA	2021	✓
MRM [217]	GL	FT	ViT+Transformer	X-ray images, Text	CLS, SEG	2023	✓
BiomedGPT [29]	GL	FT, PE	Transformer	Multimodal images, Text	VQA, CMG, CLS, NLI, TS	2023	✓
RadFM [27]	GL	FT, PE	ViT+Transformer	Multimodal images, Text	VQA, RG	2023	✓
ConVIRT [218]	CL	FT	ResNet+BERT	X-ray/Musculoskeletal images, Text	CLS, CMR	2022	✓
LoVT [219]	CL	FT	ResNet+BERT	X-ray images, Text	DET, SEG	2022	
UniBrain [220]	CL	FT	ResNet+BERT	MRI images, Text	CLS	2023	✓
M-FLAG [221]	CL	FT	ResNet+BERT	X-ray images, Text	CLS, DET, SEG	2023	✓
MGCA [222]	CL	FT	ResNet/ViT+BERT	X-ray images, Text	CLS, DET, SEG	2022	✓
MedKLIP [223]	CL	FT	ResNet/ViT+BERT	X-ray images, Text	CLS, SEG, PG	2023	✓
ETP [224]	CL	FT, PE	ResNet+BERT	ECG signals, Text	CLS	2024	
GLoRIA [225]	CL	FT, PE	ResNet+BERT	X-ray images, Text	CLS, SEG, CMR	2021	✓
IMITATE [226]	CL	FT, PE	ResNet+BERT	X-ray images, Text	CLS, SEG, DET	2023	
MedCLIP [227]	CL	FT, PE	ResNet/ViT+BERT	X-ray images, Text	CLS, CMR	2022	✓
Med-UniC [228]	CL	FT, PE	ResNet/ViT+BERT	X-ray images, Text	CLS, SEG, DET	2024	✓
CXR-CLIP [229]	CL	FT, PE	ResNet/Swin+BERT	X-ray images, Text	CLS, CMR	2023	✓
BiomedCLIP [230]	CL	FT, PE	ViT+BERT	Multimodal images, Text	CMR, CLS, VQA	2023	✓
UMCL [231]	CL	FT, PE	Swin+BERT	X-ray images, Text	CLS, CMR	2023	
KAD [232]	CL	FT, PE	ResNet+BERT	X-ray images, Text	CLS	2023	
MoleculeSTM [233]	CL	FT, PE	MegaMolBART/GIN+BERT	Molecule, Text	CMR, CMG, MPP	2023	✓
CLIP-Lung [234]	CL	PE	ResNet+Transformer	CT images, Text	CLS	2023	
BFSPR [235]	CL	PE	ResNet+Transformer	X-ray images, Text	CLS	2022	
MI-Zero [236]	CL	PE	CTransPath+BERT	Pathology images, Text	CLS	2023	✓
Clinical-BERT [237]	HL	FT	DenseNet+BERT	X-ray images, Text	RG, CLS	2022	
M^3AE [238]	HL	FT	ViT+Transformer	Multimodal images, Text	VQA, CLS, CMR	2022	✓
MedViLL [239]	HL	FT	ResNet+BERT	Multimodal images, Text	CLS, CMR, VQA, RG	2022	✓
PMC-CLIP [240]	HL	FT	ResNet+BERT	Multimodal images, Text	VQA, CLS, CMR	2023	✓
ARL [241]	HL	FT	ViT+BERT	Multimodal images, Text	VQA, CLS, CMR	2022	✓
MaCo [242]	HL	FT	ViT+BERT	X-ray images, Text	CLS, SEG, PG	2023	✓
MUMC [243]	HL	FT	ViT+BERT	Multimodal images, Text	VQA	2023	✓
T3D [244]	HL	FT	Swin+BERT	CT images, Text	CLS, SEG	2023	
GIMP [245]	HL	FT	ResNet+Transformer	Pathology images, Genomic	CLS	2023	✓
BioViL [246]	HL	FT, PE	ResNet+BERT	X-ray images, Text	NLI, CLS, SEG, PG	2022	
PIROR [247]	HL	FT, PE	ResNet+BERT	X-ray images, Text	CLS, SEG, DET, CMR	2023	✓
CONCH [248]	HL	FT, PE	ViT+Transformer	Pathology images, Text	CLS, CMR, SEG, CMG	2024	
ProteinDT [249]	HL	PE	ProtBERT+SciBERT	Protein sequences, Text	CMG, PPP	2023	✓
PubMedCLIP [250]	-	FT	CLIP	Multimodal images, Text	VQA	2023	✓
Med-PaLMM [31]	-	FT	PaLM-E	Multimodal images, Text, Genomic	VQA, RG, CLS, GVC	2023	✓
Med-Flamingo [251]	-	FT	Flamingo	Multimodal images, Text	VQA	2023	✓
LLaVA-Med [252]	-	FT	LLaVA	Multimodal images, Text	VQA	2024	✓
CheXZero [253]	-	FT, PE	CLIP	X-ray images, Text	CLS	2022	✓
QUILTNET [254]	-	FT, PE	CLIP	Pathology images, Text	CLS, CMR	2024	✓
PLIP [255]	-	FT, PE	CLIP	Pathology images, Text	CLS, CMR	2023	✓
CoOpLVT [256]	-	FT, PE	CLIP	Ophthalmology images, Text	CLS	2023	✓
RoentGen [257]	-	FT, PE	Stable diffusion	X-ray images, Text	CMR	2022	
Van Sonsbeek *et al.* [258]	-	FT, AT, PE	CLIP-ViT+GPT-2/BioMedLM/BioGPT	X-ray images, Text	VQA	2023	✓
Chambon *et al.* [259]	-	FT, AT, PE	Stable diffusion	X-ray images, Text	CMR	2022	
Qilin-Med-VL [260]	-	FT, AT, PE	CLIP-ViT+Chinese-LLaMA	Multimodal images, Text	VQA	2023	✓
PathAsst [261]	-	FT, AT	PLIP-ViT+Vicuna	Pathology images, Text	CLS, DET, SEG, CMR, CMG	2024	✓
PathChat [262]	-	FT, AT	CONCH-ViT+LLaMA-2	Pathology images, Text	VQA	2023	
Lu *et al.* [263]	-	FT, AT	ResNet+GPT/OpenLLaMA	X-ray images, Text	RG	2023	
M^3AD [264]	-	AT	M^3AE	Multimodal images, Text	VQA	2023	
I-AI [265]	-	AT	BiomedCLIP	X-ray images, Text	GMG, CLS	2024	✓
CITE [266]	-	AT, PE	CLIP-ViT+BioLinkBERT	Pathology images, Text	CLS	2023	✓
XrayGPT [267]	-	AT, PE	MedCLIP+Vicuna	X-ray images, Text	VQA	2023	✓
Xplainer [268]	-	PE	BioViL	X-ray images, Text	CLS	2023	✓
Qin *et al.* [269]	-	PE	GLIP	Multimodal images, Text	DET	2022	✓
Guo *et al.* [270]	-	PE	GLIP	Multimodal images, Text	DET	2023	

2) 适应: MFM 也采用 FT、AT 和 PE 范式进行模型适应，从而促进 MFM 在医疗实践中的广泛应用。

a) 基于 *FT* 的适应使用特定领域或任务的数据来调整 MFM 预训练模型的参数。一些方法尝试将通用领域的 MFM 结合 FL 适应到医疗领域。如，CheXZero [253]、PubMedCLIP [250]、QUILTNET [254] 和 PLIP [255] 是 CLIP [47] 专门针对胸部 X 光、放射学图像和组织病理学图像的微调版本。RoentGen [257] 和 Chambon 等人 [259] 基于 Stable Diffusion [21]，通过微调实现了医学领域的扩散模型。LLaVA-Med [252]、Med-Flamingo [251] 和 Med-PaLMM [31] 遵循了 LLaVA [274]、Flamingo [275] 和 PaLM-E [276] 模型，通过匹配或交错的医学图像-文本数据进行微调。具体来说，LLaVA-Med [252] 引入了一个两阶段课程学习策略，先后学习生物医学词汇对齐和开放式对话学习进行模型微调。MFM 还激发了探索视觉条件语言模型的兴趣通过微调实现统一的生物医学 AI 模型。例如，PathAsst [261]、PathChat [262]、Qilin-Med-VL [260] 和 XrayGPT [267] 将强大的视觉编码器骨干与开源大型语言模型相结合，实现了视觉语言交互式 AI 助手。然而，基础模型通常具有大量的参数，完全微调模型权重会导致训练时间长、过拟合风险和潜在的领域偏差。因此，Van Sonsbeek 等人 [258] 探索了 LoRA 和前缀微调方法，用于视觉语言交互模型的语言骨干，从而实现资源和数据高效的微调。Lu 等人 [263] 也利用 LoRA 将 LLM 适应到放射学报告生成任务中。

b) 基于 *AT* 的适应方法将适配器集成到预训练基础模型中，并微调这些适配器以适应特定领域或任务。MFM 中的适配器作为将一种模态中的特征转换为另一种模态的桥梁，发挥了独特的作用，从而以低成本融合多模态数据。一些方法 [258], [260]–[263], [267] 仅利用简单的投影层便将医学视觉特征转换为文本嵌入，能够作为文本编码器的视觉软提示。特别地，M^3AD [264] 将适配器嵌入两个单模态编码网络，并采用模态融合适配器增强多模态交互。总的来说，MFM 中的适配器可以轻量级且高效地实现不同模态间迁移。

c) 基于 *PE* 的适应方法也在 MFM 上被广泛。一些手工设计的提示 [224]–[227], [229], [235], [253], [267] 能够引导预训练模型与下游任务对齐。特别地，在提示设计的方法方面，BFSPR [235] 发现更详细的提示设计可以提高性能，因此它使用了来自小类别集合的各种组合来探索不同的设置。Qin 等人 [269] 的研究还表明，

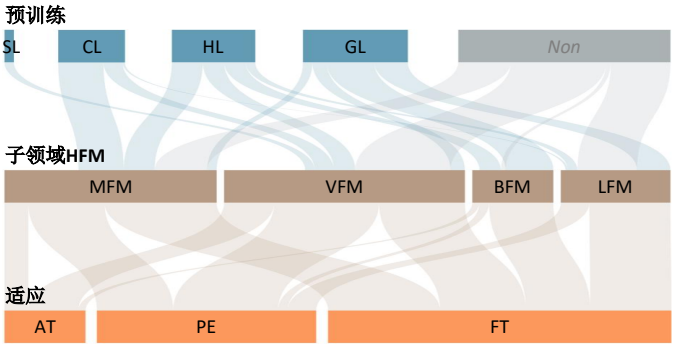

图 3. 医疗健康基础模型的桑基图展示了预训练范式、HFM 子领域和适应范式之间的关联。“Non” 代表该工作直接将现有的预训练模型用于其任务，而不是自行预训练其模型。

与默认的类别名称相比，一些基本属性，如颜色、形状和位置，可以进一步增强领域迁移能力。此外，Guo 等人 [270] 利用多个提示融合来全面描述识别对象的信息。Xplainer [268] 利用 ChatGPT 生成提示，然后请经验丰富的放射科医师对齐进行优化，以获得更好的提示性能。与其他子领域一样，MFM 也引入了可学习提示方法 [231], [234], [256], [258], [266]，利用提示微调来将预训练模型适应到不同的下游任务中，减少可训练的参数数量，同时提高对未知任务的性能。CoOpLVT [256] 和 CLIP-Lung [234] 利用提示微调来增强模态间对齐的准确性。CITE [266] 将可学习的提示 token 添加到视觉输入中，以更有效地进行病理图像分类。

E. 对 HFM 中的范式分析

如图3所示，桑基图可视化了从预训练范式到各子领域再到适应范式的论文数量流，展示了它们的各自属性和彼此关联。从这张图中我们可以得出五个对于当前 HFM 进展的观察结果：

1) 通用领域的基础模型能够适应到部分医疗健康领域。超过 1/3 的工作直接将现有的预训练模型适应于语言、视觉和多模态领域的任务，除了生物信息领域。与生物信息相比，医疗健康和通用领域中的视觉和语言相对统一，因此一些在通用领域中成功的预训练模型也拥有被推广到医疗健康任务中的能力。生物信息缺乏针对其非常特定的组学数据的通用预训练模型，因此只有两个工作 [212], [213] 研究了从 LFM 预测嵌入的适应方法。

2) 大多数 *LFM* 直接将现有的预训练语言模型适应于医疗健康任务。因为语言是由人类创造的，具有很

强的可迁移性。在通用领域预训练的 LFM 拥有这种可迁移性，获得了适应到医疗健康领域任务的能力。

3) 大多数预训练工作集中在无需标签的学习上。由于预训练数据量巨大，使用监督学习需要大量人工标注的标签，成本较大。因此大多数工作集中于自监督范式，即 GL、CL 或 HL，以低标注成本学习通用的表征。

4) 监督学习仍然用于 *VFM* 预训练。因为视觉信息的连续性使得仅通过自监督范式难以解耦图像上下文中的语义 [113]。因此，仍有一些工作尝试利用监督学习为模型提供更直接的优化目标，从而驱动模型解耦图像内部的语义并学习有意义的特征。

5) 微调仍然广泛用于适应。超过一半的工作利用微调范式，因为微调可以带来稳定的学习过程。一些新的微调技术，如 LoRA，正在探索以实现低参数、高效率的适应方法。

III. 数据集

A. 语言

医疗健康 LFM 的快速进展归功于不同的医疗健康文本数据集。尽管这些文本数据中积累了丰富的信息，但在规模和针对性方面仍然具有挑战性。许多 LFM 工作结合了多种数据集，从而创建更加全面的训练语料库，其中详细的组成部分可以在其文章中获得。[64], [66], [67], [72], [74], [76], [77]. 如表V所示，我们描述了文献、健康记录和对话中的大规模医疗健康语言数据集，这些数据集对于 LFM 在理解和处理医学术语方面非常重要。这些数据集的实例数超过 4K 或 token 数超过 10 亿，适用于 LFM 的训练和评估。

1) 医疗健康文献: 由于医学知识浓缩在文献文本中，医疗健康文献数据对 LFM 非常重要，一些大型数据集已经公开提供。它们涵盖的内容通常是广泛的，为通用领域语言基础模型注入了丰富的医学知识。PubMed[1]是一个大规模的数据库，主要包括生命科学和生物医学主题的 MEDLINE 数据库的参考文献和摘要，为开发医疗健康 LFM 提供了全面的医疗健康文献数据。MedC-I [65] 从论文、书籍、对话、Rationale QA 和知识图谱中收集了超过 79B 个 token。Guidelines [67] 由来自 17 个高质量在线医学源的 47K 个临床实践指南组成。PMC-Patients [277] 则由从 PubMed Central 的病例报告中提取的 167K 个患者摘要组成。

[1]https://pubmed.ncbi.nlm.nih.gov/download/

2) 电子病历: 电子病历包含了大量有临床价值的疾病描述和诊断信息。这些数据将使 LFM 获得理解临床场景和患者预后的能力，从而增强模型在具有挑战性的医疗健康实践中的性能。由于电子病历数据中包含大量隐私信息，它们的数据集通常比医疗健康文献数据集小得多。MIMIC-III [278] 包含了来自四万名住院于重症监护单位的患者的超过 122K 条健康记录，提供了重症监护实践中的详细临床知识，而更新版本 MIMIC-IV [279] 包含了 299K 条临床记录。EHRs [15] 包含了来自 UF Health IDR 的超过 2.9 亿条临床笔记。MD-HER [76] 包含了 100K 条记录，涵盖了一系列疾病组，而 eICU-CRDv2.0 [280] 则包含了 208 家美国医院的 ICU 和步进单元中的 200,859 个住院记录。然而，电子病历数据集仍然很少，EHRs [15] 和 MD-HER [76] 目前仍然尚未公开。

3) 医疗健康对话: 医疗健康对话数据记录了医生和患者或医生之间在医疗健康场景中的交互。这些对话对于提升模型的交流技巧和信息检索过程非常重要。已有许多公开可用的医疗健康对话数据集 [64], [68], [69], [71], [72], [84], [86], [90], [103], [104], [282]–[284], [284], [286], [288]–[291], 并且一些技术也可以将其他医疗健康文本数据或分类标签转化为对话 [292]。BianqueCorpus [69] 数据集基于包括 MedDialog [290]、IMCS-21 [281]、CHIP-MDCFNPC [293]、MedDG [285]、cMedQA2 [287] 和 CMD 在内的多个现有数据集构建而成，合成了一个包含 2.4M 个实例的中文医学对话数据集。对于英语医学对话，Medical Meadow [71] 也融合了 11 个自创建数据集和 7 个外部数据集，合成了包含 160K 个实例的大型数据集。

B. 视觉

VFM 的成功同样依赖于大规模的医学图像数据集，因此许多最近的方法和 LFM 一样通过混合了多个公开或私有的数据集来构建一个大型医学图像数据集。这些工作的详细数据集列表可在它们的文章中找到 [19], [42], [118]–[120], [123], [125], [137], [137], [162], [294]。如表III-B3所示，我们回顾了公开可用且相对较大的数据集，其中包括数据量超过 1K 的 3D 医学图像、超过 1K 的 2D 全切片图像（WSI）和超过 10k 的其他 2D 医学图像/视频帧。

1) 三维医学图像: 三维医学图像，包括三维 CT、MRI、PET 等，可以可视化人体内部信息，在临床实

表 V
具有代表性的语言数据集，其中三个数据集暂时无法获得。本文中，我们使用语言 TOKEN 数量或数据实例（INSTANCE）数量来表示数据集的规模。这里的缩写是 LM：语言建模（LANGUAGE MODELING），DIAL：对话（DIALOGUE），IR：信息检索（INFORMATION RETRIEVAL），NRE：命名实体识别（NAMED ENTITY RECOGNITION），RE：关系提取（RELATION EXTRACTION），STS：语义文本相似性（SEMANTIC TEXTUAL SIMILARITY），NLI：自然语言推理（NATURAL LANGUAGE INFERENCE），QA：问答（QUESTION ANSWERING），VQA：视觉问答（VISUAL QUESTION ANSWERING）。

