Mamba-Based Multimodal Learning for Automated Tumor Characterization and Clinical Decision Support in Precision Oncology

Chatresh Nagi
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
chatreshn@uc.edu

Nitesh Karan
Department of Computer Science, George Mason University, Fairfax, VA, USA.
niteshkaran@gmu.edu

Maritta Dowerg
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
marittadowerg@oregonstate.edu

Abstract

Precision oncology requires the synthesis of heterogeneous clinical evidence, including radiological imaging, digital pathology, genomic signatures, laboratory trends, treatment histories, and longitudinal outcomes. Although multimodal artificial intelligence has demonstrated promise for tumor characterization and clinical decision support, many existing architectures remain computationally expensive, difficult to govern, or poorly aligned with real-world clinical workflows. Mamba-based state space models have recently emerged as a compelling alternative because they can model long-range dependencies with linear-time sequence processing while preserving selective information routing. This paper examines the system-level implications of Mamba-based multimodal learning for automated tumor characterization and clinical decision support in precision oncology. It analyzes architectural trade-offs among transformers, convolutional networks, and selective state space models; discusses data infrastructure, interoperability, and secure deployment; and evaluates human factors, robustness, fairness, sustainability, and policy considerations. The central argument is that Mamba-based multimodal systems can reduce computational barriers to longitudinal oncology intelligence, but their clinical value depends less on isolated predictive accuracy than on integration with tumor board workflows, regulatory oversight, continuous monitoring, and equitable data governance. The paper concludes that precision oncology needs hybrid, auditable, and lifecycle-managed multimodal platforms rather than single-model solutions.

Keywords

Mamba, multimodal learning, precision oncology, clinical decision support, tumor characterization, governance, health equity, deployment.

1. Introduction

Precision oncology has moved from a predominantly histology-driven discipline toward an increasingly data-intensive practice in which diagnosis, prognosis, and treatment selection depend on multiple evidence streams. Radiological imaging describes anatomical extent and

treatment response, digital pathology captures cellular morphology and microenvironmental organization, genomic assays identify actionable mutations and resistance mechanisms, and clinical narratives preserve nuanced patient histories that are difficult to reduce to structured variables. The promise of artificial intelligence in this setting is not merely automated detection but the construction of a clinically meaningful representation of disease that can support oncologists, pathologists, radiologists, and patients in shared decision-making. Yet the field remains fragmented. Most models are trained on single modalities, evaluated on narrow endpoints, and deployed without the infrastructure needed for continuous oversight. This fragmentation limits generalizability and creates a gap between algorithmic performance and clinical usefulness [6], [7].

Multimodal learning has been proposed as a response to this fragmentation. By combining imaging, text, and genomics, multimodal systems can approximate the integrative reasoning of tumor boards and may detect patterns that are invisible in any single data source [5], [10]. Early successes in dermatology, breast cancer screening, and computational pathology demonstrated that deep learning can achieve expert-level performance on selected tasks [8], [9]. However, many of these systems rely on transformer-based architectures whose computational cost grows rapidly with sequence length and image resolution [1], [2], [19]. This cost becomes especially problematic when models must process whole-slide images, longitudinal radiology series, long clinical notes, and genomic panels simultaneously. The resulting infrastructure burden can exclude smaller hospitals, reduce update frequency, and increase the carbon footprint of clinical artificial intelligence.

Mamba-based state space models offer a different structural trade-off. Rather than applying attention across every pair of tokens, selective state space models maintain a recurrent state that is updated selectively as new information arrives [3], [4]. This design supports long-sequence modeling with linear-time scaling and has been adapted to vision, medical image segmentation, and multimodal representation learning [16], [17], [18]. For precision oncology, the implication is significant: a Mamba-based multimodal system could process long pathology sequences, longitudinal imaging, and narrative histories with lower memory and compute requirements than comparable transformer systems. Yet architectural efficiency alone does not ensure clinical safety. The central challenge is system-level: how to collect, curate, validate, deploy, monitor, and govern multimodal oncology models across heterogeneous institutions.

This paper argues that Mamba-based multimodal learning should be understood as a platform technology rather than a standalone algorithm. Its clinical value depends on data infrastructure, human-computer interaction, regulatory pathways, fairness auditing, and sustainable lifecycle management. Section 2 reviews the clinical and technical context of tumor characterization. Section 3 examines Mamba architectures and multimodal fusion. Section 4 discusses system architecture and data infrastructure. Section 5 analyzes clinical decision support and workflow integration. Section 6 addresses robustness, fairness, and governance. Section 7 considers deployment, sustainability, and policy implications. Section 8 concludes.