数据集	文本类型	规模	任务	连接
PubMed	Literature	18B tokens	LM	✓
MedC-I [65]	Literature	79.2B tokens	DIAL	✓
Guidelines [67]	Literature	47K instances	LM	✓
PMC-Patients [277]	Literature	167K instances	IR	✓
MIMIC-III [278]	Health record	122K instances	LM	✓
MIMIC-IV [279]	Health record	299K instances	LM	✓
eICU-CRDv2.0 [280]	Health record	200K instances	LM	✓
EHRs [15]	Health record	82B tokens	NER, RE, STS, NLI, DIAL	
MD-HER [76]	Health record	96K instances	DIAL, QA	
IMCS-21 [281]	Dialogue	4K instances	DIAL	✓
Huatuo-26M [282]	Dialogue	26M instances	QA	✓
MedInstruct-52k [68]	Dialogue	52K instances	DIAL	✓
MASH-QA [283]	Dialogue	35K instances	QA	✓
MedQuAD [284]	Dialogue	47K instances	QA	✓
MedDG [285]	Dialogue	17K instances	DIAL	✓
CMExam [286]	Dialogue	68K instances	DIAL, QA	✓
cMedQA2 [287]	Dialogue	108K instances	QA	✓
CMtMedQA [84]	Dialogue	70K instances	DIAL, QA	✓
CliCR [288]	Dialogue	100K instances	QA	✓
webMedQA [289]	Dialogue	63K instances	QA	✓
ChiMed [72]	Dialogue	1.59B tokens	QA	✓
MedDialog [290]	Dialogue	20K instances	DIAL	✓
CMD	Dialogue	882K instances	LM	✓
BianqueCorpus [69]	Dialogue	2.4M instances	DIAL	✓
MedQA [104]	Dialogue	4K instances	QA	✓
HealthcareMagic	Dialogue	100K instancess	DIAL	✓
iCliniq	Dialogue	10K instances	DIAL	✓
CMeKG-8K [103]	Dialogue	8K instances	DIAL	✓
Hybrid SFT [64]	Dialogue	226K instances	DIAL, QA	✓
VariousMedQA [91]	Dialogue	54K instances	VQA	✓
Medical Meadow [71]	Dialogue	160K instances	QA	✓
MultiMedQA [89]	Dialogue	193K instances	QA	
BiMed1.3M [90]	Dialogue	250K instances	QA	✓
OncoGPT [86]	Dialogue	180K instances	QA	✓

践中被广泛使用。医学分割十项全能挑战赛 [295] 共开放了 1,411 个 3D CT 和 1,222 个 MRI 图像，以评估 10 种器官或疾病的语义分割算法。ULS 挑战赛 [296] 进一步开放了 38,842 个 CT 图像，以评估通用病变分割，促进了病变分割的 VFM。一些其他的 CT 数据集，包括 LIDC-IDRI、TotalSegmentator [117], [297]、FLARE 2022、2023 [296]、AbdomenCT-1K [298]、CTSpine1K [299]、CTPelvic1K [300]，也开放了超过 1K 个三维 CT 图像用于分割任务。BraTS [301]–[304] 挑战已经开放了超过 2K 个带有多种序列的脑 MRI 体积，用于脑 MRI 分析。ADNI [305] 和 PPMI [306] 数据库维护阿尔茨海默病和帕金森病的脑 MRI 图像和其他临床数据，为这些疾病的临床研究做出了贡献。AutoPET 挑战 [307], [308] 开放了 1,214 个 PET-CT 对，促进了跨模态图像分析的研究。

2) 全切片图像（WSI）： WSI 是通过显微镜可视化人体组织切片的图像，被广泛用于癌症或癌前病变诊断。与其他二维医学图像不同，WSI 具有非常的分辨率（如 150,000×85,000 像素）[309]，使得现有方法难以在全局视野上对其直接进行分析。TCGA 数据库 [310] 包含了几个 WSI 数据集，包括 NSCLC、Lung、BRCA、GBM、KIRC、LUAD、LUSC、OV 等，涵盖了 33 种

癌症类型。PAIP [311] 和 TissueNet [309] 分别组织了公开挑战，用于肝癌和宫颈癌的病理分割和诊断，均包含超过 1K 的 WSI。一些其他数据 [312], [313] 从 WSI 中裁剪出更小的图像块，用于分类相关任务的研究。

3) 其他二维图像和视频: 二维医学图像或视频在医疗实践中也被广泛使用。X 射线成像被广泛用于疾病筛查和手术辅助，整个社区中积累了大量数据，因此有许多超过 10K 张图像的大型 X 射线数据集被公开 [314], [315]。ISIC 挑战 [316] 开放了超过 30K 张皮肤镜图像，促进了皮肤病的诊断。AIROG 挑战 [317] 开放了超过 100K 张眼底照片，用于青光眼筛查。Retinal OCT-C8 数据集 [318] 融合了多种来源的数据，开放了 24K 张眼底 OCT 图像，用于视网膜疾病的诊断。对于超声（Ultrasound, US）图像，超声神经分割挑战赛 [319] 公开了 11K 张图像用于臂丛神经分割。Fetal planes 数据集 [320] 开放了 12,400 张 US 图像，用于胎儿筛查。US 图像也用于心脏疾病分析，EchoNet-Dynamic [321] 构建了一个大型心脏 US 视频数据集，用于心脏功能评估。内窥镜视频广泛用于胃肠疾病检测和手术，已有多个公开数据集 [322]–[327] 在内窥镜场景中开放了大量的视频帧数据。

C. 生物信息

高通量测序已经成为生物领域的基本技术，已拥有超过十年的发展 [214]。因此，大量的 DNA、RNA、蛋白质和单细胞 RNA 测序数据被大规模扫描，为生物研究提供了丰富的数据信息。公开可用的测序数据在社区内被快速积累，使得研究人员能够用于训练 BFM。如表III-B3所示，我们列出了一些包含数百万表达值、序列或结构的大规模生物数据集。

1) 基因组学和单细胞组学数据: 基因组学和单细胞组学数据提供了全面的遗传信息、基因表达模式和细胞功能的洞见。DNA 序列代表生物体的遗传地图，而单细胞 RNA 测序则在单个细胞中分析基因表达，揭示功能多样性和细胞响应。NCBI GenBank [339] 是所有公开可用 DNA 序列的带标签集合。目前，它包含了多达 37 亿条序列。GenCode [341] 是一个旨在注释整个人类基因组中所有基于证据的基因特征的科学项目。其目标是识别所有基因特征，包括编码蛋白质序列、非编码 RNA、假基因及其变异体。基因组序列也包含在该数据集中。CellxGene Corpus [338] 是一个影响广泛的单细胞语料库，包含来自 1,219 个数据集的 789 个细胞类型，总细胞表达值数量超过 7200 万。UK Biobank [352] 提供了 50 万研究参与者的所有信息的摘要，包括成像、遗传学、健康联系、生物标志物、活动监测器、在线问卷、重复基线评估等。SCP [340] 汇总了来自 645 个不同研究的超过 4 千万个细胞。该平台由各种生物数据组成，包括 14 个物种、83 种疾病、104 个器官和 160 种不同的细胞类型。该数据集提供了对各种生物背景和条件下的细胞行为的宝贵洞见。其他数据集，如 Human Cell Atlas [342]、10x Genomics、Allen Brain Cell Atlas 等，也提供了丰富的生物基因组数据。

2) 转录组学和蛋白质组学数据: 转录组学和蛋白质组学共同阐明了从遗传信息到功能蛋白质的过程，揭示了复杂的细胞过程。转录组学关注 RNA 序列以了解基因表达，而蛋白质组学分析蛋白质结构，揭示了复杂的细胞机制。Ensembl 项目 [345] 提供了人类以及其他生物医学感兴趣的物种的自动标注的基因组序列。研究人员可以在该数据库中找到基因组序列和相应的蛋白质序列。RNAcentral [346] 是一个非编码 RNA（ncRNA）序列数据库，汇集了多个用于 ncRNA 研究的数据集。该数据集提供了一个入口点，能够访问来自所有生物体的所有 ncRNA 类型的 ncRNA 序列。它目前包含来自 53 个数据库的超过 3600 万个 ncRNA 序列。AlphaFold 蛋白质结构数据库 [25] 公开了超过 2 亿个蛋白质结构。Protein Data Bank in Europe（PDBe）[347] 和 UniProt [348] 也提供了数百万个蛋白质结构。

3) 其他大规模生物数据库: 还有其他大规模的生物数据库可用于疾病研究和药物反应。LINCS L1000 数据集 [349] 包括有关不同类型的人类细胞在对各种扰动时作出的反应信息，例如暴露于药物、毒素或基因修饰。它测量了约 1,000 个被精心选择以代表整个人类基因组的标志性基因的表达水平，涵盖了超过 41k 个小分子。Genomics of Drug Sensitivity in Cancer（GDSC）数据集 [350] 包含 1,000 个人类癌细胞系，并使用大约 400 种化合物对其进行筛选。还有其他一些数据库（表III-D2），如 Cancer Cell Line Encyclopedia（CCLE）[351] 和 The Cancer Genome Atlas Program（TCGA）[310]，可用于验证癌症靶点和定义药物疗效，Chinese Glioma Genome Atlas（CGGA）[353] 可用于胶质瘤相关疾病研究。UK Biobank [352] 是一个大规模的生物医学数据库和研究资源，包含来自 50 万英国参与者的深入遗传和健康信息。独特而丰富的数据资源，包括遗传、生活方式和健康信息，为研究遗传、环境和生活方

表 VI
公开可获得的视觉数据集。这里的缩写是 CLS：分类（CLASSIFICATION），SEG：分割（SEGMENTATION），DET：检测（DETECTION），REG：配准（REGISTRATION），US：超声（ULTRASOUND）。“LINICAL STUDY”意味着这是一个没有明确任务指导的综合数据集。

数据集	模态	规模	任务	连接
LIMUC [322]	Endoscopy	1043 videos (11,276 frames)	DET	✓
SUN [323]	Endoscopy	1018 videos (158,690 frames)	DET	✓
Kvasir-Capsule [324]	Endoscopy	117 videos (4,741,504 frames)	DET	✓
EndoSLAM [325]	Endoscopy	1020 videos (158,690 frames)	DET, REG	✓
LDPolypVideo [328]	Endoscopy	263 videos (895,284 frames)	DET	✓
HyperKvasir [326]	Endoscopy	374 videos (1,059,519 frames)	DET	✓
CholecT45 [327]	Endoscopy	45 videos (90,489 frames)	SEG, CLS	✓
DeepLesion [329]	CT slices (2D)	32,735 images	RET, CLS	✓
LIDC-IDRI [330]	3D CT	1,018 volumes	SEG	✓
TotalSegmentator [117]	3D CT	1,204 volumes	SEG	✓
TotalSegmentatorv2 [297]	3D CT	1,228 volumes	SEG	✓
AutoPET [307], [308]	3D CT, 3D PET	1,214 PET-CT pairs	SEG	✓
ULS	3D CT	38,842 volumes	SEG	✓
FLARE 2022 [296]	3D CT	2,300 volumes	SEG	✓
FLARE 2023	3D CT	4,500 volumes	SEG	✓
AbdomenCT-1K [298]	3D CT	1,112 volumes	SEG	✓
CTSpine1K [299]	3D CT	1,005 volumes	SEG	✓
CTPelvic1K [300]	3D CT	1,184 volumes	SEG	✓
MSD [295]	3D CT, 3D MRI	1,411 CT, 1,222 MRI	SEG	✓
BraTS21 [301]–[303]	3D MRI	2,040 volumes	SEG	✓
BraTS2023-MEN [304]	3D MRI	1,650 volumes	SEG	✓
ADNI [305]	3D MRI	-	Clinical study	✓
PPMI [306]	3D MRI	-	Clinical study	✓
ATLAS v2.0 [331]	3D MRI	1,271 volumes	SEG	✓
PI-CAI [332]	3D MRI	1,500 volumes	SEG	✓
MRNet [333]	3D MRI	1,370 volumes	DET, SEG	✓
Retinal OCT-C8 [318]	2D OCT	24,000 imsges	CLS	✓
Ultrasound Nerve Segmentation [319]	US	11,143 images	SEG	✓
Fetal Planes [320]	US	12,400 images	CLS	✓
EchoNet-LVH [334]	US	12,000 videos	DET, Clinical study	✓
EchoNet-Dynamic [321]	US	10,030 videos	Function assessment	✓
AIROGS [317]	CFP	113,893 images	CLS	✓
ISIC 2020 [316]	Dermoscopy	33,126 images	CLS	✓
LC25000 [312]	Pathology	25,000 images	CLS	✓
DeepLIIF [335]	Pathology	1,667 WSIs	SEG	✓
PAIP [311]	Pathology	2,457 WSIs	SEG	✓
TissueNet [309]	Pathology	1,016 WSIs	CLS	✓
NLST [336]	3D CT, Pathology	26,254 CT, 451 WSIs	Clinical study	✓
CRC [313]	Pathology	100k images	CLS	✓
MURA [315]	X-ray	40,895 images	DET	✓
ChestX-ray14 [314]	X-ray	112,120 images	DET	✓
SNOW [337]	Synthetic pathology	20K image tiles	SEG	✓

式在决定健康结果方面的复杂相互作用提供了前所未有的机会。

D. 多模态

多模态医疗数据的积累和大规模多模态数据集的整理构建是医疗多模态框架成功的基础。然而，由于医疗图像和文本数据的可访问性，目前大部分多模态医疗数据集仍然局限于视觉和语言。如表III-D所示，我们总结了公开可用且相对较大的医疗多模态数据集，包括图像、文本描述、脑电图（EEG）信号、蛋白质和分子信息。在此，我们将讨论视觉语言数据集和超越视觉和语言的多模态数据集。

1) 视觉语言数据： 由于医学图像模态间具有不同的特性，现有医疗视觉语言数据集的规模和组成也是多

表 VII
公开可获得的生物信息数据集。“研究”意味着这是一个没有明确任务指导的生物信息综合数据集。

Dataset	Modalities	Scale	Tasks	Link
CellxGene Corpus [338]	scRNA-seq	over 72M scRNA-seq data	Single cell omics study	✓
NCBI GenBank [339]	DNA	3.7B sequences	Genomics study	✓
SCP [340]	scRNA-seq	over 40M scRNA-seq data	Single cell omics study	✓
GenCode [341]	DNA	-	Genomics study	✓
10x Genomics	scRNA-seq, DNA	-	Single cell omics and genomics study	✓
ABC Atlas	scRNA-seq	over 15M scRNA-seq data	Single cell omics study	✓
Human Cell Atlas [342]	scRNA-seq	over 50M scRNA-seq data	Single cell omics study	✓
UCSC Genome Browser [343]	DNA	-	Genomics study	✓
CPTAC [344]	DNA, RNA, protein	-	Genomics and proteomics study	✓
Ensembl Project [345]	Protein	-	Proteomics study	✓
RNAcentral database [346]	RNA	36M sequences	Transcriptomics study	✓
AlphaFold DB [25]	Protein	214M structures	Proteomics study	✓
PDBe [347]	Protein	-	Proteomics study	✓
UniProt [348]	Protein	over 250M sequences	Proteomics study	✓
LINCS L1000 [349]	Small molecules	1,000 genes with 41k small molecules	Disease research, drug response	✓
GDSC [350]	Small molecules	1,000 cancer cells with 400 compounds	Disease research, drug response	✓
CCLE [351]	-	-	Bioinformatics study	✓

表 VIII
公开可获得的多模态数据集。这里的缩写是 QA：问答（question answering），VQA：视觉问答（visual question answering）。“Multimodal learning”表示这是一个没有明确任务指导的综合数据集。

数据集	模态	规模	任务	连接
MIMIC-CXR [354]	X-ray images, Medical report	377K images, 227K texts	Vision-Language learning	✓
PadChest [355]	X-ray images, Medical report	160K images, 109K texts	Vision-Language learning	✓
CheXpert [356]	X-ray images, Medical report	224K images, 224K texts	Vision-Language learning	✓
ImageCLEF2018 [357]	Multimodal images, Captions	232K images, 232K texts	Image captioning	✓
OpenPath [255]	Pathology images, Tweets	208K images, 208K texts	Vision-Language learning	✓
PathVQA [358]	Pathology images, QA	4K images, 32K QA pairs	VQA	✓
Quilt-1M [254]	Pathology images, Mixed-source text	1M images, 1M texts	Vision-Language learning	✓
PatchGastricADC22 [359]	Pathology images, Captions	991 WSIs, 991 texts	Image captioning	✓
PTB-XL [360]	ECG signals, Medical report	21K records, 21K texts	Vision-Language learning	✓
ROCO [361]	Multimodal images, Captions	87K images, 87K texts	Vision-Language learning	✓
MedICaT [362]	Multimodal images, Captions	217K images, 217K texts	Vision-Language learning	✓
PMC-OA [240]	Multimodal images, Captions	1.6M images, 1.6M texts	Vision-Language learning	✓
ChiMed-VL [260]	Multimodal images, Medical report	580K images, 580K texts	Vision-Language learning	✓
PMC-VQA [363]	Multimodal images, QA	149K images, 227K QA pairs	VQA	✓
SwissProtCLAP [249]	Protein Sequence, Text	441K protein sequence, 441K texts	Protein-Language learning	✓
Duke Breast Cancer MRI [364]	Genomic, MRI images, Clinical data	922 patients	Multimodal learning	✓
I-SPY2 [365]	MRI images, Clinical data	719 patients	Multimodal learning	✓

样的。对于 X 射线成像，MIMIC-CXR [354] 是最常用的 MFM 预训练数据集。它包含 377,110 张胸部 X 射线图像和相应的 227,835 份医疗报告。PadChest [355] 和 CheXpert [356] 也包括胸部 X 射线图像和相应的医疗报告，丰富了胸部 X 射线图像的种类和数量。对于病理学成像，OpenPath [255] 爬取了 Twitter 上人们发出病理图像，构建了包含超过 200K 张由医疗专业人员描述的病理图像-文本数据集。PathVQA [358] 是一个病理 VQA 数据集，包括 4,998 张图像和 32,799 个 QA 对。Quilt-1M [366] 是一个大型组织病理学数据集，包括 1M 个图像-文本对。PatchGastricADC22 [359] 包括从 991 个 WSI 上提取的 262,777 个图像块，以及相关的诊断字幕。此外，还有一些具有多个医学图像模态的视觉语言数据集，包括 ROCO [361]、MedICaT [362]、PMC-OA [240]、ChiMed-VL [260]。这些数据集提供了各种图像模态，涵盖了放射学和组织学领域。其

中，PMC-VQA [363] 是一个多模态医学视觉问答数据集，包含总共 227K 个视觉问答对和 149K 张图像。

2) 其他多模态数据: 除了视觉语言数据，还有一些其他公开可用的医疗多模态数据集。SwissProtCLAP [249] 包括 441,000 个蛋白质序列-文本对。与视觉语言领域中的图像-文本对相比，它的规模相对较小。PTB-X [224] 包含 21,837 个 EEG 信号及其相应的医疗报告。Duke Breast Cancer MRI [364] 包括来自 922 名经活检确认的乳腺癌患者的多序列 MRI 图像和病理学、临床治疗和基因组数据。I-SPY2 [364] 包括来自 719 名乳腺癌患者的超过 4TB 的 MRI 和临床数据。此外，还有一些大规模的综合数据库（表III-D2）包含不同模态的大量医疗数据。TCGA [310] 是一个里程碑式的癌症基因组学计划，包括 2.5PB 的基因组、病理图像、病理报告和其他多模态数据。

IV. 应用

A. 语言

由于文本在医疗实践中的广泛应用，LFM 在诊断、教育、咨询等方面已经取得了显著的应用。特别是随着一些著名的 LLM 被广泛使用，如 ChatGPT [39]，LFM 的临床应用潜力受到了广泛关注并被进一步探索。一些通用的医疗语言模型，如 BianQue [69] 和 Med-PaLMM [31]，已经在医疗场景中取得了成功。

1) 医疗诊断: 基于 LFM 的医疗诊断模型可以通过医疗测试和患者描述来预测最可能的疾病，对于治疗和预防并发症至关重要 [368]。最近，LFM 已被用于提升医学诊断，并在不同疾病上展现出了通用能力 [64], [75], [76], [92]。Ueda 等人 [369] 利用患者病史和影像检查结果通过 ChatGPT 诊断胸腔积液。Wu 等人 [370] 也在甲状腺结节的诊断中评估了三个 LFM，展示了 LFM 在提高医学影像诊断方面的应用潜力。尽管 LFM 已经展示了诊断能力，但临床医生需要追踪和理解每个诊断决策背后的逻辑，缺乏透明度仍然是一个巨大的挑战(在第V节中讨论)。

2) 报告生成: LFM 已经表现出了在生成医疗报告方面的潜力，包括放射学报告 [89]、出院总结 [39] 和转诊信 [371]。这些模型擅长从多种信息源（如电子病历、医学文献和临床指南）中融合信息，生成连贯和信息丰富的报告。医生通常认为写医疗报告是繁琐和耗时的工作，因此利用医学语言模型可以减轻他们的工作负担。一种方法是将诊断结果输入到语言模型中，然后将其作为总结工具生成报告 [93]。因此，可以在不进行手动编辑的情况下生成合理的报告。另一些方法通过输入一些图像描述，使得放射科医生能够快速诊断 [370] 并生成报告。这样的模型包括 ChatCAD [93]、ChatCAD+ [94]、Visual Med-Alpaca [91] 和 MedAgents [99]。

3) 医疗健康教育: 对于从业人员和普通公众来说，LFM 也在医疗健康教育中扮演着重要的角色 [372]。对于医学生，这些模型能够生成医学问题，增强他们对医学知识的理解 [373]。LFM 也可以扮演医学教师的角色，为学生的临床问题提供专业的答案。Kung 等人 [374] 已经评估了 ChatGPT 在医学教育中的潜力。一些用于医学教育的模型，如专注于中医考试的 HuatuoGPT-II [66]，也已经开发，帮助医疗教育的发展。对于普通公众，LFM 也可以将复杂的医学术语翻译成易于理解的语言，促进公共医疗教育 [373]。一项研究 [375] 探讨了 ChatGPT 和电子健康素养的结合，说明了 LFM 在提高健康服务的可访问性和质量方面的显着潜力。

4) 医疗咨询: LFM 可以提高医疗咨询的质量 [376]，对于医疗健康来说十分重要。这些模型可以利用其内部所学知识和医学网站（如健康论坛和教科书）的信息，为患者提供自我诊断或其他目的的医学信息。此外，这些模型还可以作为聊天机器人，为患者提供心理健康支持 [377]，从而提高他们的情绪并减轻心理健康专业人员的负担。现有的几个用于医疗咨询的模型，包括 BenTsao [63]、MedPaLM [16]、MedPaLM 2 [89] 等 [68]–[70], [74], [74], [76]，都展示了利用 LFM 提高医疗咨询和相关服务的质量和效率的可行性和有效性。

B. 视觉

VFM 在分割、分类、检测等任务中也取得了成功，展示了它们在赋能放射科医生、外科医生或临床医生方面，以及协助诊断、预后、手术或其他医疗实践的工作流程中的巨大应用前景。

1) 医疗诊断: VFM 也在基于医学图像的诊断方面展示了应用潜力 [27]。它们可以在一些低风险图像上进行自动疾病筛查，并协助检测和识别不清晰的目标解剖结构，从而减轻放射科医生的工作负担并提高他们的诊断准确性。分割和检测 VFM 提供医学图像中的位置信息，包括器官 [125], [135], [137], [156], [162], [177]、肿瘤 [42], [149] 和病变 [19], [135], [165]，从而协助放射科医生将图像解耦成语义区域并发现感兴趣的区域。分类 VFM 也通过直接预测输入图像的类别来促进自动疾病

表 IX
包含来自多个子领域的大规模医疗健康综合数据库。

数据库	描述	连接
CGGA [353]	中华神经胶质瘤基因组图谱数据库（CGGA）包含来自中国的超过 2,000 个脑肿瘤样本的临床和测序数据。	✓
UK Biobank [352]	UK Biobank 是一个大规模的生物医学数据库，包含来自 50 万英国参与者的去识别化的遗传、生活方式、健康信息以及生物样本。	✓
TCGA [310]	癌症基因组图谱计划（TCGA）对超过 20,000 个原发性癌症和 33 种癌症类型的正常样本进行了分子表征，并生成了超过 2.5PB 的基因组、表观基因组、转录组和蛋白质组数据。	✓
TCIA [367]	癌症影像档案库（TCIA）是一项用于识别和托管大型公共癌症医学影像地档案库。	✓

诊断 [19], [41], [125], [126], [133], [134], [141]，从而有效降低像体检筛查这样的低风险图像的成本。然而，由于可信度的限制，对于一些高风险的诊断应用，如肿瘤分级，仍然具有挑战性。

2) 疾病预后: 一些 VFM 在疾病预后方面也取得了有前途的结果，能够提供一些生物标志物来预测疾病的可能性或预期发展。因此，临床医生或放射科医生能够根据预后结果为患者制定干预计划。一些大规模预训练的 VFM，如 RETFound [19] 和 VisionFM [122]，能够从视网膜图像中提取与眼科疾病相关的代表性特征，因此这些特征有可能作为生物标志物代表疾病的进展。一些分割或检测 VFM [42], [162] 也可以提供病变（例如肿瘤）的形状、大小和位置，作为潜在的生物标志物。然而，由于构建大规模随访数据集存在巨大困难，直接为生存预测等预后应用构建基础模型仍然具有挑战。

3) 手术计划和辅助: 手术是 VFM 的另一个潜在应用场景，它构建了即插即用的医学图像处理工具，用于手术规划或手术辅助，而无需在传统范式中进行额外的数据收集和模型训练。对于手术规划，外科医生能够通过一些 3D 分割 VFM（如 SAM-Med-3D [119]）从医学图像（如 CT 和 MRI）中分割出三维物体，从而可视化感兴趣的物体进行规划。在手术过程中，像 SP-SAM [150]（一种分割 VFM）这样的 VFM 也可以在内窥镜视野中分割工具或感兴趣的区域，从而协助手术并提高手术结果。然而，在外科医生手术时无法操作机器的情况下，医生和 VFM 之间的交互仍然存在挑战。

4) 其他应用: VFM 在医疗健康领域的应用不仅限于诊断、预后和手术。例如，CTransPath [131] 在检索和医疗软件原型开发方面表现出了有效性，促进相关医学图像的高效检索，并协助开发医疗设备和系统的原型。USFM [125] 则为图像增强技术做出了贡献，提高了医学图像的质量、清晰度和可解释性，从而协助精准的诊断和治疗计划的制定。通过它们多样化的应用，VFM 在推动医学成像技术的发展和提高患者护理方面发挥着关键作用，覆盖各种临床领域。

C. 生物信息

BFM 可用于多种下游任务，包括序列分析、交互分析、结构和功能分析以及疾病和药物研究。这些应用场景结合 BFM 的计算能力揭示了生命的复杂性，在生物学洞见方面发挥着重要作用。

1) 序列分析: BFM 在序列分析上得到了广泛的应用。研究人员利用 BFM 对 DNA 序列，RNA 序列和蛋白质序列进行分析，从而推进了对基因组和转录组的理解，揭示了复杂的生物过程和分子的相互作用。大量的工作通过 BFM 进行此类分析研究，推进了基因组和转录组学的进展 [23], [46], [186], [187], [190], [191], [193], [194], [199], [202], [204], [206], [208]。一些工作将 BFM 用于启动子检测任务 [23], [46], [190], [191], [199], [208]，从而识别 DNA 序列中的启动子。这些启动子是启动基因转录的关键元素，是控制基因表达的关键调节序列。

2) 交互分析: 交互分析是 BFM 的另一个潜在研究应用。它可以帮助人们了解细胞系统内复杂的相互作用和调节机制。目前，BFM 已经有效地分析了基因之间的相互作用 [189], [203], [212]，蛋白质和 RNA 之间的相互作用 [26] 以及蛋白质之间的相互作用 [212]。例如，利用 RNA-FM [26] 编码的嵌入序列，研究者取得了交互分析的最佳性能。模型预测的性能可与带有真实二级结构的序列相媲美，这表明 RNA-FM 的嵌入提供了足够的信息，可以替代真实的二级结构。

3) 结构和功能分析: BFM 也已经在结构和功能分析中得到了应用，该应用揭示了分子结构和生物功能之间的复杂关系，增强了人们对细胞行为和遗传变异的理解。在基因组学 [23], [46]、转录组学 [26], [192], [193], [197], [204], [206], [209]、蛋白质组学 [25], [187], [195], [196], [198], [205], [210] 和单细胞组学 [188], [201],

[203], [207] 等领域，已经取得了显著的成功。蛋白质结构预测是蛋白质组学中最受欢迎的任务之一，著名的 Alphafold [25] 在多序列比对数据集上预训练，从而指示了序列之间的功能、结构或进化关系。xTrimoPGLM [210] 模型也实现了蛋白质结构预测和生成，在大规模参数中显著优于其他先进的基线模型。

4) 疾病研究和药物反应: 疾病研究和药物反应在推进医学知识方面起着关键作用，促进了创新疗法和治愈方法的发展，从而改善人类健康状况并延长了寿命。BFM 可以进行药物敏感性预测 [189], [200], [207]、转录因子剂量敏感性预测 [189], [212]、药物反应预测 [200]、疾病风险评估 [208]、细胞干扰响应预测 [200] 和 COVID 变异分类 [190]，为基因因素和治疗结果之间复杂相互作用提供了洞见，从而促进了个性化医学和对疾病机制的认识。例如，scGPT [203] 利用已知实验中细胞响应所获得的知识，并将其推广到预测未知响应。在基因维度上利用自注意机制，可以对受扰动的基因和其他基因之间的复杂相互作用关系进行编码，从而预测在组合空间中潜在的基因扰动情况 [203]。

D. 多模态

MFM 能够融合来自不同模态的信息，从而提高基于单一模态的某些应用的性能（例如诊断），并且还实现某些模态间的跨模态生成（例如报告生成）。

1) 医疗诊断: 虽然 LFM 和 VFM 在医学诊断方面展示了很有前途的应用潜力，MFM 进一步整合了来自患者的多种数据来源，利用了 AI 模型在诊断方面的能力，协助医生做出更准确的诊断决策。特别是，一些基于 CLIP 的预训练 MFM（例如 Gloria [225]、ETP [224]、Expert [253] 和 Visual [255]）能够通过提示实现零样本分类，适用于开放式的疾病诊断环境。与 VFM 相比，医疗文本中的上下文信息能够进一步丰富特征表示，尤其是在视觉特征不明显的情况下，实现有效的诊断。对于不同模态的医学图像，MFM（例如 RadFM [27]）也已经实现了融合信息的能力，根据不同成像条件的特征促进诊断结果。像 Qilin-Med-VL [260]、Med-Flamingo [251]、LLaVA-Med [252] 这样的视觉语言 MFM 具有医疗对话能力，能够通过提问预测疾病描述，例如疾病类型和医学图像上的位置。

2) 报告生成: 与 LFM 中的医学报告生成不同，视觉语言 MFM 从医学图像中生成放射学报告，加速了基于影像的诊断效率。一些大规模的 MFM，例如 RadFM [27]、Clinical-BERT [237] 和 MedViLL [239] 利用图像和文本信息，实现了对患者数据更全面的理解，并生成来自医学图像的报告。与传统的手工撰写相比，它展现出了解放放射科医师繁琐报告写作的能力。此外，最近一些 MFM 也尝试支持交互式报告生成，基于提示以按照特定模板和格式创建医学报告 [267]。因此，它进一步展现出减轻医疗专业人员工作负担和提高报告输出稳定性的双重优势。

3) 生物科学研究: MFM 提供了一个用于连接生命语言（例如 DNA、RNA、蛋白质等）和人类自然语言的解决方案，展示了其在生物科学中的潜在应用。分子-语言 MFM [233] 在不需要额外分子数据和注释的情况下，使得分子编辑可以通过文本提示实施，促进了生物化学研究。MFM（例如 BioMedGPT-10B [378]）也能够扮演生物信息学专家的角色，并通过语言与人类进行生物医学研究和开发的对话。研究人员可以上传生物数据，例如分子结构和蛋白质序列，并用自然语言查询这些数据实例。基于与 MFM 的对话，它将激发研究人员的灵感，提高发现新型分子结构、阐明蛋白质功能和推进药物研究和开发的效率。

4) 医疗咨询: 除了上述为专业医生或研究人员所构建的聊天机器人外，MFM 也可以应用为医疗咨询的患者聊天机器人。由于医疗健康信息通常分散且对没有医学背景的人难以理解，MFM 将为患者提供初步的医疗咨询。一些 MFM，例如 Qilin-Med-VL [260]、XrayGPT [258]、LLaVA-Med [252] 和 Med-Flamingo [251]，具有文本和视觉的理解能力，可以基于患者输入的查询信息和图像提供初步的诊断和治疗建议，引导患者寻求治疗。由于医疗聊天机器人可以进行持续的对话，能够帮助患者更好地了解自己的病情，并对其日常生活、运动计划、饮食习惯和其他生活方式因素做出必要的调整，对于患有慢性疾病的患者展现出巨大的潜力。然而，医疗聊天机器人的应用仍然具有挑战性，因为一个错误的建议可能会对缺乏医学知识的患者造成危险。我们将在第V节中讨论这一挑战。

V. 挑战

如图4所示，数据、算法和计算基础设施作为 AI 的三大支柱 [379]，为 HFM 的产生提供了契机，但由于其目前存在的不足，依然是 HFM 所面临的各种挑战的根源。

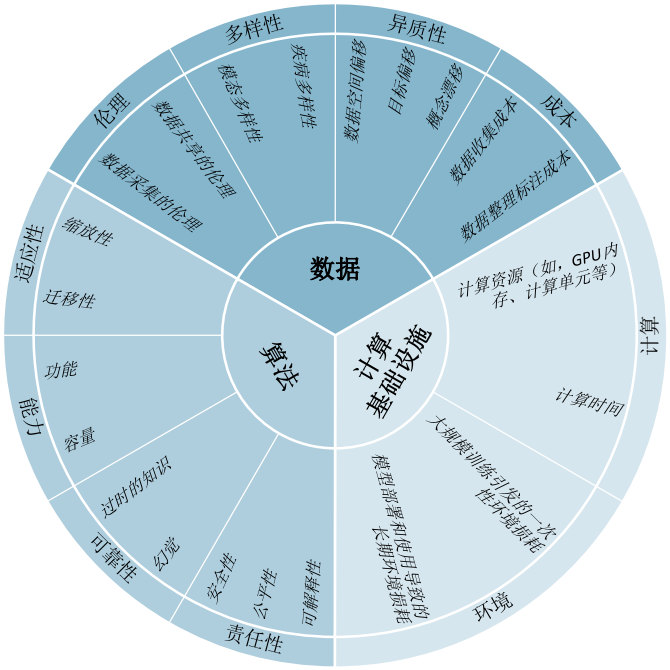

图 4. 医疗基础模型在数据、算法和计算基础设施方面的挑战。

A. 数据

数据缺乏是 HFM 面临的核心挑战。基础模型的通用能力依赖于对大规模、多样化数据集的学习 [9]。然而，医疗健康数据中存在的伦理、多样性、异质性、高成本等固有问题，阻碍了大规模数据集的构建，并且也带来了伦理、社会和经济上的挑战。因此，“如何构建大规模医疗健康数据集” [33] 是我们必须回答的第一个问题，以解决 HFM 所面临的挑战。

1) 伦理: 医疗健康数据的伦理问题使构建大型数据集的一个至关重要的挑战 [380]。具体而言，**a)** 医疗健康数据的获取必须符合伦理要求。医疗健康数据是从人体扫描获取的，而某些扫描协议或模态会对人体造成伤害，例如 CT 成像数据 [381]。尽管这些伤害对疾病的治疗可能并不重要，但为了构建 AI 训练集而扫描人体是不道德的。因此，这些特殊数据将无法像一些现有的数据收集方法 [20] 那样通过主动采集用于构建大型数据集，从而阻碍了某些 HFM 任务的训练。**b)** 医疗健康数据的使用和共享也受到伦理限制 [380]。医疗健康数据包含大量来自人体的私人信息，这些信息是敏感和有风险的，例如基因。这些数据的使用和共享受到法律和数据所有者的严格限制。一旦在没有治理的情况下大规模收集并用于基础模型的训练，将会很危险。在 HFM 的进一步应用中，不可控的外部环境也会放大这种风险。例如，语言模型可能会泄露或滥用敏感的医疗数据，如个人健康记录、测试结果、基因信息等。因此，这加剧了数据挑战，使得 HFM 和它们的应用的数据集构建面临着重重障碍。尽管社区已经做出了一些初步的努力 [380], [382]，但仍有很长的路要走。

2) 多样性: 由于医疗健康数据处于长尾分布 [57]，数据的多样性已成为 HFM 中的另一个重要挑战。**a)** 医疗实践的不断进步使得不同模态的医疗数据呈现出长尾分布。例如，在医学影像方面，虽然广泛可用的胸部 X 线和胸部 CT 图像可用于一般胸部疾病的诊断，但其他像光学相干断层扫描（OCT）、数字减影血管造影（DSA）、正电子发射断层扫描（PET）等特定临床任务所需的成像模态却很少且昂贵。这使得这些任务特定的模态在大型数据集中相对于普通模态而言很少，限制了训练 HFM 的泛化能力。**b)** 一些疾病的发生也呈现出长尾分布 [383]，因此这些疾病的图像、生物信息数据或文本记录很少。这意味着许多极为罕见的疾病数据在 HFM 的训练数据集中无法被覆盖或者数量稀少，限制了 HFM 在这些疾病上的泛化能力。因此，许多重要但专业化的任务由于相关数据的多样性有限而无法被 HFM 处理，限制了这些应用的潜在范围，成为其通用化道路上的新障碍。