2. Clinical and Technical Context for Multimodal Tumor Characterization

Tumor characterization in precision oncology is inherently multimodal. A single patient may generate computed tomography, magnetic resonance imaging, positron emission tomography, whole-slide pathology images, immunohistochemistry, next-generation sequencing reports, laboratory panels, medication histories, and unstructured progress notes. Each modality has

different spatial and temporal scales. Imaging is volumetric and often longitudinal; pathology is gigapixel and hierarchical; genomics is high-dimensional but sparse; clinical notes are narrative and context-dependent. A clinically useful model must reconcile these differences without collapsing important distinctions. For example, a radiomic feature associated with tumor heterogeneity may mean something different when paired with a specific mutational profile than when paired with an inflammatory pathology pattern. Multimodal fusion therefore requires not only technical alignment but also clinical semantics.

Transformer architectures have advanced multimodal learning by enabling flexible attention across modalities [1], [2], [11]. Vision-language models such as CLIP demonstrated that contrastive pretraining can align images and text at scale [11], while clinical language models such as ClinicalBERT showed the value of domain-specific pretraining on medical narratives [12]. Large language models have further demonstrated encoded clinical knowledge and reasoning potential [13]. Foundation models for medicine have been proposed as general-purpose systems that can be adapted to many downstream tasks [14], and retinal foundation models have shown that large-scale pretraining can improve generalization [15]. These developments suggest that oncology artificial intelligence will increasingly rely on pretrained multimodal backbones rather than task-specific models trained from scratch.

Nevertheless, oncology presents conditions that stress conventional architectures. Whole-slide images can contain tens of thousands of patches; radiology series can include hundreds of slices over multiple time points; clinical histories can span years; genomic panels can include hundreds of genes and variants. Attention mechanisms that compare all tokens to all other tokens become computationally expensive and memory-intensive. This cost is not merely an engineering inconvenience. It affects the frequency of model updates, the feasibility of on-premises deployment, the ability to run uncertainty estimation, and the capacity to audit model behavior. Hospitals with limited computational resources may be unable to adopt large multimodal systems, thereby widening the gap between academic medical centers and community oncology practices. A more efficient architecture could reduce this gap, but only if it is embedded in a broader system of validation and governance [6], [7].

Clinical decision support in oncology is also different from many other medical artificial intelligence tasks. The decisions are high-stakes, often irreversible, and shaped by patient preferences, comorbidities, and goals of care. A model that predicts survival or treatment response does not automatically produce a recommendation. It must be contextualized within guidelines, multidisciplinary discussion, and the patient's values. Tumor boards are social and epistemic institutions as much as technical workflows. Automated tumor characterization can support them by summarizing evidence, highlighting discordance, and prioritizing cases for review, but it can also introduce automation bias, obscure uncertainty, or shift responsibility away from clinicians. The design of decision support must therefore account for cognitive load, explainability, and accountability [7], [20].

3. Mamba Architectures and Multimodal Representation Learning

Mamba-based models derive from state space sequence modeling, in which a latent state is updated as a sequence is processed. The selective mechanism allows the model to decide which information to retain, propagate, or forget, making it more adaptive than earlier linear recurrent models [3]. Structured state space duality further connects these models to attention-like formulations while preserving efficient computation [4]. In vision, Mamba variants have been adapted to image classification, segmentation, and detection by scanning images along multiple directions or hierarchies [18]. In medical image segmentation, UNet-like Mamba

designs have shown competitive performance with lower computational overhead [16], [17]. These developments are relevant to oncology because tumor characterization often requires long-range spatial and temporal dependency modeling.

From a system perspective, the appeal of Mamba is not that it replaces every other architecture. Convolutional neural networks remain strong for local texture and edge detection, transformers remain powerful for global semantic alignment, and graph neural networks are useful for molecular and pathway relationships [18], [19]. The more plausible architecture is hybrid. A multimodal oncology platform might use convolutional encoders for local pathology and radiology features, Mamba blocks for long-sequence integration across slides, slices, and time points, and transformer components for cross-modal alignment between images, text, and genomics. The required reference [16] illustrates how dynamic scanning and UNet-style medical segmentation can be adapted within this broader design space. In a multimodal setting, such scanning mechanisms could be extended to clinical narratives and longitudinal laboratory sequences, but this extension must be validated carefully because medical sequences are not homogeneous.