3) 异质性: 医疗健康数据的特征在不同人群、地区和医疗中心间存在差异，这使得 HFM 在实际应用中的数据具有异质性 [3], [384]。因此，在 HFM 的测试数据和训练数据之间存在潜在的分布不匹配问题 [385]。**a)** 数据空间偏移是一个紧迫的问题。数据获取协议的变化、传感器配置的变化等会导致收集到的医疗健康数据分布的变化。因此，针对原始数据训练的 HFM 在变化的情况下难以适应新的数据形式。**b)** 当数据从不同的受试者、背景、人群等中采样时，结果的异质性也会出现，从而在 HFM 中出现目标偏移。不同疾病的发病率会随着个人特征和行为方式的改变而改变 [386]，如遗传疾病、吸烟、饮食习惯等，这将极大地限制 HFM 在个性化和精确医疗方面的能力。**c)** 从长期的角度来看，概念漂移 [387] 是数据异质性的另一个重要因素。随着医疗领域的发展，一些新的概念将出现，一些错误的概念将被纠正。因此，HFM 很难跟随这些概念的变化，让输入和输出之间对应关系随之变化。

4) 成本: 长期以来，数据成本一直是医疗健康 AI 面临的重要挑战，而基础模型对大规模数据的依赖进一步加大了 HFM 中的这一挑战 [10]。**a)** 在数据收集 [33]

方面，由于某些医疗健康数据模态的扫描方法专业和扫描设备昂贵，导致其获取成本极高。例如，在美国，CT 扫描的价格可以从 300 美元到 6750 美元不等[2]。因此，这使得为了基础模型的训练而构建大规模的医疗健康数据集变得即为昂贵，产生难以想象的成本，使得一些机构难以独立实施。**b)** 尽管许多 HFM 工作侧重于自监督学习而无需注释，但组织整理大规模数据集仍需要大量的专业人力，使其这成为了另一个重要挑战 [388], [389]。医疗健康数据的专业性要求熟练的专业人员进行数据的过滤或标注。这使得利用众包 [20] 进行医疗健康数据集的标注和整理过程变得不切实际。因此，在这些重复任务上花费大量专业时间是低效且昂贵的。

B. 算法

尽管算法已经随着 AI 的发展研究了几十年 [2]，但在基础模型时代 [9]，前所未有的数据量、模型规模和应用范围暴露出了算法面临的新挑战。在这里，我们分析了医疗健康中四个最重要的算法挑战，包括责任性、可靠性、能力和适应性。

1) 责任性: 基础模型的问责问题仍然是一个重要的关注点，特别是由于医疗健康与人类生命的密切关系，使其在 HFM 中变得尤为重要 [390]。**a)** 其中最重要的方面之一是可解释性。由于神经网络的“黑盒”属性 [391] 和更大量的隐藏层神经元数量，对 HFM 的行为的解释变得极其困难 [392]。因此，医疗健康专家将难以理解 HFM 作出回答的基础，对伦理和安全问题有着重要的担忧。**b)** 公平性 [393] 是责任性的另一个方面。由于训练数据集中存在分布偏差，HFM 可能会受到数据集固有偏见的影响，破坏结果的公平性。一些工作已经揭示了 LFM 中普遍存在的偏见和刻板印象 [394], [395]。这是危险的，因为不公平的预测可能会增加潜在的歧视，并破坏医疗健康中人类生命的平等性，引发潜在的社会冲突。**c)** 安全性也是一个重要的问题。一些 LFM 已经被记录生成仇恨言论 [396]，导致冒犯和心理上的伤害，并在极端情况下煽动暴力。这对 HFM 的用户来说非常危险，成为潜在的社会不稳定因素。一些越狱攻击 [397] 甚至会导致 LFM 的输出包含私人和敏感信息，对数据提供者构成威胁。尽管已经有一些关于负责任的 AI 的研究 [398]，但基础模型的前所未有的大规模和应用范围使得这些技术变得更加具有挑战性 [399]。

[2]https://www.newchoicehealth.com/ct-scan/cost

2) 可靠性: 可靠性在医疗健康中尤为关键 [400]，这对 HFM 的提出了巨大的可靠性要求，使其成为了巨大的挑战。**a)** 基础模型中的幻觉 [401] 问题越来越受到关注。模型容易输出基于非事实或不准确信息的内容。例如，在与 LFM 进行医疗对话时，模型提供与事实相矛盾的临床知识或结论 [376]。这引发了人们对使用 HFM 得出的结论可靠性的担忧。**b)** 另一个可靠性挑战来自于过时的知识。正如在数据异构性V-A3中所讨论的，医疗健康领域的发展将构建一些新知识，并纠正一些错误。因此，一旦基础模型落后于该领域的发展，它们可能潜在地基于过时的知识误导使用者 [402]。尽管一些现有的努力试图开发模型编辑技术来更新模型的具体行为或所学习到的知识 [403]，但这是昂贵的且缺乏针对性的，可能会导致意想不到的副作用 [404]。因此这仍然是 HFM 长期可靠性的一个重要算法挑战。

3) 能力: HFM 的能力决定了它们在应用中的表现，因此这是最受关注的挑战之一。**a)** 基础模型的容量 [405] 近年来一直是研究人员的关注重点。它指的是模型表示和记忆大量知识的能力 [9]。一些先进的网络架构，如 ViT [406] 和 Swin Transformer [407]，试图构建一个大容量的骨干网络，使得基础模型能够从大规模数据集中学习大量的知识。然而，增加模型容量也会增加计算和内存，导致成本和碳排放量增加 [12]。特别是对于医疗健康领域，其数据可能非常庞大，如 3D CT 体积，因此如何在较少的计算消耗下提高 HFM 的容量仍然是一个长期的问题。**b)** HFM 的功能仍然显得单调，很难满足一些复杂的临床需求。例如，慢性病管理将涉及多个部门和各种模态的信息 [408]，并需要实施多种临床程序，如诊断、干预、预后等。尽管一些通才 HFM 在各种临床环境中表现出了很好的性能 [10], [31]，但它们仍然难以满足这种复杂的多模态、多任务需求。

4) 适应性: 基础模型的适应性决定了它们适应下游场景的能力，这仍然是 HFM 中的一个重要挑战。**a)** 一个方面是模型对下游任务的迁移能力。现有的 HFM 仍然难以解决现实世界中的数据异构性（V-A3）问题 [384]，在某些非常特定的领域表现不佳 [43]。尽管一些方法利用 FT 或 AT 将基础模型应用于下游领域，但它们的适应性仍然受到原始预训练模型的限制，构建服务于下游任务的适应数据集仍然需要花费较大的成本 [42], [161]。因此，如何有效地激发 HFM 在现实场景中的潜在能力仍然是一个紧迫的问题。**b)** 另一个方面是 HFM 在下游设备上的可扩展性。在一些潜在的资源有

限的临床场景中，例如可穿戴医疗设备 [409]，原始的非常大基础模型无法直接部署。因此，HFM 需要有效的缩放方法来适应潜在的操作环境。尽管已经研究了一些关于 AI 模型的模型压缩和加速技术 [410]，但缩小这样的大型基础模型并保持其通用性仍然具有挑战性。

C. 计算基础设施

基础模型的大规模属性（包括参数量和数据量）[12] 使得训练和推理过程需要前所未有的大规模计算基础设施，由此带来了新的挑战。在这里，我们分析了两个最重要的计算基础设施挑战，包括计算和环境。

1) 计算: 在训练或者适应 HFM 时，模型训练所需的时间和计算资源成本过高，超出了大多数研究人员和组织的预算 [411]。**a)** 基础模型的大量参数需要大量的计算资源，如 GPU 内存和计算单元，进行训练或适应，这在大多数医院和机构中几乎不可行。举个例子，直接微调 GPT-3 需要训练大约 175,255 亿个参数 [411]，将导致巨大的计算开销。**b)** 在大规模数据集上进行 HFM 训练也会消耗大量的计算时间。例如，LLaMA 使用 2048 个 A100 GPU 和 80GB 的 RAM 花费了大约 21 天的时间，才实现从 1.4T 的 token 中训练 65B 参数的模型 [79]。这大大延长了 HFM 产品的开发周期，并极大地增加了试错的时间成本，最终增加了软件构建的风险。研究人员迫切需要先进的 GPU 设备来推进 HFM 的探索，但当前 GPU 芯片仍然严重短缺 [412]。

2) 环境: 开发如此大规模的基础模型会产生相当大的环境成本 [36]。**a)** 由于 HFM 的规模极大，它们在大规模数据上的训练将消耗大量的电力，从而产生重大的一次性环境成本。一项研究估计，基于 BERT 的模型一次训练所产生的碳排放需要通过种植 40 棵树花费 10 年时间来抵消 [9]。增加的碳排放 [413] 将对环境产生负面影响。**b)** 在广泛的部署和推理运行过程中，这样巨大的模型将进一步带来巨大的长期环境成本 [414]。因此，在基础模型的构建和部署中减少环境影响，促进可持续发展已成为一个重要需求。然而，相关技术和政策仍然滞后。

VI. 未来方向

HFM 的发展代表了模型应用从具体任务到通用任务的进步 [10]。它使得 AI 具备了更通用的能力，能够应对现实世界中各种需求和复杂环境。如图5所示，我们探索了 HFM 在角色、实施方式、应用和关注点的四个未来方向。

A. 超越人工智能替代人类的模式

尽管传统范式侧重于利用人工智能自动完成某些医疗健康任务，从而取代人工手动工作 [7], [8]，但在 HFM 时代，AI-人类协作的设置 [5] 展示了它们的新的机会和实际应用价值。特别是，在 HFM 中，AI-人类协作有三个重要的目标。

1) 提高医疗健康能力: 它的目标是让人工智能与医生扮演协作角色，而不是取代人类，从而激发人们在人工智能的支持下完成更具挑战性的医疗健康任务。这创造了一种协作过程，在这个过程中，人工智能协助人类快速完成复杂任务中繁琐和耗时的部分，而人类提供专业判断并纠正人工智能可能犯的错误。因此，它将使人类和人工智能能够高效准确地完成更具挑战性的医疗健康任务。与传统的独立人工智能范式和单个人类专家相比，AI-人类协作具有更强大的能力 [415], [416]，在具有挑战性的医疗任务中展现出了潜力。

2) 符合医疗健康需求: 它的目标是使人工智能在人类专家的监督下运行，从而满足实际医疗健康实践的责任性要求 [398]。由于医疗健康实践与人类生命密切相关，因此对于与健康相关的事件需要有问责机制。尽管人工智能已经在某些临床场景中展示了巨大的潜力 [6], [8]，甚至超越了人类专家，但在现实生活中应用仍然很困难。因为在一些临床突发事件发生时，独立的人工智能模型是无法对其负责。而与人类专家的协作将增加了人们对人工智能决策的信任，使得人工智能在临床场景中拥有更大的应用机会。潜在的临床突发事件也将实现问责机制，为患者提供更多的法律保障。

3) 优化合作过程: 它的目标是通过 HFM 的支持来改善协作方法。基于人类的提示，HFM 的涌现能力将能够输出从大规模数据中学习到的广泛而合理的反馈 [108]。因此，设计协作方法以有效地挖掘 HFM 内部的广泛知识并减少人类交互成本非常重要。其中一个重要的未来工作是人工智能和人类之间的分工。一些研究发现，分配不当会导致 AI-人类协作的表现比独立的人工智能更差 [417]，而相比于高级医生，初级医生可以从人工智能中获益更多 [418]。因此，如何在协作中制定合理的任务分配规则以最大化医疗健康任务的性能仍不清楚。提示的设计是另一个未来的工作。仍有许多医疗健康场景缺乏有效的提示或交互方法。一些提示策略，如 SAM 中的点或边界框 [20]，在用户必须保持专注的临床场景中（如手术期间）很难应用。因此，设计更多适应场景的提示以实现更高效的交互非常重要。

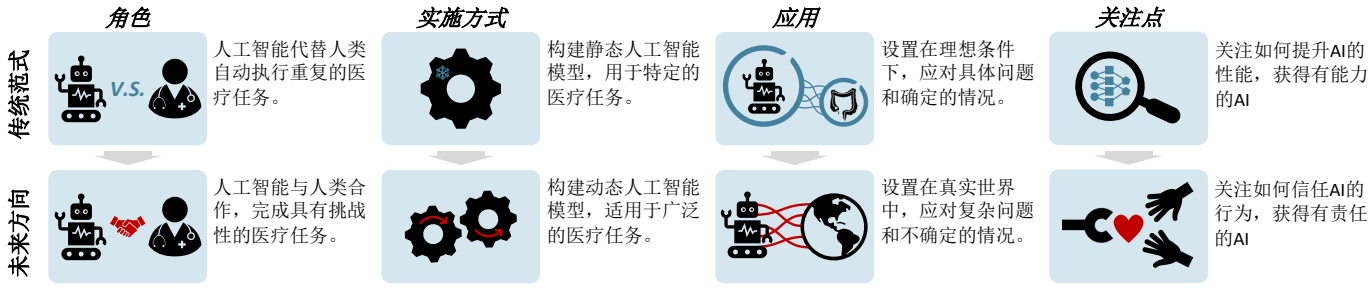

图 5. 医疗健康基础模型的未来方向，本文中，我们将讨论未来 HFM 在角色、实施方式、应用和关注点方面的转变。

B. 从静态模型到动态模型

尽管静态人工智能模型已经在特定的医疗健康任务中表现出了有效性，但现实世界的医疗健康实践总是需要协调来自多个部门的不同临床要求和多样化的数据模态。因此，构建动态人工智能模型是 HFM 中一个重要的未来方向 [419]。具体来说，

1) 表征能力: 由于现实世界中数据分布的变化，HFM 的内在表征能力对它们适应医疗健康场景非常重要。因此需要构建强大的动态神经网络结构，例如最近非常著名的注意力机制（例如 Transformer）[420]、专家混合模型（例如 MoE）[421], [422]、选择性状态空间模型（例如 Mamba）[423] 等，以表征更广泛的数据分布。它将使 HFM 能够动态适应各种数据分布和临床情况 [419]，推动 HFM 在医疗健康应用中实现复杂的设计。此外，为了解决幻觉和过时知识可能带来的挑战，进一步研究持续学习和模型编辑 [403], [424] 从而更新模型表征也非常重要。这将有助于 HFM 在其生命周期内保持长期的有效性。

2) 任务适应能力: 在医疗健康实践中，为了将模型应用于各种医疗健康场景和任务，HFM 的任务适应能力非常重要。其中一个重要方面是降低适应成本，使用更少的数据和计算来提高基础模型的灵活性 [388], [389], [411]。一旦成功，用户将更容易将这些模型应用于他们的任务中，这对于 HFM 获得更广泛的适用性至关重要。另一方面，设计更有效的方法提高 HFM 的涌现能力 [9]，以激发从大规模数据中学习到的丰富知识 [108] 非常重要。现有的研究已经表明，无需进行任何额外的训练 [425]，精心设计的提示模板即可显著提高 LFM 模型在中目标任务上的性能。然而，在其他子领域中，更强大和灵活的提示方法仍然是紧迫的需求。

3) 缩放能力: 正如在V-B3中讨论的那样，下游设备的可扩展性对于在资源有限的临床场景中部署 HFM 非常重要。特别是，在医疗中心已经拥有大量昂贵但计算受限的设备的情况下，将 HFM 运行在这些设备上成为了一个巨大的挑战。因此，开发类似于学习基因 [426] 的针对基础模型的有效的缩放方法非常重要，从而动态适应计算环境并在这些设备上实现高效推理。

C. 从理想设置到复杂的真实世界

以往的医疗健康人工智能应用 [3]–[5] 是在特定问题和一些确定的理想条件下运行的，无法应对现实世界的复杂性和不确定性。因此，探索 HFM 在实际医疗健康实践中的应用已成为一个重要的未来方向。

1) 从单域到多域: 正如在V-A3中讨论的那样，由于人口、地区和医疗中心的差异，医疗健康数据面临着严重的异质性挑战，即 “域” [384]。因此，HFM 必须拥有学习和推广到多个域的能力，以实现广泛的应用。尽管一些域适应 [384] 和域泛化 [256] 算法已经被研究以帮助模型适应数据异质性，但是他们在基础模型上的研究仍让需要进一步推进。同时，它们在实际医疗健康应用中的有效性仍需要验证。此外，由于医疗健康数据的隐私性，越来越多研究关注用于 HFM [427], [428] 的联邦学习，从而构建隐私保护的大规模跨域学习系统。这些研究采用分布式训练机制，使 HFM 能够在受保护的情况下学习多个域的医疗健康数据，避免了隐私泄露的风险。然而，对于基础模型来说，进行如此大规模的分布式训练（大参数和数据量）是极具挑战性的。

2) 从单/封闭任务到多/开放任务: 由于医疗场景的多样性，实际应用中需要 HFM [10] 能够满足广泛的医疗健康任务。与为单一任务设计的传统范式不同，HFM 面临着具有多个任务的医疗健康场景，例如不同的器官、疾病、临床目标等。因此，HFM 需要在不同的医疗健康场景中获得更强大的多任务能力。一些动态模型技术，例如 MoE [421]，已经证明了它们在基础模

型的多任务预测中的有效性，展现出一条迈向通用人工智能的潜在途径 [429]。另一方面，现实世界的不确定性进一步引入了 HFM 的开放集问题 [430]。由于在医疗健康任务上，模型的不负责任的预测可能会危及人们生命安全，因此该问题尤为重要。对于一些潜在的未知的输入，HFM 必须建立机制来处理超出其能力的要求，以获得合理和安全的医疗健康预测结果 [401]。

3) 从单模态到多模态: 在现实世界中，医疗健康场景往往同时运用多种模态 [3]，因此为在多模态设置下构建统一的 HFM 提出了巨大需求。与传统的单模态范式相比，多模态设置将融合来自不同模态的具有代表性的特征，从而使模型能够输出精确可靠的结果 [30]。正如在II-D中讨论的那样，虽然 HFM 在多模态设置中取得了初步的成功，但大部分的努力仍然集中在语言和视觉模态上。因此，如何将更多的模态整合到实际的医疗健康实践中，仍然是一个开放的问题。此外，多模态数据的学习算法已经成为一个研究热点。一些现有的研究尝试通过跨模态生成 [378]、跨模态自监督学习 [48]、多模态知识蒸馏 [431] 等方式来刺激模态之间的学习能力。然而，当更多的模态被纳入大规模 HFM 的学习时，如何利用不同模态的优势和互补性仍然是一个长期的问题。在多模态设置中，缺失模态的挑战 [432] 也引起了关注。因为在真实世界中，案例之间处于不同的模态情况，例如对于不同的患者，由于其治疗计划不同，数据也将从不同的模态组合情况下被采集。因此，多模态 HFM 难以通过完整的数据模态以进行推理，需要适应不同的模态组合。

D. 从关注能力到关注信任

随着基础模型在医疗健康中的发展，人工智能的角色、实施方式和应用也将不断发展，因此人们的关注点也将随之从探索它们的能力转向信任它们的行为。正如在第V节中讨论的那样，在医疗健康中信任并深入地应用 HFM 仍然具有挑战性。在这里，我们讨论三个重要方面。

1) 可解释地 HFM: 如第V-B1节所讨论地，神经网络的“黑匣子”属性使人们难以理解它们的行为。因此，在医疗健康中，解释 HFM 结果背后的内在意图对于人们信任其行为至关重要 [391]。未来的任务之一是促进基础模型的机器学习理论研究 [433]。通过分析基础模型的机器学习属性能够揭示它们的独特行为模式，从而为研究人员设计更加可理解的模型并提高研究效率提供指导。然而，由于 HFM 的巨大参数和数据量极大地超出了机器学习理论的理想设置，现有的表征、优化或泛化理论仍然难以适用于 HFM [434]。另一个有前途的方向是发现更有效的证据。尽管现有的工作利用了一些热力图 [391]，包括注意力图、类激活图、不确定性图等，用于解释个体行为，但仍亟需进一步探索更高层次的解释证据来帮助人们理解模型地行为。此外，进一步利用可解释的 HFM 进行科学探索，如药物发现 [435]，也是一个具有潜力的未来方向。

2) 安全的 HFM: HFM 的安全性是人们信任和使用它们的基础，因此这也是重要的未来方向之一 [436]。安全性的一个方面是 HFM 抵御外部攻击的能力。例如，一些越狱攻击 [397] 会导致 LFM 输出一些隐私和敏感的信息，对数据提供者构成威胁。因此，应建立防御机制，如对抗攻击 [437]，以应对潜在的恶意用户，从而在生命周期内保护 HFM 安全。另一个方面是 HFM 本身的安全性。在医疗健康任务中，只有安全的输出值得信任，因为它们与人类生命的关系密切。因此，除了构建探索方法 [391] 来衡量可靠性之外，还需要引入更多的数据和设计更强大的方法来增强鲁棒性和准确性。最后，应在 HFM 的应用过程中建立合理的问责机制 [438]，以增强用户的使用过程中谨慎性和医疗健康实践中的法律安全性。

3) 可持续的 HFM: HFM 需要采用一种可持续的方式来进一步地深入发展 [36], [439]。能源和环境危机正伴随着基础模型的一起发展。大规模的训练会导致大量的能源消耗和碳排放，使得基于环境代价的 HFM 的发展变得不可持续 [440]。因此，迫切需要研究低功耗的基础模型训练和部署策略，包括构建更环保的芯片技术和模型架构。随着 HFM 应用范围的扩大，成本问题也是限制可持续性的另一个重点（在第V-A4节和第V-C1节中讨论）。因此，进一步研究高效的学习算法 [441] 和硬件设施对于 HFM 未来发展至关重要。减少成本，包括数据收集和处理 [33], [388], [389]，模型训练和推断 [411]，将刺激基础模型的商业优势，从而增强它们的可持续性。

VII. 结论

本文对于“我们能否构建 AI 模型来惠及多种医疗健康任务?”这个问题提供了一个潜在的答案。更多的医疗健康实践将从 HFM 的发展中受益，实现先进的智能医疗健康服务。尽管 HFM 正在逐渐展示其巨大的应

用价值，但人们仍缺乏对于基础模型在医疗健康实践中所面临的挑战、新机遇以及潜在未来方向的清晰认识。因此，本文首先对 HFM 进行了全面的概述和分析，包括方法、数据和应用，有助于理解 HFM 的当前进展。然后，我们对数据、算法和计算基础设施中的关键挑战进行了深入讨论，阐明了 HFM 的不足之处。最后，我们展望关于角色、实施方式、应用和关注点的未来方向，展望了未来 HFM 在推进医疗健康方面的巨大前景。

致谢

这项工作得到香港创新及科技基金（项目编号为 MHP/002/22 和 PRP/034/22FX）、深圳市科技创新委员会基金（项目编号为 SGDX20210823103201011）、香港特别行政区肺尘埃沉着病补偿基金委员会（项目编号为 PCFB22EG01）和中国香港特别行政区研究资助局（项目编号为 R6003-22 和 C4024-22GF）的支持。

参考文献

[1] N. J. Nilsson, *Principles of artificial intelligence*. Springer Science & Business Media, 1982.

[2] Y. LeCun *et al.*, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[3] X. Gu *et al.*, "Beyond supervised learning for pervasive healthcare," *IEEE Rev. Biomed. Eng.*, 2023.

[4] F. Jiang *et al.*, "Artificial intelligence in healthcare: past, present and future," *Stroke and vascular neurology*, vol. 2, no. 4, 2017.

[5] P. Rajpurkar *et al.*, "Ai in health and medicine," *Nat. Med.*, vol. 28, no. 1, pp. 31–38, 2022.

[6] K. Cao *et al.*, "Large-scale pancreatic cancer detection via non-contrast ct and deep learning," *Nat. Med.*, pp. 1–11, 2023.

[7] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nat. Med.*, vol. 24, no. 9, pp. 1342–1350, 2018.

[8] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[9] R. Bommasani *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[10] M. Moor *et al.*, "Foundation models for generalist medical artificial intelligence," *Nature*, vol. 616, no. 7956, pp. 259–265, 2023.

[11] B. Azad *et al.*, "Foundational models in medical imaging: A comprehensive survey and future vision," *arXiv preprint arXiv:2310.18689*, 2023.

[12] J. Qiu *et al.*, "Large ai models in health informatics: Applications, challenges, and the future," *IEEE J. Biomed. Health Inform.*, 2023.

[13] A. J. Thirunavukarasu *et al.*, "Large language models in medicine," *Nat. Med.*, vol. 29, no. 8, pp. 1930–1940, 2023.

[14] K. He *et al.*, "A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics," *arXiv preprint arXiv:2310.05694*, 2023.

[15] X. Yang *et al.*, "A large language model for electronic health records," *NPJ Digit. Med.*, vol. 5, no. 1, p. 194, 2022.

[16] K. Singhal *et al.*, "Large language models encode clinical knowledge," *Nature*, vol. 620, no. 7972, pp. 172–180, 2023.

[17] Z. Li *et al.*, "D-lmbmap: a fully automated deep-learning pipeline for whole-brain profiling of neural circuitry," *Nat. Methods*, vol. 20, no. 10, pp. 1593–1604, 2023.

[18] Z. Wang *et al.*, "Foundation model for endoscopy video analysis via large-scale self-supervised pre-train," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2023, pp. 101–111.

[19] Y. Zhou *et al.*, "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, no. 7981, pp. 156–163, 2023.

[20] A. Kirillov *et al.*, "Segment anything," in *Proc. IEEE Int. Conf. Comput. Vis.*, October 2023, pp. 4015–4026.

[21] R. Rombach *et al.*, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 10 684–10 695.

[22] X. Shen and X. Li, "Omnina: A foundation model for nucleotide sequences," *bioRxiv*, pp. 2024–01, 2024.

[23] H. Dalla-Torre *et al.*, "The nucleotide transformer: Building and evaluating robust foundation models for human genomics," *bioRxiv*, pp. 2023–01, 2023.

[24] N. Brandes *et al.*, "Proteinbert: a universal deep-learning model of protein sequence and function," *Bioinformatics*, vol. 38, no. 8, pp. 2102–2110, 2022.

[25] J. Jumper *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021. [Online]. Available: https://doi.org/10.1038/s41586-021-03819-2

[26] J. Chen *et al.*, "Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions," *bioRxiv*, pp. 2022–08, 2022.

[27] C. Wu *et al.*, "Towards generalist foundation model for radiology," *arXiv preprint arXiv:2308.02463*, 2023.

[28] N. Fei *et al.*, "Towards artificial general intelligence via a multimodal foundation model," *Nat. Commun.*, vol. 13, no. 1, p. 3094, 2022.

[29] K. Zhang *et al.*, "Biomedgpt: A unified and generalist biomedical generative pre-trained transformer for vision, language, and multimodal tasks," *arXiv preprint arXiv:2305.17100*, 2023.

[30] J. N. Acosta *et al.*, "Multimodal biomedical ai," *Nat. Med.*, vol. 28, no. 9, pp. 1773–1784, 2022.

[31] T. Tu *et al.*, "Towards generalist biomedical ai," *NEJM AI*, vol. 1, no. 3, p. AIoa2300138, 2024.

[32] P. Shrestha *et al.*, "Medical vision language pretraining: A survey," *arXiv preprint arXiv:2312.06224*, 2023.

[33] M. J. Willemink *et al.*, "Preparing medical imaging data for machine learning," *Radiology*, vol. 295, no. 1, pp. 4–15, 2020.

[34] J. J. Hatherley, "Limits of trust in medical ai," *Journal of medical ethics*, 2020.

[35] A. F. Markus *et al.*, "The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies," *Journal of biomedical informatics*, vol. 113, p. 103655, 2021.

[36] C.-J. Wu *et al.*, "Sustainable ai: Environmental implications, challenges and opportunities," *Proceedings of Machine Learning and Systems*, vol. 4, pp. 795–813, 2022.

[37] J. Devlin *et al.*, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

[38] J. Lee *et al.*, "Biobert: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 09 2019. [Online]. Available: https://doi.org/10.1093/bioinformatics/btz682

[39] S. B. Patel and K. Lam, "Chatgpt: the future of discharge summaries?" *The Lancet Digital Health*, vol. 5, no. 3, pp. e107–e108, 2023.

[40] Y. He *et al.*, "Geometric visual similarity learning in 3d medical image self-supervised pre-training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 9538–9547.

[41] Z. Zhou *et al.*, "Models genesis," *Med. Image Anal.*, vol. 67, p. 101840, 2021.

[42] J. Ma *et al.*, "Segment anything in medical images," *Nat. Commun.*, vol. 15, no. 1, p. 654, 2024.

[43] M. A. Mazurowski *et al.*, "Segment anything model for medical image analysis: an experimental study," *Med. Image Anal.*, vol. 89, p. 102918, 2023.

[44] M. Baharoon *et al.*, "Towards general purpose vision foundation models for medical image analysis: An experimental study of dinov2 on radiology benchmarks," *arXiv preprint arXiv:2312.02366*, 2023.

[45] X. Wang *et al.*, "Uni-rna: universal pre-trained models revolutionize rna research," *bioRxiv*, pp. 2023–07, 2023.

[46] Y. Ji *et al.*, "Dnabert: pre-trained bidirectional encoder representations from transformers model for dna-language in genome," *Bioinformatics*, vol. 37, no. 15, pp. 2112–2120, 8 2021. [Online]. Available: https://doi.org/10.1093/bioinformatics/btab083

[47] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2021, pp. 8748–8763.

[48] Z. Zhao *et al.*, "Clip in medical imaging: A comprehensive survey," *arXiv preprint arXiv:2312.07353*, 2023.

[49] B. Wang *et al.*, "Pre-trained language models in biomedical domain: A systematic survey," *ACM Computing Surveys*, vol. 56, no. 3, pp. 1–52, 2023.

[50] H. Zhou, B. Gu, X. Zou, Y. Li, S. S. Chen, P. Zhou, J. Liu, Y. Hua, C. Mao, X. Wu *et al.*, "A survey of large language models in medicine: Progress, application, and challenge," *arXiv preprint arXiv:2311.05112*, 2023.

[51] M. Yuan *et al.*, "Large language models illuminate a progressive pathway to artificial healthcare assistant: A review," *arXiv preprint arXiv:2311.01918*, 2023.

[52] H. H. Lee *et al.*, "Foundation models for biomedical image segmentation: A survey," *arXiv preprint arXiv:2401.07654*, 2024.

[53] Q. Li *et al.*, "Progress and opportunities of foundation models in bioinformatics," *arXiv preprint arXiv:2402.04286*, 2024.

[54] J. Liu *et al.*, "Large language models in bioinformatics: applications and perspectives," *arXiv preprint arXiv:2401.04155*, 2024.

[55] Y. Qiu *et al.*, "Pre-training in medical data: A survey," *Machine Intelligence Research*, vol. 20, no. 2, pp. 147–179, 2023.

[56] D.-Q. Wang *et al.*, "Accelerating the integration of chatgpt and other large-scale ai models into biomedical research and healthcare," *MedComm–Future Medicine*, vol. 2, no. 2, p. e43, 2023.

[57] S. Zhang and D. Metaxas, "On the challenges and perspectives of foundation models for medical image analysis," *Med. Image Anal.*, vol. 91, p. 102996, 2024.

[58] Y. Zhang *et al.*, "Data-centric foundation models in computational healthcare: A survey," *arXiv preprint arXiv:2401.02458*, 2024.

[59] A. Radford *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[60] Q. Jin *et al.*, "Medcpt: Contrastive pre-trained transformers with large-scale pubmed search logs for zero-shot biomedical information retrieval," *Bioinformatics*, vol. 39, no. 11, p. btad651, 2023.

[61] C. Peng *et al.*, "A study of generative large language model for medical research and healthcare," *NPJ Digit. Med.*, vol. 6, 2023. [Online]. Available: https://doi.org/10.1038/s41746-023-00958-w

[62] Y. Gu *et al.*, "Domain-specific language model pretraining for biomedical natural language processing," *ACM Transactions on Computing for Healthcare*, vol. 3, no. 1, pp. 1–23, 2021.

[63] H. Wang *et al.*, "Huatuo: Tuning llama model with chinese medical knowledge," *arXiv preprint arXiv:2304.06975*, 2023.

[64] H. Zhang *et al.*, "Huatuogpt, towards taming language model to be a doctor," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, pp. 10 859–10 885.

[65] C. Wu *et al.*, "Pmc-llama: Towards building open-source language models for medicine," *arXiv preprint arXiv:2305.10415*, vol. 6, 2023.

[66] J. Chen *et al.*, "Huatuogpt-ii, one-stage training for medical adaption of llms," *arXiv preprint arXiv:2311.09774*, 2023.

[67] Z. Chen *et al.*, "Meditron-70b: Scaling medical pretraining for large language models," *arXiv preprint arXiv:2311.16079*, 2023.

[68] X. Zhang *et al.*, "Alpacare: Instruction-tuned large language models for medical application," *arXiv preprint arXiv:2310.14558*, 2023.

[69] Y. Chen *et al.*, "Bianque: Balancing the questioning and suggestion ability of health llms with multi-turn health conversations polished by chatgpt," *arXiv preprint arXiv:2310.15896*, 2023.

[70] Y. Li *et al.*, "Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge," *Cureus*, vol. 15, no. 6, 2023.

[71] T. Han *et al.*, "Medalpaca–an open-source collection of medical conversational ai models and training data," *arXiv preprint arXiv:2304.08247*, 2023.

[72] Q. Ye *et al.*, "Qilin-med: Multi-stage knowledge injection advanced medical large language model," *arXiv preprint arXiv:2310.09089*, 2023.

[73] L. Luo *et al.*, "Taiyi: a bilingual fine-tuned large language model for diverse biomedical tasks," *Journal of the American Medical Informatics Association*, p. ocae037, 02 2024.

[74] W. Wang *et al.*, "Gpt-doctor: Customizing large language models for medical consultation," *arXiv preprint arXiv:2312.10225*, 2023.

[75] H. Xiong *et al.*, "Doctorglm: Fine-tuning your chinese doctor is not a herculean task," *arXiv preprint arXiv:2304.01097*, 2023.

[76] G. Wang *et al.*, "Clinicalgpt: Large language models finetuned with diverse medical data and comprehensive evaluation," *arXiv preprint arXiv:2306.09968*, 2023.

[77] Q. Li *et al.*, "From beginner to expert: Modeling medical knowledge into general llms," *arXiv preprint arXiv:2312.01040*, 2023.

[78] Y. Labrak *et al.*, "Biomistral: A collection of open-source pre-trained large language models for medical domains," *arXiv preprint arXiv:2402.10373*, 2024.

[79] H. Touvron *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[80] E. Alsentzer *et al.*, "Publicly available clinical BERT embeddings," in *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Minneapolis, Minnesota, USA: Association for Computational Linguistics, Jun. 2019, pp. 72–78.

[81] Y.-P. Chen *et al.*, "Modified bidirectional encoder representations from transformers extractive summarization model for hospital information systems based on character-level tokens (alphabert): development and performance evaluation," *JMIR medical informatics*, vol. 8, no. 4, p. e17787, 2020.

[82] Y. Li *et al.*, "Behrt: transformer for electronic health records," *Scientific reports*, vol. 10, no. 1, p. 7155, 2020.

[83] H. Yuan *et al.*, "BioBART: Pretraining and evaluation of a biomedical generative language model," in *Proceedings of the 21st Workshop on Biomedical Language Processing*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 97–109. [Online]. Available: https://aclanthology.org/2022.bionlp-1.9

[84] S. Yang *et al.*, "Zhongjing: Enhancing the chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 17, 2024, pp. 19 368–19 376.

[85] Q. Xie *et al.*, "Me llama: Foundation large language models for medical applications," *arXiv preprint arXiv:2402.12749*, 2024.

[86] F. Jia *et al.*, "Oncogpt: A medical conversational model tailored with oncology domain expertise on a large language model meta-ai (llama)," *arXiv preprint arXiv:2402.16810*, 2024.

[87] J. Wang *et al.*, "Jmlr: Joint medical llm and retrieval training for enhancing reasoning and professional question answering capability," *arXiv preprint arXiv:2402.17887*, 2024.

[88] Singhal *et al.*, "Large language models encode clinical knowledge," *Nature*, vol. 620, no. 7973, p. E19, 2023.

[89] K. Singhal *et al.*, "Towards expert-level medical question answering with large language models," *arXiv preprint arXiv:2305.09617*, 2023.

[90] S. Pieri *et al.*, "Bimedix: Bilingual medical mixture of experts llm," *arXiv preprint arXiv:2402.13253*, 2024.

[91] C. Shu, B. Chen, F. Liu, Z. Fu, E. Shareghi, and N. Collier, "Visual med-alpaca: A parameter-efficient biomedical llm with visual capabilities," 2023.

[92] W. Gao *et al.*, "Ophglm: Training an ophthalmology large language-and-vision assistant based on instructions and dialogue," *arXiv preprint arXiv:2306.12174*, 2023.

[93] S. Wang *et al.*, "Chatcad: Interactive computer-aided diagnosis on medical image using large language models," *arXiv preprint arXiv:2302.07257*, 2023.

[94] Z. Zhao *et al.*, "Chatcad+: Towards a universal and reliable interactive cad using llms," *arXiv preprint arXiv:2305.15964*, 2023.

[95] Z. Liu *et al.*, "Deid-gpt: Zero-shot medical text de-identification by gpt-4," *arXiv preprint arXiv:2303.11032*, 2023.

[96] Y. Gao *et al.*, "Leveraging a medical knowledge graph into large language models for diagnosis prediction," *arXiv preprint arXiv:2308.14321*, 2023.

[97] H. Nori *et al.*, "Can generalist foundation models outcompete special-purpose tuning? case study in medicine," *Medicine*, vol. 84, no. 88.3, pp. 77–3, 2023.

[98] S. Sivarajkumar and Y. Wang, "Healthprompt: A zero-shot learning paradigm for clinical natural language processing," in *AMIA Annual Symposium Proceedings*, vol. 2022. American Medical Informatics Association, 2022, p. 972.