Multimodal fusion can occur at different levels. Early fusion combines raw or lightly processed features, which preserves fine-grained interactions but is sensitive to missing modalities and scale differences. Intermediate fusion aligns modality-specific representations before joint reasoning, which is often more practical in clinical settings. Late fusion combines modality-specific predictions, which improves interpretability but may miss cross-modal interactions. Mamba-based models can contribute at each level. At the intermediate level, selective state updates can maintain a shared latent representation that is updated as new evidence arrives, which mirrors the sequential way clinicians revise diagnostic impressions. This is particularly relevant for treatment response monitoring, where a model must update its characterization as imaging, pathology, and laboratory results change over time.

Despite these advantages, Mamba-based multimodal learning faces unresolved issues. Selective state space models can be difficult to interpret because the retained state is distributed and not easily mapped to clinical concepts. Their performance under domain shift, missing modalities, and irregular sampling has not been fully characterized in oncology. They may also inherit biases from pretraining data and institutional workflows. A model that efficiently processes long sequences can still produce confident but wrong outputs if the data are biased or the task is mis-specified. Therefore, the evaluation of Mamba-based systems must include calibration, uncertainty estimation, subgroup performance, and failure-mode analysis. Architectural efficiency should be treated as an enabling condition for rigorous validation, not as a substitute for it [5], [14].

4. System Architecture and Data Infrastructure

A clinical multimodal oncology platform requires more than a model. It needs a data infrastructure that can ingest, de-identify, harmonize, store, and version heterogeneous records. Imaging systems use DICOM, pathology uses whole-slide image formats, genomics uses variant call formats, and electronic health records use a mixture of structured codes and unstructured notes. Interoperability standards such as FHIR and DICOM provide partial solutions, but oncology workflows often rely on proprietary systems and manual reconciliation. A Mamba-based platform must be designed around these realities. It should not assume that all modalities are present, aligned, or equally reliable. Missingness is a clinical signal, not merely a preprocessing problem. A patient may lack genomic testing

because of cost, tissue availability, or physician discretion, and that absence can correlate with socioeconomic status and access to care.

The infrastructure should therefore include a multimodal data layer that preserves provenance, timestamps, and uncertainty. Each feature should be traceable to its source, acquisition parameters, and preprocessing steps. Versioning is essential because models may be updated while patient data continue to evolve. Reproducibility requires the ability to reconstruct the exact data and model state used for a prior prediction. This is difficult in many machine learning systems, but it is necessary for clinical accountability [6], [14]. A Mamba-based system may be more efficient than a transformer system, but it still requires substantial storage for images, sequences, and model checkpoints. Cloud, on-premises, and hybrid deployments each have trade-offs in cost, privacy, latency, and control. Hospitals may prefer on-premises inference for sensitive data, while training may occur in secure cloud environments. Federated learning can reduce data movement, but it introduces coordination costs and heterogeneity challenges.

Data governance must also address consent, secondary use, and equity. Oncology data are highly sensitive, and patients may be identifiable even after de-identification when genomic and imaging data are combined [6]. Governance frameworks should specify who can access data, for what purposes, and under what oversight. They should also ensure that the benefits of research and deployment are shared with the communities that contribute data. A platform that improves tumor characterization only for patients in well-resourced academic centers would reproduce existing inequities. Therefore, infrastructure design must include mechanisms for external validation, community engagement, and transparent reporting. The FAIR principles for data management provide a useful foundation, but clinical artificial intelligence requires additional attention to consent, privacy, and distributive justice [6], [20].

5. Clinical Decision Support, Human Factors, and Workflow Integration

Automated tumor characterization can support clinical decision-making in several ways. It can summarize multimodal evidence before a tumor board, flag discordance between radiology and pathology, estimate prognosis, suggest relevant clinical trials, or monitor treatment response over time. However, these functions are only useful if they fit into existing workflows. Oncologists are already burdened by documentation, prior authorizations, and fragmented information systems. A decision support tool that adds alerts without reducing cognitive load may be ignored or disabled. A tool that presents a risk score without explaining its basis may be trusted too much or too little. Human factors research shows that trust in automation is shaped by transparency, reliability, and perceived control. A Mamba-based system should therefore provide concise explanations, confidence estimates, and links to source evidence rather than a single opaque prediction [7], [14].

Explainability in multimodal oncology is difficult. Post hoc methods can highlight image regions or text snippets, but they may not explain how modalities interacted. For example, a model might rely on a spurious correlation between a radiology report phrase and a treatment outcome rather than on the tumor's biological features. State space models add further interpretability challenges because their internal state is not easily decomposed into human-readable concepts. One practical approach is to design the system as a set of modular components with separate outputs for each modality, followed by a fusion layer whose behavior is audited. Another approach is to use counterfactual explanations that show how the prediction changes when a specific modality is removed or altered. Neither approach is perfect, but both can support clinician judgment better than a single saliency map [20].

Workflow integration also requires attention to roles and responsibilities. A tumor board may include medical oncologists, surgical oncologists, radiation oncologists, pathologists, radiologists, genetic counselors, nurses, and research coordinators. Each has different information needs and different levels of tolerance for uncertainty. A decision support system should be configurable to these roles. It should also define escalation pathways when the model is uncertain or when it disagrees with the clinical team. Accountability cannot be delegated to the algorithm. If a model contributes to a treatment recommendation, the responsible clinician must understand its limitations and have the authority to override it. This means that training, credentialing, and documentation must evolve alongside the technology [7], [13].

6. Robustness, Fairness, and Governance

Robustness is a prerequisite for clinical deployment. Oncology data are subject to distribution shift across scanners, staining protocols, sequencing platforms, coding practices, and patient populations. A model trained at one institution may perform poorly at another even if it uses the same architecture. Mamba-based models are not immune to this problem. Their efficiency may allow more frequent retraining, but retraining without careful validation can introduce instability. Robustness testing should include external validation, stress tests with missing modalities, adversarial perturbations, and temporal validation in which the model is tested on data collected after its training period. Uncertainty estimation should be calibrated so that clinicians can distinguish confident predictions from uncertain ones [5], [14].

Fairness is equally important. Precision oncology has a history of disparities in screening, diagnosis, treatment access, and clinical trial enrollment. Algorithms trained on historical data can encode these disparities. For example, a model that predicts benefit from an expensive therapy may learn to recommend it less often for populations that historically received less care. A model that relies on documentation intensity may underrepresent patients with limited access to specialists. The well-known case of racial bias in a population health algorithm illustrates how a technically accurate model can produce inequitable outcomes when the target variable is a proxy for cost rather than need [20]. Multimodal oncology models must therefore be evaluated across demographic, socioeconomic, and geographic subgroups. They should also be assessed for intersectional disparities, because a patient may belong to multiple marginalized groups.

Governance must cover the entire lifecycle. Before deployment, models should undergo regulatory review appropriate to their risk level. In the United States, the Food and Drug Administration has developed pathways for software as a medical device, including predetermined change control plans for adaptive algorithms. In Europe, the Medical Device Regulation and the Artificial Intelligence Act create overlapping obligations. These frameworks vary in specificity, and oncology decision support may fall between device regulation, clinical decision support software, and research use. Institutions should establish local oversight committees that include clinicians, data scientists, ethicists, patient representatives, and legal experts. They should also require model facts labels that summarize intended use, training data, performance, limitations, and monitoring plans [7], [14], [20].

Monitoring after deployment is often neglected. A model that performs well in a validation study may degrade when clinical practice changes, when a new scanner is installed, or when a new treatment guideline is adopted. Continuous monitoring should track input data drift, output calibration, subgroup performance, and clinician override rates. It should also include mechanisms for reporting adverse events and near misses. Governance should not be a one-

time approval but a dynamic process. A Mamba-based platform can support this process by making model updates more computationally feasible, but it also increases the need for version control and audit trails. The goal is not to eliminate risk but to make it visible, bounded, and manageable [6], [14].

7. Deployment, Sustainability, and Policy Implications

Deployment choices have clinical, economic, and environmental consequences. Large multimodal models require substantial compute for training and inference. Transformer-based systems can be expensive to run at scale, especially when processing whole-slide images or long sequences. Mamba-based architectures may reduce energy consumption and latency, which is important for sustainability and for deployment in resource-constrained settings [3], [16]. However, efficiency claims must be evaluated in real clinical pipelines, not only in benchmark experiments. Preprocessing, data transfer, storage, and retraining can dominate the total cost. A system that is efficient at inference but requires frequent full retraining may still have a large footprint. Therefore, sustainability should be assessed across the lifecycle, including hardware procurement, cooling, and electronic waste.