[99] X. Tang *et al.*, "Medagents: Large language models as collaborators for zero-shot medical reasoning," *arXiv preprint arXiv:2311.10537*, 2023.

[100] A. Elfrink *et al.*, "Soft-prompt tuning to predict lung cancer using primary care free-text dutch medical notes," in *International Conference on Artificial Intelligence in Medicine*. Springer, 2023, pp. 193–198.

[101] M. Abaho *et al.*, "Position-based prompting for health outcome generation," in *Proceedings of the 21st Workshop on Biomedical Language Processing*, 2022, pp. 26–36.

[102] S. Lee *et al.*, "Clinical decision transformer: Intended treatment recommendation through goal prompting," *arXiv preprint arXiv:2302.00612*, 2023.

[103] O. Byambasuren *et al.*, "Preliminary study on the construction of chinese medical knowledge graph," *Journal of Chinese Information Processing*, vol. 33, no. 10, pp. 1–9, 2019.

[104] D. Jin *et al.*, "What disease does this patient have? a large-scale open domain question answering dataset from medical exams," *Applied Sciences*, vol. 11, no. 14, p. 6421, 2021.

[105] D. A. Lindberg *et al.*, "The unified medical language system," *Yearbook of medical informatics*, vol. 2, no. 01, pp. 41–51, 1993.

[106] J. Li *et al.*, "Pre-trained language models for text generation: A survey," *ACM Comput. Surv.*, mar 2024, just Accepted. [Online]. Available: https://doi.org/10.1145/3649449

[107] E. J. Hu *et al.*, "Lora: Low-rank adaptation of large language models," in *Proc. Int. Conf. Learn. Represent.*, 2021.

[108] J. Wang *et al.*, "Prompt engineering for healthcare: Methodologies and applications," *arXiv preprint arXiv:2304.14670*, 2023.

[109] J. Wei *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 24 824–24 837, 2022.

[110] I. Beltagy *et al.*, "Scibert: A pretrained language model for scientific text," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 3615–3620.

[111] S. Verkijk and P. Vossen, "Medroberta. nl: a language model for dutch electronic health records," *Computational Linguistics in the Netherlands Journal*, vol. 11, pp. 141–159, 2021.

[112] M. Awais *et al.*, "Foundational models defining a new era in vision: A survey and outlook," *arXiv preprint arXiv:2307.13721*, 2023.

[113] K. He *et al.*, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 9729–9738.

[114] J. Ma and B. Wang, "Towards foundation models of biological image segmentation," *Nat. Methods*, vol. 20, no. 7, pp. 953–955, 2023.

[115] S. Chen *et al.*, "Med3d: Transfer learning for 3d medical image analysis," *arXiv preprint arXiv:1904.00625*, 2019.

[116] Z. Huang *et al.*, "Stu-net: Scalable and transferable medical image segmentation models empowered by large-scale supervised pre-training," *arXiv preprint arXiv:2304.06716*, 2023.

[117] J. Wasserthal *et al.*, "Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images," *Radiology: Artificial Intelligence*, vol. 5, no. 5, 2023.

[118] V. I. Butoi *et al.*, "Universeg: Universal medical image segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

[119] H. Wang *et al.*, "Sam-med3d," *arXiv preprint arXiv:2310.15161*, 2023.

[120] J. Ye *et al.*, "Sa-med2d-20m dataset: Segment anything in 2d medical imaging with 20 million masks," *arXiv preprint arXiv:2311.11969*, 2023.

[121] H.-Y. Zhou *et al.*, "A unified visual information preservation framework for self-supervised pre-training in medical image analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023.

[122] J. Qiu *et al.*, "Visionfm: a multi-modal multi-task vision foundation model for generalist ophthalmic artificial intelligence," *arXiv preprint arXiv:2310.04992*, 2023.

[123] Y. Du *et al.*, "Segvol: Universal and interactive volumetric medical image segmentation," *arXiv preprint arXiv:2311.13385*, 2023.

[124] Q. Kang *et al.*, "Deblurring masked autoencoder is better recipe for ultrasound image recognition," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 352–362.

[125] J. Jiao *et al.*, "Usfm: A universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis," *arXiv preprint arXiv:2401.00153*, 2024.

[126] Z. Zhou *et al.*, "Models genesis: Generic autodidactic models for 3d medical image analysis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2019, pp. 384–393.

[127] J. Zhou *et al.*, "Image bert pre-training with online tokenizer," in *International Conference on Learning Representations*, 2021.

[128] K. He *et al.*, "Masked autoencoders are scalable vision learners," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16 000–16 009.

[129] Z. Xie, Z. Zhang, Y. Cao, Y. Lin, J. Bao, Z. Yao, Q. Dai, and H. Hu, "Simmim: A simple framework for masked image modeling," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 9653–9663.

[130] T. Chen *et al.*, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2020, pp. 1597–1607.

[131] X. Wang *et al.*, "Transformer-based unsupervised contrastive learning for histopathological image classification," *Med. Image Anal.*, vol. 81, p. 102559, 2022.

[132] O. Ciga *et al.*, "Self supervised contrastive learning for digital histopathology," *Machine Learning with Applications*, vol. 7, p. 100198, 2022.

[133] H. Sowrirajan *et al.*, "Moco pretraining improves representation and transferability of chest x-ray models," in *Proc. Int. Conf. Medical Imaging Deep Learn.* PMLR, 2021, pp. 728–744.

[134] H.-Y. Zhou *et al.*, "Comparing to learn: Surpassing imagenet pretraining on radiographs by comparing image representations," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2020, pp. 398–407.

[135] D. M. Nguyen *et al.*, "Lvm-med: Learning large-scale self-supervised vision models for medical imaging via second-order graph matching," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[136] L. Wu, *et al.*, "Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis," in *IEEE Conf. Comput. Vis. Pattern Recog.*, 2024.

[137] G. Wang *et al.*, "Mis-fm: 3d medical image segmentation using foundation models pretrained on a large-scale unannotated dataset," *arXiv preprint arXiv:2306.16925*, 2023.

[138] F. C. Ghesu *et al.*, "Contrastive self-supervised learning from 100 million medical images with optional supervision," *Journal of Medical Imaging*, vol. 9, no. 6, pp. 064 503–064 503, 2022.

[139] E. Vorontsov *et al.*, "Virchow: A million-slide digital pathology foundation model," *arXiv preprint arXiv:2309.07778*, 2023.

[140] R. J. Chen *et al.*, "Towards a general-purpose foundation model for computational pathology," *Nature Medicine*, 2024.

[141] J. Dippel *et al.*, "Rudolfv: A foundation model by pathologists for pathologists," *arXiv preprint arXiv:2401.04079*, 2024.

[142] M. Oquab *et al.*, "Dinov2: Learning robust visual features without supervision," *Transactions on Machine Learning Research*, 2023.

[143] Y. Wu *et al.*, "Brow: Better features for whole slide image based on self-distillation," *arXiv preprint arXiv:2309.08259*, 2023.

[144] F. Haghighi *et al.*, "Transferable visual words: Exploiting the semantics of anatomical patterns for self-supervised learning," *IEEE transactions on medical imaging*, vol. 40, no. 10, pp. 2857–2868, 2021.

[145] G. Campanella *et al.*, "Computational pathology at health system scale–self-supervised foundation models from billions of images," in *AAAI 2024 Spring Symposium on Clinical Foundation Models*, 2024.

[146] Y. Tang *et al.*, "Self-supervised pre-training of swin transformers for 3d medical image analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 20 730–20 740.

[147] C. Chen *et al.*, "Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation," *arXiv preprint arXiv:2309.08842*, 2023.

[148] S. Pandey *et al.*, "Comprehensive multimodal segmentation in medical imaging: Combining yolov8 with sam and hq-sam models," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2023, pp. 2592–2598.

[149] S. Gong *et al.*, "3dsam-adapter: Holistic adaptation of sam from 2d to 3d for promptable medical image segmentation," *arXiv preprint arXiv:2306.13465*, 2023.

[150] W. Yue *et al.*, "Part to whole: Collaborative prompting for surgical instrument segmentation," *arXiv preprint arXiv:2312.14481*, 2023.

[151] M. Hu *et al.*, "Skinsam: Empowering skin cancer segmentation with segment anything model," *arXiv preprint arXiv:2304.13973*, 2023.

[152] Y. Li *et al.*, "Polyp-sam: Transfer sam for polyp segmentation," *arXiv preprint arXiv:2305.00293*, 2023.

[153] C. Wang *et al.*, "Sam-octa: A fine-tuning strategy for applying foundation model to octa image segmentation tasks," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 1771–1775.

[154] K. Zhang and D. Liu, "Customized segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.13785*, 2023.

[155] S. Chai *et al.*, "Ladder fine-tuning approach for sam integrating complementary network," *arXiv preprint arXiv:2306.12737*, 2023.

[156] W. Feng *et al.*, "Cheap lunch for medical image segmentation by fine-tuning sam on few exemplars," *arXiv preprint arXiv:2308.14133*, 2023.

[157] Y. Zhang *et al.*, "Semisam: Exploring sam for enhancing semi-supervised medical image segmentation with extremely limited annotations," *arXiv preprint arXiv:2312.06316*, 2023.

[158] X. Yan *et al.*, "After-sam: Adapting sam with axial fusion transformer for medical imaging segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 7975–7984.

[159] X. Xiong *et al.*, "Mammo-sam: Adapting foundation segment anything model for automatic breast mass segmentation in whole mammograms," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2023, pp. 176–185.

[160] H. Li *et al.*, "Promise: Prompt-driven 3d medical image segmentation using pretrained image foundation models," *arXiv preprint arXiv:2310.19721*, 2023.

[161] J. Wu *et al.*, "Medical sam adapter: Adapting segment anything model for medical image segmentation," *arXiv preprint arXiv:2304.12620*, 2023.

[162] J. Cheng *et al.*, "Sam-med2d," *arXiv preprint arXiv:2308.16184*, 2023.

[163] J. N. Paranjape *et al.*, "Adaptivesam: Towards efficient tuning of sam for surgical scene segmentation," in *Medical Imaging with Deep Learning*, 2024.

[164] S. Kim *et al.*, "Medivista-sam: Zero-shot medical video analysis with spatio-temporal sam adaptation," *arXiv preprint arXiv:2309.13539*, 2023.

[165] X. Lin *et al.*, "Samus: Adapting segment anything model for clinically-friendly and generalizable ultrasound image segmentation," *arXiv preprint arXiv:2309.06824*, 2023.

[166] H. Gu *et al.*, "Segmentanybone: A universal model that segments any bone at any location on mri," *arXiv preprint arXiv:2401.12974*, 2024.

[167] Z. Feng *et al.*, "Swinsam: Fine-grained polyp segmentation in colonoscopy images via segment anything model integrated with a swin transformer decoder," *Available at SSRN 4673046*.

[168] Y. Zhang *et al.*, "Input augmentation with sam: Boosting medical image segmentation with segmentation foundation model," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2023, pp. 129–139.

[169] T. Shaharabany *et al.*, "Autosam: Adapting sam to medical images by overloading the prompt encoder," *arXiv preprint arXiv:2306.06370*, 2023.

[170] Y. Gao *et al.*, "Desam: Decoupling segment anything model for generalizable medical image segmentation," *arXiv preprint arXiv:2306.00499*, 2023.

[171] U. Israel *et al.*, "A foundation model for cell segmentation," *bioRxiv*, pp. 2023–11, 2023.

[172] G. Deng *et al.*, "Sam-u: Multi-box prompts triggered uncertainty estimation for reliable sam in medical image," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 368–377.

[173] J. Zhang *et al.*, "Sam-path: A segment anything model for semantic segmentation in digital pathology," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 161–170.

[174] C. Cui and R. Deng, "All-in-sam: from weak annotation to pixel-wise nuclei segmentation with prompt-based finetuning," in *Asia Conference on Computers and Communications, ACCC*, 2023.

[175] W. Yue *et al.*, "Surgicalsam: Efficient class promptable surgical instrument segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[176] R. Biswas, "Polyp-sam++: Can a text guided sam perform better for polyp segmentation?" *arXiv preprint arXiv:2308.06623*, 2023.

[177] Y. Zhang *et al.*, "Segment anything model with uncertainty rectification for auto-prompting medical image segmentation," *arXiv preprint arXiv:2311.10529*, 2023.

[178] W. Lei *et al.*, "Medlsam: Localize and segment anything model for 3d medical images," *arXiv preprint arXiv:2306.14752*, 2023.

[179] Y. Li *et al.*, "nnsam: Plug-and-play segment anything model improves nnunet performance," *arXiv preprint arXiv:2309.16967*, 2023.

[180] Y. Xu *et al.*, "Eviprompt: A training-free evidential prompt generation method for segment anything model in medical images," *arXiv preprint arXiv:2311.06400*, 2023.

[181] D. Anand *et al.*, "One-shot localization and segmentation of medical images with foundation models," in *R0-FoMo: Robustness of Few-shot and Zero-shot Learning in Large Foundation Models*, 2023.

[182] Y. Liu *et al.*, "Samm (segment any medical model): A 3d slicer integration to sam," *arXiv preprint arXiv:2304.05622*, 2023.

[183] R. Sathish *et al.*, "Task-driven prompt evolution for foundation models," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2023, pp. 256–264.

[184] M. Fischer *et al.*, "Prompt tuning for parameter-efficient medical image segmentation," *Medical Image Analysis*, vol. 91, p. 103024, 2024.

[185] T. Chen *et al.*, "Sam fails to segment anything?–sam-adapter: Adapting sam in underperformed scenes: Camouflage, shadow, and more," *arXiv preprint arXiv:2304.09148*, 2023.

[186] A. Madani *et al.*, "Large language models generate functional protein sequences across diverse families," *Nat. Biotechnol.*, pp. 1–8, 2023.

[187] E. Nijkamp *et al.*, "Progen2: Exploring the boundaries of protein language models," *Cell Systems*, vol. 14, pp. 968–978.e3, 11 2023, doi: 10.1016/j.cels.2023.10.002.

[188] F. Yang *et al.*, "scbert as a large-scale pretrained deep language model for cell type annotation of single-cell rna-seq data," *Nat. Mach. Intell*, vol. 4, pp. 852–866, 2022. [Online]. Available: https://doi.org/10.1038/s42256-022-00534-z

[189] C. V. Theodoris *et al.*, "Transfer learning enables predictions in network biology," *Nature*, vol. 618, pp. 616–624, 2023. [Online]. Available: https://doi.org/10.1038/s41586-023-06139-9

[190] Z. Zhou *et al.*, "Dnabert-2: Efficient foundation model and benchmark for multi-species genomes," in *Proc. Int. Conf. Learn. Represent.*, 2023.

[191] V. Fishman *et al.*, "Gena-lm: A family of open-source foundational models for long dna sequences," *bioRxiv*, p. 2023.06.12.544594, 1 2023. [Online]. Available: http://biorxiv.org/content/early/2023/06/13/2023.06.12.544594.abstract

[192] Y. Zhang *et al.*, "Multiple sequence-alignment-based rna language model and its application to structural inference," *bioRxiv*, p. 2023.03.15.532863, 1 2023. [Online]. Available: http://biorxiv.org/content/early/2023/03/16/2023.03.15.532863.abstract

[193] K. Chen *et al.*, "Self-supervised learning on millions of pre-mrna sequences improves sequence-based rna splicing prediction," *bioRxiv*, p. 2023.01.31.526427, 1 2023. [Online]. Available: http://biorxiv.org/content/early/2023/02/03/2023.01.31.526427.abstract

[194] Y. Yang *et al.*, "Deciphering 3' utr mediated gene regulation using interpretable deep representation learning," *bioRxiv*, p. 2023.09.08.556883, 1 2023. [Online]. Available: http://biorxiv.org/content/early/2023/09/12/2023.09.08.556883.abstract

[195] Z. Lin *et al.*, "Evolutionary-scale prediction of atomic-level protein structure with a language model," *Science*, vol. 379, pp. 1123–1130, 3 2023, doi: 10.1126/science.ade2574.

[196] A. Elnaggar *et al.*, "Prottrans: Toward understanding the language of life through self-supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, pp. 7112–7127, 2022.

[197] R. M. Rao *et al.*, "Msa transformer," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 8844–8856.

[198] A. Rives *et al.*, "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences," *Proc. Natl. Acad. Sci.*, vol. 118, p. e2016239118, 4 2021, doi: 10.1073/pnas.2016239118. [Online]. Available: https://doi.org/10.1073/pnas.2016239118

[199] E. Nguyen *et al.*, "Hyenadna: Long-range genomic sequence modeling at single nucleotide resolution," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[200] M. Hao *et al.*, "Large scale foundation model on single-cell transcriptomics," *bioRxiv*, p. 2023.05.29.542705, 1 2023. [Online]. Available: http://biorxiv.org/content/early/2023/06/15/2023.05.29.542705.abstract

[201] Y. Rosen *et al.*, "Universal cell embeddings: A foundation model for cell biology," *bioRxiv*, p. 2023.11.28.568918, 1 2023. [Online]. Available: http://biorxiv.org/content/early/2023/11/29/2023.11.28.568918.abstract

[202] D. Zhang *et al.*, "Dnagpt: A generalized pretrained tool for multiple dna sequence analysis tasks," *arXiv preprint arXiv:2307.05628*, 2023.

[203] H. Cui *et al.*, "scgpt: towards building a foundation model for single-cell multi-omics using generative ai," *bioRxiv*, pp. 2023–04, 2023.

[204] M. Akiyama and Y. Sakakibara, "Informative rna base embedding for rna structural alignment and clustering by deep representation learning," *NAR Genomics and Bioinformatics*, vol. 4, p. lqac012, 3 2022. [Online]. Available: https://doi.org/10.1093/nargab/lqac012

[205] R. Chowdhury *et al.*, "Single-sequence protein structure prediction using a language model and deep learning," *Nature Biotechnol.*, vol. 40, no. 11, pp. 1617–1623, 2022.

[206] Y. Chu *et al.*, "A 5' utr language model for decoding untranslated regions of mrna and function predictions," *bioRxiv*, p. 2023.10.11.561938, 1 2023. [Online]. Available: http://biorxiv.org/content/early/2023/10/14/2023.10.11.561938.abstract

[207] S. Zhao *et al.*, "Large-scale cell representation learning via divide-and-conquer contrastive learning," *arXiv preprint arXiv:2306.04371*, 2023.

[208] S. Mo *et al.*, "Multi-modal self-supervised pre-training for regulatory genome across cell types," *arXiv preprint arXiv:2110.05231*, 2021.

[209] S. Li *et al.*, "Codonbert: Large language models for mrna design and optimization," in *NeurIPS 2023 Generative AI and Biology (GenBio) Workshop*, 2023.