Policy implications extend beyond environmental impact. Reimbursement models must align with value. If hospitals invest in multimodal infrastructure but are not reimbursed for improved characterization or reduced diagnostic delays, adoption will be uneven. Regulatory pathways must balance innovation with safety. Privacy laws must protect patients while enabling beneficial research. Competition policy must prevent a small number of vendors from controlling the data and models that underpin clinical care. International cooperation is needed because oncology data and patients cross borders, while privacy rules differ. A Mamba-based platform could be designed for modularity, allowing institutions to share pretrained components without sharing raw data. Federated learning, secure enclaves, and synthetic data are potential strategies, but each has limitations [6], [7], [14].

The most sustainable path is likely hybrid and incremental. Instead of replacing entire clinical workflows, Mamba-based multimodal systems can be introduced as focused tools for specific tasks, such as triaging pathology slides, summarizing tumor board evidence, or monitoring treatment response. Each tool should be evaluated prospectively, integrated into existing record systems, and governed by clear accountability structures. Over time, these tools can be connected into a learning health system in which clinical outcomes inform model updates and model outputs inform clinical decisions. Such a system would require standardized data collection, transparent reporting, and patient engagement. It would also require a workforce trained in both clinical oncology and data science. The technology is necessary but not sufficient. The limiting factor is often institutional capacity, not algorithmic novelty [5], [14].

8. Conclusion

Mamba-based multimodal learning offers a credible technical pathway for automated tumor characterization and clinical decision support in precision oncology. Its selective state space mechanisms can process long imaging, pathology, genomic, and narrative sequences with greater efficiency than many attention-based alternatives, making multimodal oncology intelligence more feasible for diverse deployment settings. Yet the clinical value of these models will be determined by system-level factors: data infrastructure, interoperability, validation, explainability, workflow fit, fairness, governance, and lifecycle monitoring. A model that is accurate in a benchmark but opaque, biased, or disconnected from tumor board practice will not improve care. A model that is efficient but poorly governed may scale

inequity more quickly. The appropriate research agenda is therefore interdisciplinary. It must combine architectural innovation with prospective clinical evaluation, participatory design, regulatory science, and sustainable infrastructure. Mamba-based multimodal systems should be developed as auditable platforms that support rather than replace clinical judgment. In precision oncology, the goal is not automation for its own sake but more accurate, equitable, and accountable decisions for patients facing complex cancer care.

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations.

3. Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. arXiv. https://arxiv.org/abs/2312.00752

4. Dao, T., & Gu, A. (2024). Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. Proceedings of the 41st International Conference on Machine Learning, 10041–10071.

5. Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. Nature Medicine, 28(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2

6. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

7. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

8. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056

9. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., ... Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

10. Chen, R. J., Lu, M. Y., Weng, W. H., Chen, T. Y., Williamson, D. F., Manz, T., Shady, M., & Mahmood, F. (2021). Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. Proceedings of the IEEE/CVF International Conference on Computer Vision, 4015–4025. https://doi.org/10.1109/ICCV48922.2021.00399

11. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning, 8748–8763.

12. Huang, K., Altosaar, J., & Ranganath, R. (2019). ClinicalBERT: Modeling clinical notes and predicting hospital readmission. arXiv. https://arxiv.org/abs/1904.05342

13. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gambhir, S. S., Nirschl, C., & Natarajan, V. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180. https://doi.org/10.1038/s41586-023-06291-2

14. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259–265. https://doi.org/10.1038/s41586-023-05881-4

15. Zhou, Y., Chia, M. A., Wagner, S. K., Ayhan, M. S., Williamson, D. J., Struyven, R. R., Liu, T., Xu, M., Lozano, M. G., Woodward-Court, P., Kihara, Y., Allen, N., Alexander, D. C., Keane, P. A., & Denniston, A. K. (2023). A foundation model for generalizable disease detection from retinal images. Nature, 622(7981), 156–163. https://doi.org/10.1038/s41586-023-06555-x

16. Zhang, Z., Ma, Q., Zhang, T., Chen, J., Zheng, H., & Gao, W. (2025). Switch-UMamba: Dynamic scanning vision Mamba UNet for medical image segmentation. Medical Image Analysis, 103792.

17. Wang, Z., Zheng, J. Q., Zhang, Y., Cui, G., & Li, L. (2024). Mamba-UNet: UNet-like pure visual Mamba for medical image segmentation. arXiv. https://arxiv.org/abs/2402.05079

18. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., & Liu, Y. (2024). VMamba: Visual state space model. arXiv. https://arxiv.org/abs/2401.10166

19. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00986

20. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453. https://doi.org/10.1126/science.aax2342