[210] B. Chen *et al.*, "xtrimopglm: Unified 100b-scale pre-trained transformer for deciphering the language of protein," *bioRxiv*, p. 2023.07.05.547496, 1 2024. [Online]. Available: http://biorxiv.org/content/early/2024/01/11/2023.07.05.547496.abstract

[211] Z. Du *et al.*, "GLM: General language model pretraining with autoregressive blank infilling," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Dublin, Ireland: Association for Computational Linguistics, may 2022, pp. 320–335.

[212] Y. T. Chen and J. Zou, "Genept: A simple but hard-to-beat foundation model for genes and cells built from chatgpt," *bioRxiv*, p. 2023.10.16.562533, 1 2023. [Online]. Available: http://biorxiv.org/content/early/2023/10/19/2023.10.16.562533.abstract

[213] T. Liu *et al.*, "scelmo: Embeddings from language models are good learners for single-cell data analysis," *bioRxiv*, 2024. [Online]. Available: https://www.biorxiv.org/content/early/2024/03/03/2023.12.07.569910

[214] B. E. Slatko *et al.*, "Overview of next-generation sequencing technologies," *Current Protocols in Molecular Biology*, vol. 122, no. 1, p. e59, 2018. [Online]. Available: https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpmb.59

[215] T. Wu *et al.*, "A brief overview of chatgpt: The history, status quo and potential future development," *IEEE/CAA J. Autom. Sin.*, vol. 10, no. 5, pp. 1122–1136, 2023.

[216] Y. Khare *et al.*, "Mmbert: Multimodal bert pretraining for improved medical vqa," in *Proc. IEEE Int. Symp. Biomed. Imaging*. IEEE, 2021, pp. 1033–1036.

[217] H.-Y. Zhou *et al.*, "Advancing radiograph representation learning with masked record modeling," *The Eleventh International Conference on Learning Representations*, 2023.

[218] Y. Zhang *et al.*, "Contrastive learning of medical visual representations from paired images and text," in *Machine Learning for Healthcare Conference*. PMLR, 2022, pp. 2–25.

[219] P. Müller *et al.*, "Joint learning of localized representations from medical images and reports," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2022, pp. 685–701.

[220] J. Lei *et al.*, "Unibrain: Universal brain mri diagnosis with hierarchical knowledge-enhanced pre-training," *arXiv preprint arXiv:2309.06828*, 2023.

[221] C. Liu *et al.*, "M-flag: Medical vision-language pre-training with frozen language models and latent space geometry optimization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2023, pp. 637–647.

[222] F. Wang *et al.*, "Multi-granularity cross-modal alignment for generalized medical visual representation learning," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 33 536–33 549, 2022.

[223] C. Wu *et al.*, "Medklip: Medical knowledge enhanced language-image pre-training," *medRxiv*, pp. 2023–01, 2023.

[224] C. Liu *et al.*, "Etp: Learning transferable ecg representations via ecg-text pre-training," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 8230–8234.

[225] S.-C. Huang *et al.*, "Gloria: A multimodal global-local representation learning framework for label-efficient medical image recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 3942–3951.

[226] C. Liu *et al.*, "Imitate: Clinical prior guided hierarchical vision-language pre-training," *arXiv preprint arXiv:2310.07355*, 2023.

[227] Z. Wang *et al.*, "Medclip: Contrastive learning from unpaired medical images and text," in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 2022, pp. 3876–3887.

[228] Z. Wan *et al.*, "Med-unic: Unifying cross-lingual medical vision-language pre-training by diminishing bias," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[229] K. You *et al.*, "Cxr-clip: Toward large scale chest x-ray language-image pre-training," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2023, pp. 101–111.

[230] S. Zhang *et al.*, "Large-scale domain-specific pretraining for biomedical vision-language processing," *arXiv preprint arXiv:2303.00915*, 2023.

[231] Y. Wang and G. Wang, "Umcl: Unified medical image-text-label contrastive learning with continuous prompt," in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2023, pp. 2285–2289.

[232] X. Zhang *et al.*, "Knowledge-enhanced visual-language pre-training on chest radiology images," *Nat. Commun.*, vol. 14, no. 1, p. 4542, 2023.

[233] S. Liu *et al.*, "Multi-modal molecule structure–text model for text-based retrieval and editing," *Nature Machine Intelligence*, vol. 5, no. 12, pp. 1447–1457, 2023.

[234] Y. Lei *et al.*, "Clip-lung: Textual knowledge-guided lung nodule malignancy prediction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*. Cham: Springer Nature Switzerland, 2023, pp. 403–412.

[235] C. Seibold *et al.*, "Breaking with fixed set pathology recognition through report-guided contrastive training," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2022, pp. 690–700.

[236] M. Y. Lu *et al.*, "Visual language pretrained multiple instance zero-shot transfer for histopathology images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 19 764–19 775.

[237] B. Yan and M. Pei, "Clinical-bert: Vision-language pre-training for radiograph diagnosis and reports generation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 3, 2022, pp. 2982–2990.

[238] Z. Chen *et al.*, "Multi-modal masked autoencoders for medical vision-and-language pre-training," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2022, pp. 679–689.

[239] J. H. Moon *et al.*, "Multi-modal understanding and generation for medical images and text via vision-language pre-training," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 12, pp. 6070–6080, 2022.

[240] W. Lin and other, "Pmc-clip: Contrastive language-image pre-training using biomedical documents," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*. Cham: Springer Nature Switzerland, 2023, pp. 525–536.

[241] Z. Chen *et al.*, "Align, reason and learn: Enhancing medical vision-and-language pre-training with knowledge," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 5152–5161.

[242] W. Huang *et al.*, "Enhancing representation in radiography-reports foundation model: A granular alignment algorithm using masked contrastive learning," *arXiv preprint arXiv:2309.05904*, 2023.

[243] P. Li *et al.*, "Masked vision and language pre-training with unimodal and multimodal contrastive losses for medical visual question answering," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2023, pp. 374–383.

[244] C. Liu *et al.*, "T3d: Towards 3d medical image understanding through vision-language pre-training," *arXiv preprint arXiv:2312.01529*, 2023.

[245] T. Jin and Others, "Gene-induced multimodal pre-training for image-omic classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2023, pp. 508–517.

[246] B. Boecking *et al.*, "Making the most of text semantics to improve biomedical vision–language processing," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2022, pp. 1–21.

[247] P. Cheng *et al.*, "Prior: Prototype representation joint learning from medical images and reports," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2023, pp. 21 361–21 371.

[248] M. Y. Lu *et al.*, "A visual-language foundation model for computational pathology," *Nature Medicine*, 2024.

[249] S. Liu *et al.*, "A text-guided protein design framework," *arXiv preprint arXiv:2302.04611*, 2023.

[250] S. Eslami *et al.*, "Pubmedclip: How much does clip benefit visual question answering in the medical domain?" in *Findings of the Association for Computational Linguistics: EACL 2023*, 2023, pp. 1181–1193.

[251] M. Moor *et al.*, "Med-flamingo: a multimodal medical few-shot learner," in *Machine Learning for Health*. PMLR, 2023, pp. 353–367.

[252] C. Li *et al.*, "Llava-med: Training a large language-and-vision assistant for biomedicine in one day," *Advances in Neural Information Processing Systems*, 2024.

[253] E. Tiu *et al.*, "Expert-level detection of pathologies from unannotated chest x-ray images via self-supervised learning," *Nat. Biomed. Eng.*, vol. 6, no. 12, pp. 1399–1406, 2022.

[254] W. Ikezogwo *et al.*, "Quilt-1m: One million image-text pairs for histopathology," *Advances in Neural Information Processing Systems*, 2024.

[255] Z. Huang *et al.*, "A visual–language foundation model for pathology image analysis using medical twitter," *Nat. Med.*, vol. 29, no. 9, pp. 2307–2316, 2023.

[256] S. Baliah *et al.*, "Exploring the transfer learning capabilities of clip in domain generalization for diabetic retinopathy," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2023, pp. 444–453.

[257] P. Chambon *et al.*, "Roentgen: vision-language foundation model for chest x-ray generation," *arXiv preprint arXiv:2211.12737*, 2022.

[258] T. Van Sonsbeek *et al.*, "Open-ended medical visual question answering through prefix tuning of language models," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 726–736.

[259] P. Chambon *et al.*, "Adapting pretrained vision-language foundational models to medical imaging domains," in *NeurIPS 2022 Foundation Models for Decision Making Workshop*, 2022.

[260] J. Liu *et al.*, "Qilin-med-vl: Towards chinese large vision-language model for general healthcare," *arXiv preprint arXiv:2310.17956*, 2023.

[261] Y. Sun *et al.*, "Pathasst: Redefining pathology through generative foundation ai assistant for pathology," *Proc. AAAI Conf. Artif. Intell.*, 2024.

[262] M. Y. Lu *et al.*, "A foundational multimodal vision language ai assistant for human pathology," *arXiv preprint arXiv:2312.07814*, 2023.

[263] Y. Lu *et al.*, "Effectively fine-tune to improve large multimodal models for radiology report generation," in *Deep Generative Models for Health Workshop NeurIPS 2023*, 2023.

[264] Z. Yu *et al.*, "Multi-modal adapter for medical vision-and-language learning," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2023, pp. 393–402.

[265] T. T. Pham *et al.*, "I-ai: A controllable & interpretable ai system for decoding radiologists' intense focus for accurate cxr diagnoses," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 7850–7859.

[266] Y. Zhang *et al.*, "Text-guided foundation model adaptation for pathological image classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2023, pp. 272–282.

[267] O. Thawkar *et al.*, "Xraygpt: Chest radiographs summarization using medical vision-language models," *arXiv preprint arXiv:2306.07971*, 2023.

[268] C. Pellegrini *et al.*, "Xplainer: From x-ray observations to explainable zero-shot diagnosis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 420–429.

[269] Z. Qin *et al.*, "Medical image understanding with pretrained vision language models: A comprehensive study," in *The Eleventh International Conference on Learning Representations*, 2022.

[270] M. Guo and Others, "Multiple prompt fusion for zero-shot lesion detection using vision-language models," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2023, pp. 283–292.

[271] Z. Wang *et al.*, "Biobridge: Bridging biomedical foundation models via knowledge graph," *arXiv preprint arXiv:2310.03320*, 2023.

[272] Y. Li *et al.*, "A comparison of pre-trained vision-and-language models for multimodal representation learning across medical images and reports," in *2020 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2020, pp. 1999–2004.

[273] J. Yu *et al.*, "Coca: Contrastive captioners are image-text foundation models," *arXiv preprint arXiv:2205.01917*, 2022.

[274] H. Liu *et al.*, "Visual instruction tuning," *Advances in neural information processing systems*, vol. 36, 2024.

[275] J.-B. Alayrac *et al.*, "Flamingo: a visual language model for few-shot learning," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 23 716–23 736, 2022.

[276] D. Driess *et al.*, "Palm-e: An embodied multimodal language model," in *International Conference on Machine Learning*. PMLR, 2023, pp. 8469–8488.

[277] Z. Zhao *et al.*, "A large-scale dataset of patient summaries for retrieval-based clinical decision support systems." *Scientific data*, vol. 10 1, p. 909, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:266360591

[278] A. E. Johnson *et al.*, "Mimic-iii, a freely accessible critical care database," *Sci. Data*, vol. 3, no. 1, pp. 1–9, 2016.

[279] A. Johnson *et al.*, "Mimic-iv, a freely accessible electronic health record dataset," *Scientific data*, vol. 10, no. 1, p. 1, 2023.

[280] T. J. Pollard *et al.*, "The eicu collaborative research database, a freely available multi-center database for critical care research," *Scientific data*, vol. 5, no. 1, pp. 1–13, 2018.

[281] W. Chen *et al.*, "A benchmark for automatic medical consultation system: frameworks, tasks and datasets," *Bioinformatics*, vol. 39, no. 1, p. btac817, 2023.

[282] J. Li *et al.*, "Huatuo-26m, a large-scale chinese medical qa dataset," *arXiv preprint arXiv:2305.01526*, 2023.

[283] M. Zhu *et al.*, "Question answering with long multiple-span answers," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, T. Cohn, Y. He, and Y. Liu, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 3840–3849. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.342

[284] A. Ben Abacha and Others, "A question-entailment approach to question answering," *BMC bioinformatics*, vol. 20, pp. 1–23, 2019.

[285] W. Liu *et al.*, "Meddg: An entity-centric medical consultation dataset for entity-aware medical dialogue generation," in *Natural Language Processing and Chinese Computing*. Cham: Springer International Publishing, 2022, pp. 447–459.

[286] J. Liu *et al.*, "Benchmarking large language models on cmexam-a comprehensive chinese medical exam dataset," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[287] S. Zhang *et al.*, "Multi-scale attentive interaction networks for chinese medical question answer selection," *IEEE Access*, vol. 6, pp. 74 061–74 071, 2018.

[288] S. Suster and W. Daelemans, "Clicr: a dataset of clinical case reports for machine reading comprehension," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 1551–1563.

[289] J. He *et al.*, "Applying deep matching networks to chinese medical question answering: a study and a dataset," *BMC medical informatics and decision making*, vol. 19, pp. 91–100, 2019.

[290] G. Zeng *et al.*, "Meddialog: Large-scale medical dialogue datasets," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 9241–9250.

[291] M. Zhu *et al.*, "A hierarchical attention retrieval model for healthcare question answering," in *The World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 2472–2482. [Online]. Available: https://doi.org/10.1145/3308558.3313699

[292] Y. Hu *et al.*, "Omnimedvqa: A new large-scale comprehensive evaluation benchmark for medical lvlm," *arXiv preprint arXiv:2402.09181*, 2024.

[293] N. Zhang *et al.*, "Cblue: A chinese biomedical language understanding evaluation benchmark," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2022, pp. 7888–7915.

[294] D. Wang *et al.*, "A real-world dataset and benchmark for foundation model adaptation in medical image classification," *Scientific Data*, vol. 10, no. 1, p. 574, 2023.

[295] M. Antonelli *et al.*, "The medical segmentation decathlon," *Nat. Commun.*, vol. 13, no. 1, p. 4128, 2022.

[296] J. Ma *et al.*, "Unleashing the strengths of unlabeled data in pan-cancer abdominal organ quantification: the flare22 challenge," *arXiv preprint arXiv:2308.05862*, 2023.

[297] J. Wasserthal *et al.*, "Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images," *Radiology: Artificial Intelligence*, vol. 5, no. 5, p. e230024, 2023.

[298] J. Ma *et al.*, "Abdomenct-1k: Is abdominal organ segmentation a solved problem?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6695–6714, 2022.

[299] Y. Deng *et al.*, "Ctspine1k: A large-scale dataset for spinal vertebrae segmentation in computed tomography," *arXiv preprint arXiv:2105.14711*, 2021.

[300] P. Liu *et al.*, "Deep learning to segment pelvic bones: large-scale ct datasets and baseline models," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 5, p. 749, 2021.

[301] U. Baid *et al.*, "The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint arXiv:2107.02314*, 2021.

[302] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[303] S. Bakas *et al.*, "Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features," *Sci. Data*, vol. 4, no. 1, p. 170117, 2017.

[304] D. LaBella *et al.*, "The asnr-miccai brain tumor segmentation (brats) challenge 2023: Intracranial meningioma," *arXiv preprint arXiv:2305.07642*, 2023.

[305] R. C. Petersen *et al.*, "Alzheimer's disease neuroimaging initiative (adni): clinical characterization," *Neurology*, vol. 74, no. 3, pp. 201–209, 2010.

[306] K. Marek *et al.*, "The parkinson progression marker initiative (ppmi)," *Progress in neurobiology*, vol. 95, no. 4, pp. 629–635, 2011.

[307] S. Gatidis *et al.*, "A whole-body fdg-pet/ct dataset with manually annotated tumor lesions," *Sci. Data*, vol. 9, no. 1, p. 601, 2022.

[308] ——, "The autopet challenge: Towards fully automated lesion segmentation in oncologic pet/ct imaging," 2023.

[309] N. F. Greenwald *et al.*, "Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning," *Nature biotechnology*, vol. 40, no. 4, pp. 555–565, 2022.

[310] K. Chang *et al.*, "The cancer genome atlas pan-cancer analysis project," *Nature Genetics*, vol. 45, pp. 1113–1120, 2013. [Online]. Available: https://doi.org/10.1038/ng.2764

[311] Y. J. Kim *et al.*, "Paip 2019: Liver cancer segmentation challenge," *Med. Image Anal.*, vol. 67, p. 101854, 2021.

[312] A. A. Borkowski *et al.*, "Lung and colon cancer histopathological image dataset (lc25000)," *arXiv preprint arXiv:1912.12142*, 2019.

[313] J. N. Kather *et al.*, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLoS Med.*, vol. 16, no. 1, p. e1002730, 2019.

[314] X. Wang, *et al.*, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3462–3471.

[315] P. Rajpurkar *et al.*, "Mura: Large dataset for abnormality detection in musculoskeletal radiographs," in *Medical Imaging with Deep Learning*, 2022.

[316] V. Rotemberg *et al.*, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Sci. Data*, vol. 8, no. 1, p. 34, 2021.

[317] C. De Vente *et al.*, "Airogs: artificial intelligence for robust glaucoma screening challenge," *IEEE transactions on medical imaging*, 2023.

[318] M. Subramanian *et al.*, "Classification of retinal oct images using deep learning," in *2022 International Conference on Computer Communication and Informatics (ICCCI)*, 2022, pp. 1–7.

[319] A. Montoya *et al.*, "Ultrasound nerve segmentation," 2016. [Online]. Available: https://kaggle.com/competitions/ultrasound-nerve-segmentation

[320] X. P. Burgos-Artizzu *et al.*, "Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes," *Sci. Rep.*, vol. 10, no. 1, p. 10200, 2020.

[321] D. Ouyang *et al.*, "Video-based ai for beat-to-beat assessment of cardiac function," *Nature*, vol. 580, no. 7802, pp. 252–256, 2020.

[322] G. Polat *et al.*, "Improving the computer-aided estimation of ulcerative colitis severity according to mayo endoscopic score by using regression-based deep learning," *Nes. Nutr. Ws.*, p. izac226, 2022.

[323] M. Misawa *et al.*, "Development of a computer-aided detection system for colonoscopy and a publicly accessible large colonoscopy video database (with video)," *Gastrointestinal endoscopy*, vol. 93, no. 4, pp. 960–967, 2021.

[324] P. H. Smedsrud *et al.*, "Kvasir-capsule, a video capsule endoscopy dataset," *Sci. Data*, vol. 8, no. 1, p. 142, 2021.

[325] K. B. Ozyoruk *et al.*, "Endoslam dataset and an unsupervised monocular visual odometry and depth estimation approach for endoscopic videos," *Med. Image Anal.*, vol. 71, p. 102058, 2021.

[326] H. Borgli *et al.*, "Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Sci. Data*, vol. 7, no. 1, pp. 1–14, 2020.

[327] C. I. Nwoye and N. Padoy, "Data splits and metrics for method benchmarking on surgical action triplet datasets," *arXiv preprint arXiv:2204.05235*, 2022.

[328] Y. Ma *et al.*, "Ldpolypvideo benchmark: a large-scale colonoscopy video dataset of diverse polyps," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*. Springer, 2021, pp. 387–396.

[329] K. Yan *et al.*, "Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of medical imaging*, vol. 5, no. 3, pp. 036 501–036 501, 2018.

[330] S. G. Armato III *et al.*, "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical physics*, vol. 38, no. 2, pp. 915–931, 2011.

[331] S.-L. Liew *et al.*, "A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms," *Sci. Data*, vol. 9, no. 1, p. 320, 2022.

[332] A. Saha *et al.*, "Artificial intelligence and radiologists at prostate cancer detection in mri—the pi-cai challenge," in *Medical Imaging with Deep Learning, short paper track*, 2023.

[333] N. Bien *et al.*, "Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet," *PLoS medicine*, vol. 15, no. 11, p. e1002699, 2018.

[334] G. Duffy *et al.*, "High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning," *JAMA cardiology*, vol. 7, no. 4, pp. 386–395, 2022.

[335] P. Ghahremani *et al.*, "Deep learning-inferred multiplex immunofluorescence for immunohistochemical image quantification," *Nature machine intelligence*, vol. 4, no. 4, pp. 401–412, 2022.

[336] N. L. S. T. R. Team, "The national lung screening trial: overview and study design," *Radiology*, vol. 258, no. 1, pp. 243–253, 2011.

[337] K. Ding *et al.*, "A large-scale synthetic pathological dataset for deep learning-enabled segmentation of breast cancer," *Sci. Data*, vol. 10, no. 1, p. 231, 2023.

[338] C. S.-C. Biology *et al.*, "Cz cell×gene discover: A single-cell data platform for scalable exploration, analysis and modeling of aggregated data," *bioRxiv*, pp. 2023–10, 2023.

[339] D. A. Benson *et al.*, "GenBank," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D36–D42, 11 2012. [Online]. Available: https://doi.org/10.1093/nar/gks1195

[340] L. Tarhan *et al.*, "Single cell portal: an interactive home for single-cell genomics data," *bioRxiv*, 2023.

[341] A. Frankish *et al.*, "GENCODE reference annotation for the human and mouse genomes," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D766–D773, 10 2018. [Online]. Available: https://doi.org/10.1093/nar/gky955

[342] A. Regev *et al.*, "Science forum: The human cell atlas," *eLife*, vol. 6, p. e27041, dec 2017. [Online]. Available: https://doi.org/10.7554/eLife.27041

[343] B. J. Raney *et al.*, "The UCSC Genome Browser database: 2024 update," *Nucleic Acids Res.*, vol. 52, no. D1, pp. D1082–D1088, 11 2023. [Online]. Available: https://doi.org/10.1093/nar/gkad987

[344] N. J. Edwards *et al.*, "The cptac data portal: A resource for cancer proteomics research," *Journal of Proteome Research*, vol. 14, no. 6, pp. 2707–2713, 2015.

[345] F. J. Martin *et al.*, "Ensembl 2023," *Nucleic Acids Res.*, vol. 51, no. D1, pp. D933–D941, 2023.

[346] The RNAcentral Consortium, "RNAcentral: a hub of information for non-coding RNA sequences," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D221–D229, 11 2018. [Online]. Available: https://doi.org/10.1093/nar/gky1034

[347] D. R. Armstrong *et al.*, "Pdbe: improved findability of macromolecular structure data in the pdb," *Nucleic acids research*, vol. 48, p. D335—D343, 1 2020. [Online]. Available: https://europepmc.org/articles/PMC7145656

[348] T. U. Consortium, "Uniprot: the universal protein knowledgebase in 2023," *Nucleic Acids Research*, vol. 51, pp. D523–D531, 1 2023. [Online]. Available: https://doi.org/10.1093/nar/gkac1052

[349] I. NeuroLINCS (University of California, "imn (exp 2) - als, sma and control (unaffected) imn cell lines differentiated from ips cell lines using a long differentiation protocol - rna-seq," 2017. [Online]. Available: http://lincsportal.ccs.miami.edu/datasets/#/view/LDS-1398

[350] W. Yang *et al.*, "Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells," *Nucleic Acids Research*, vol. 41, no. D1, pp. D955–D961, 11 2012.

[351] M. Ghandi *et al.*, "Next-generation characterization of the cancer cell line encyclopedia," *Nature*, vol. 569, pp. 503–508, 2019. [Online]. Available: https://doi.org/10.1038/s41586-019-1186-3

[352] C. Bycroft *et al.*, "The uk biobank resource with deep phenotyping and genomic data," *Nature*, vol. 562, no. 7726, pp. 203–209, 2018.

[353] Z. Zhao *et al.*, "Chinese glioma genome atlas (cgga): A comprehensive resource with functional genomic data from chinese glioma patients," *Genomics, Proteomics & Bioinformatics*, vol. 19, pp. 1–12, 2021.

[354] A. E. Johnson *et al.*, "Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports," *Sci. Data*, vol. 6, no. 1, p. 317, 2019.

[355] A. Bustos *et al.*, "Padchest: A large chest x-ray image dataset with multi-label annotated reports," *Med. Image Anal.*, vol. 66, p. 101797, 2020.

[356] J. Irvin *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, 2019, pp. 590–597.

[357] A. García Seco de Herrera *et al.*, "Overview of the imageclef 2018 caption prediction tasks," in *Working Notes of CLEF 2018-Conference and Labs of the Evaluation Forum (CLEF 2018), Avignon, France, September 10-14, 2018.*, vol. 2125. CEUR Workshop Proceedings, 2018.

[358] X. He, Y. Zhang, L. Mou, E. Xing, and P. Xie, "Pathvqa: 30000+ questions for medical visual question answering," *arXiv preprint arXiv:2003.10286*, 2020.

[359] M. Tsuneki and F. Kanavati, "Inference of captions from histopathological patches," in *Proc. Int. Conf. Medical Imaging Deep Learn.* PMLR, 2022, pp. 1235–1250.

[360] P. Wagner *et al.*, "Ptb-xl, a large publicly available electrocardiography dataset," *Sci. Data*, vol. 7, no. 1, p. 154, 2020.

[361] O. Pelka *et al.*, "Radiology objects in context (roco): a multimodal image dataset," in *Intravascular Imaging and Computer Assisted Stenting and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis: 7th Joint International Workshop, CVII-STENT 2018 and Third International Workshop, LABELS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3*. Springer, 2018, pp. 180–189.

[362] S. Subramanian *et al.*, "Medicat: A dataset of medical images, captions, and textual references," in *Findings of the Association for Computational Linguistics, ACL 2020: EMNLP 2020*. Association for Computational Linguistics (ACL), 2020, pp. 2112–2120.

[363] X. Zhang *et al.*, "Pmc-vqa: Visual instruction tuning for medical visual question answering," *arXiv preprint arXiv:2305.10415*, 2023.

[364] A. Saha *et al.*, "A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 dce-mri features," *British journal of cancer*, vol. 119, no. 4, pp. 508–516, 2018.

[365] W. Li *et al.*, "I-SPY 2 Breast Dynamic Contrast Enhanced MRI Trial (ISPY2)." [Online]. Available: https://doi.org/10.7937/TCIA.D8Z0-9T85

[366] J. Gamper and N. Rajpoot, "Multiple instance captioning: Learning representations from histopathology textbooks and articles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2021, pp. 16 549–16 559.

[367] K. Clark *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, pp. 1045–1057, 2013.

[368] E. P. Balogh *et al.*, *Improving diagnosis in health care*. National Academies Press (US), 2015.

[369] D. Ueda *et al.*, "Diagnostic performance of chatgpt from patient history and imaging findings on the diagnosis please quizzes," *Radiology*, vol. 308, no. 1, p. e231040, 2023.

[370] S.-H. Wu *et al.*, "Collaborative enhancement of consistency and accuracy in us diagnosis of thyroid nodules using large language models," *Radiology*, vol. 310, no. 3, p. e232255, 2024.

[371] S. R. Ali *et al.*, "Using chatgpt to write patient clinic letters," *The Lancet Digital Health*, vol. 5, no. 4, pp. e179–e181, 2023.

[372] A. Abd-Alrazaq *et al.*, "Large language models in medical education: Opportunities, challenges, and future directions," *JMIR Medical Education*, vol. 9, no. 1, p. e48291, 2023.

[373] M. Karabacak *et al.*, "The advent of generative language models in medical education," *JMIR Medical Education*, vol. 9, p. e48163, 2023.

[374] T. H. Kung *et al.*, "Performance of chatgpt on usmle: Potential for ai-assisted medical education using large language models," *PLoS Digital Health*, vol. 2, no. 2, p. e0000198, 2023.

[375] A. B. Coşkun *et al.*, "Integration of chatgpt and e-health literacy: Opportunities, challenges, and a look towards the future," *Journal of Health Reports and Technology*, vol. 10, no. 1, 2024.

[376] P. Lee *et al.*, "Benefits, limits, and risks of gpt-4 as an ai chatbot for medicine," *New Engl. J. Med.*, vol. 388, no. 13, pp. 1233–1239, 2023.

[377] Y. Chen *et al.*, "Soulchat: Improving llms' empathy, listening, and comfort abilities through fine-tuning with multi-turn empathy conversations," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, 2023, pp. 1170–1183.

[378] Y. Luo *et al.*, "Biomedgpt: Open multimodal generative pre-trained transformer for biomedicine," *arXiv preprint arXiv:2308.09442*, 2023.

[379] L. Huawei Technologies Co., "A general introduction to artificial intelligence," in *Artificial Intelligence Technology*. Springer, 2022, pp. 1–41.

[380] D. B. Larson *et al.*, "Ethics of using and sharing clinical imaging data for artificial intelligence: a proposed framework," *Radiology*, vol. 295, no. 3, pp. 675–682, 2020.

[381] S. Salerno *et al.*, "Overdiagnosis and overimaging: an ethical issue for radiological protection," *La radiologia medica*, vol. 124, pp. 714–720, 2019.

[382] D. Kaur *et al.*, "Trustworthy artificial intelligence: a review," *ACM Computing Surveys (CSUR)*, vol. 55, no. 2, pp. 1–38, 2022.

[383] M. Haendel *et al.*, "How many rare diseases are there?" *Nat. Rev. Drug Discov.*, vol. 19, no. 2, pp. 77–78, 2020.

[384] H. Guan and M. Liu, "Domain adaptation for medical image analysis: a survey," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 3, pp. 1173–1185, 2021.

[385] Z. Liu and K. He, "A decade's battle on dataset bias: Are we there yet?" *arXiv preprint arXiv:2403.08632*, 2024.

[386] A. Cassidy *et al.*, "Lung cancer risk prediction: a tool for early detection," *Int. J. Cancer*, vol. 120, no. 1, pp. 1–6, 2007.

[387] J. Gama *et al.*, "A survey on concept drift adaptation," *ACM computing surveys (CSUR)*, vol. 46, no. 4, pp. 1–37, 2014.

[388] S. Wang *et al.*, "Annotation-efficient deep learning for automatic medical image segmentation," *Nat. Commun.*, vol. 12, no. 1, p. 5915, 2021.

[389] N. Tajbakhsh *et al.*, "Guest editorial annotation-efficient deep learning: the holy grail of medical imaging," *IEEE Trans. Med. Imaging*, vol. 40, no. 10, pp. 2526–2533, 2021.

[390] L. Sun *et al.*, "Trustllm: Trustworthiness in large language models," *arXiv preprint arXiv:2401.05561*, 2024.

[391] K. Sokol and P. Flach, "One explanation does not fit all: The promise of interactive explanations for machine learning transparency," *KI-Künstliche Intelligenz*, vol. 34, no. 2, pp. 235–250, 2020.

[392] R. Bommasani *et al.*, "The foundation model transparency index," *arXiv preprint arXiv:2310.12941*, 2023.

[393] R. J. Chen *et al.*, "Algorithmic fairness in artificial intelligence for medicine and healthcare," *Nat. Biomed. Eng.*, vol. 7, no. 6, pp. 719–742, 2023.

[394] F. Motoki *et al.*, "More human than human: Measuring chatgpt political bias," *Available at SSRN 4372349*, 2023.

[395] V. Felkner *et al.*, "Winoqueer: A community-in-the-loop benchmark for anti-lgbtq+ bias in large language models," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2023, pp. 9126–9140.

[396] S. Gehman *et al.*, "Realtoxicityprompts: Evaluating neural toxic degeneration in language models," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 3356–3369.

[397] A. Wei *et al.*, "Jailbroken: How does llm safety training fail?" *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[398] K. Bærøe *et al.*, "How to achieve trustworthy artificial intelligence for health," *Bull. World Health Organ.*, vol. 98, no. 4, p. 257, 2020.

[399] P.-Y. Chen and C. Xiao, "Trustworthy ai in the era of foundation models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023.

[400] M. Dwyer-White *et al.*, "High reliability in healthcare," in *Patient Safety: A Case-based Innovative Playbook for Safer Care*. Springer, 2023, pp. 3–13.

[401] V. Rawte, A. Sheth, and A. Das, "A survey of hallucination in large foundation models," *arXiv preprint arXiv:2309.05922*, 2023.

[402] C. Li and J. Flanigan, "Task contamination: Language models may not be few-shot anymore," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 16, 2024, pp. 18 471–18 480.

[403] Y. Yao *et al.*, "Editing large language models: Problems, methods, and opportunities," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 10 222–10 240.

[404] J. Hoelscher-Obermaier *et al.*, "Detecting edit failures in large language models: An improved specificity benchmark," in *Findings of the Association for Computational Linguistics: ACL 2023*, 2023, pp. 11 548–11 559.

[405] M. Raghu *et al.*, "On the expressive power of deep neural networks," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2017, pp. 2847–2854.

[406] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent.*, 2020.

[407] Z. Liu *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2021, pp. 10 012–10 022.

[408] S. Zhao *et al.*, "Elements of chronic disease management service system: an empirical study from large hospitals in china," *Sci. Rep.*, vol. 12, no. 1, p. 5693, 2022.

[409] C. Chen *et al.*, "Deep learning on computational-resource-limited platforms: a survey," *Mob. Inf. Syst.*, vol. 2020, pp. 1–19, 2020.

[410] L. Deng *et al.*, "Model compression and hardware acceleration for neural networks: A comprehensive survey," *Proc. IEEE*, vol. 108, no. 4, pp. 485–532, 2020.

[411] N. Ding *et al.*, "Parameter-efficient fine-tuning of large-scale pre-trained language models," *Nat. Mach. Intell*, vol. 5, no. 3, pp. 220–235, 2023.

[412] E. Griffith, "The desperate hunt for the ai boom's most indispensable prize." *International New York Times*, pp. NA–NA, 2023.

[413] U. Gupta *et al.*, "Chasing carbon: The elusive environmental footprint of computing," in *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. IEEE, 2021, pp. 854–867.

[414] P. Henderson *et al.*, "Towards the systematic reporting of the energy and carbon footprints of machine learning," *J. Mach. Learn. Res.*, vol. 21, no. 1, pp. 10 039–10 081, 2020.

[415] A. Park *et al.*, "Deep learning–assisted diagnosis of cerebral aneurysms using the headxnet model," *JAMA network open*, vol. 2, no. 6, pp. e195 600–e195 600, 2019.

[416] D. F. Steiner *et al.*, "Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer," *Am. J. Surg. Pathol.*, vol. 42, no. 12, p. 1636, 2018.

[417] H.-E. Kim *et al.*, "Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study," *The Lancet Digital Health*, vol. 2, no. 3, pp. e138–e148, 2020.

[418] P. Tschandl *et al.*, "Human–computer collaboration for skin cancer recognition," *Nat. Med.*, vol. 26, no. 8, pp. 1229–1234, 2020.

[419] Y. Han *et al.*, "Dynamic neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7436–7456, 2021.

[420] A. Vaswani *et al.*, "Attention is all you need," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[421] N. Shazeer *et al.*, "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," in *Proc. Int. Conf. Learn. Represent.*, 2016.

[422] C. You *et al.*, "Implicit anatomical rendering for medical image segmentation with stochastic experts," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 561–571.

[423] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[424] H. Yi, Z. Qin, Q. Lao, W. Xu, Z. Jiang, D. Wang, S. Zhang, and K. Li, "Towards general purpose medical ai: Continual learning medical foundation model," *arXiv preprint arXiv:2303.06580*, 2023.

[425] T. Kojima *et al.*, "Large language models are zero-shot reasoners," *Adv. Neur. In.*, vol. 35, pp. 22 199–22 213, 2022.

[426] Q.-F. Wang *et al.*, "Learngene: From open-world to your learning task," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 8, 2022, pp. 8557–8565.

[427] Y. Tan *et al.*, "Federated learning from pre-trained models: A contrastive learning approach," *Adv. Neur. In.*, vol. 35, pp. 19 332–19 344, 2022.

[428] W. Zhuang *et al.*, "When foundation model meets federated learning: Motivations, challenges, and future directions," *arXiv preprint arXiv:2306.15546*, 2023.

[429] J. Zhu *et al.*, "Uni-perceiver-moe: Learning sparse generalist models with conditional moes," *Adv. Neur. In.*, vol. 35, pp. 2664–2678, 2022.

[430] C. Geng *et al.*, "Recent advances in open set recognition: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3614–3631, 2020.

[431] Y. Li *et al.*, "Scaling language-image pre-training via masking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 23 390–23 400.

[432] M. Ma *et al.*, "Are multimodal transformers robust to missing modality?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 18 177–18 186.

[433] M. Mohri *et al.*, *Foundations of machine learning*. MIT press, 2018.

[434] Y. Yuan, "On the power of foundation models," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2023, pp. 40 519–40 530.

[435] J. Jiménez-Luna *et al.*, "Drug discovery with explainable artificial intelligence," *Nat. Mach. Intell*, vol. 2, no. 10, pp. 573–584, 2020.

[436] A. Qayyum *et al.*, "Secure and robust machine learning for healthcare: A survey," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 156–180, 2020.

[437] C. Schlarmann and M. Hein, "On the adversarial robustness of multi-modal foundation models," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2023, pp. 3677–3685.

[438] I. Habli *et al.*, "Artificial intelligence in health care: accountability and safety," *Bull. World Health Organ.*, vol. 98, no. 4, p. 251, 2020.

[439] R. Vinuesa *et al.*, "The role of artificial intelligence in achieving the sustainable development goals," *Nat. Commun.*, vol. 11, no. 1, pp. 1–10, 2020.

[440] L. H. Kaack *et al.*, "Aligning artificial intelligence with climate change mitigation," *Nat. Clim. Change*, vol. 12, no. 6, pp. 518–527, 2022.

[441] G. Menghani, "Efficient deep learning: A survey on making deep learning models smaller, faster, and better," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–37, 2023.