Highlights of GAO-22-104629, a report to congressional requesters

September 2022

Why GAO did this study

Diagnostic errors affect more than 12 million Americans each year, with aggregate costs likely in excess of $100 billion, according to a report by the Society to Improve Diagnosis in Medicine. ML, a subfield of artificial intelligence, has emerged as a powerful tool for solving complex problems in diverse domains, including medical diagnostics. However, challenges to the development and use of machine learning technologies in medical diagnostics raise technological, economic, and regulatory questions.

GAO was asked to conduct a technology assessment on the current and emerging uses of machine learning in medical diagnostics, as well as the challenges and policy implications of these technologies. This report discusses (1) currently available ML medical diagnostic technologies for five selected diseases, (2) emerging ML medical diagnostic technologies, (3) challenges affecting the development and adoption of ML technologies for medical diagnosis, and (4) policy options to help address these challenges.

GAO assessed available and emerging ML technologies; interviewed stakeholders from government, industry, and academia; convened a meeting of experts in collaboration with the National Academy of Medicine; and reviewed reports and scientific literature. GAO is identifying policy options in this report.

View GAO-22-104629. For more information, contact Karen L. Howard at (202) 512-6888 or howardk@gao.gov.

TECHNOLOGY ASSESSMENT

ARTIFICIAL INTELLIGENCE IN HEALTH CARE

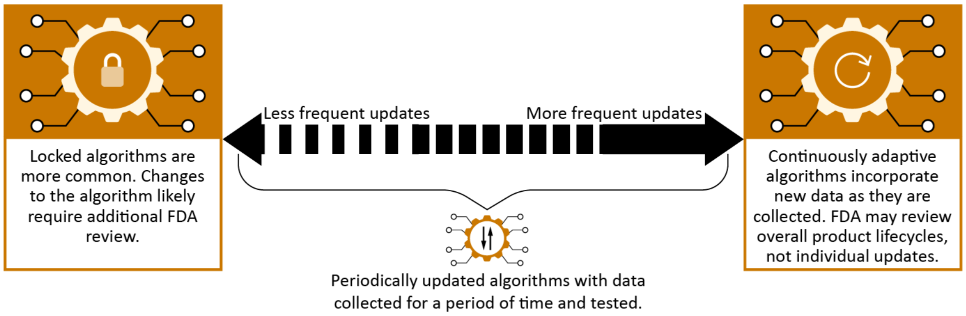
Benefits and Challenges of Machine Learning Technologies for Medical Diagnostics

What GAO found

Several machine learning (ML) technologies are available in the U.S. to assist with the diagnostic process. The resulting benefits include earlier detection of diseases; more consistent analysis of medical data; and increased access to care, particularly for underserved populations. GAO identified a variety of ML-based technologies for five selected diseases — certain cancers, diabetic retinopathy, Alzheimer's disease, heart disease, and COVID-19 —with most technologies relying on data from imaging such as x-rays or magnetic resonance imaging (MRI). However, these ML technologies have generally not been widely adopted.

Academic, government, and private sector researchers are working to expand the capabilities of ML-based medical diagnostic technologies. In addition, GAO identified three broader emerging approaches—autonomous, adaptive, and consumer-oriented ML-diagnostics—that can be applied to diagnose a variety of diseases. These advances could enhance medical professionals' capabilities and improve patient treatments but also have certain limitations. For example, adaptive technologies may improve accuracy by incorporating additional data to update themselves, but automatic incorporation of low-quality data may lead to inconsistent or poorer algorithmic performance.

Spectrum of adaptive algorithms



Source: GAO analysis of Food and Drug Administration (FDA) information. | GAO-22-104629

We identified several challenges affecting the development and adoption of ML in medical diagnostics:

- Demonstrating real-world performance across diverse clinical settings and in rigorous studies.
- Meeting clinical needs, such as developing technologies that integrate into clinical workflows.
- Addressing regulatory gaps, such as providing clear guidance for the development of adaptive algorithms.

These challenges affect various stakeholders including technology developers, medical providers, and patients, and may slow the development and adoption of these technologies.

GAO developed three policy options that could help address these challenges or enhance the benefits of ML diagnostic technologies. These policy options identify possible actions by policymakers, which include Congress, federal agencies, state and local governments, academic and research institutions, and industry. See below for a summary of the policy options and relevant opportunities and considerations.

Policy Options to Help Address Challenges or Enhance Benefits of ML Diagnostic Technologies

	Opportunities	Considerations
Evaluation (report page 28) Policymakers could create incentives, guidance, or policies to encourage or require the evaluation of ML diagnostic technologies across a range of deployment conditions and demographics representative of the intended use. *This policy option could help address the challenge of demonstrating real world performance.*	• Stakeholders could better understand the performance of these technologies across diverse conditions and help to identify biases, limitations, and opportunities for improvement. • Could inform providers' adoption decisions, potentially leading to increased adoption by enhancing trust. • Information from evaluations can help inform the decisions of policymakers, such as decisions about regulatory requirements.	• May be time-intensive, which could delay the movement of these technologies into the marketplace, potentially affecting patients and professionals who could benefit from these technologies. • More rigorous evaluation will likely lead to extra costs, such as direct costs for funding the studies. Developers may not be incentivized to conduct these evaluations if it could show their products in a negative light, so policymakers could consider whether evaluations should be conducted or reviewed by independent parties, according to industry officials.
Data Access (report page 29) Policymakers could develop or expand access to high-quality medical data to develop and test ML medical diagnostic technologies. Examples include standards for collecting and sharing data, creating data commons, or using incentives to encourage data sharing. *This policy option could help address the challenge of demonstrating real world performance.*	• Developing or expanding access to high-quality datasets could help facilitate training and testing ML technologies across diverse and representative conditions. This could improve the technologies' performance and generalizability, help developers understand their performance and areas for improvement, and help to build trust and adoption in these technologies. • Expanding access could enable developers to save time in the development process, which could shorten the time it takes for these technologies to be available for adoption.	• Entities that own data may be reluctant to share them for a number of reasons. For example, these entities may consider their data valuable or proprietary. Some entities may also be concerned about the privacy of their patients and the intended use and security of their data. • Data sharing mechanisms may be of limited use to researchers and developers depending on the quality and interoperability of these data, and curating and storing data could be expensive and may require public and private resources.
Collaboration (report page 30) Policymakers could promote collaboration among developers, providers, and regulators in the development and adoption of ML diagnostic technologies. For example, policymakers could convene multidisciplinary experts together in the design and development of these technologies through workshops and conferences. *This policy option could help address the challenges of meeting medical needs and addressing regulatory gaps.*	• Collaboration between ML developers and providers could help ensure that the technologies address clinical needs. For example, collaboration between developers and medical professionals could help developers create ML technologies that integrate into medical professionals' workflows, and minimize time, effort, and disruption. • Collaboration among developers and medical providers could help in the creation and access of ML ready data, according to NIH officials.	• As previously reported, providers may not have time to both collaborate with developers and treat patients; however, organizations can provide protected time for employees to engage in innovation activities such as collaboration. • If developers only collaborate with providers in specific settings, their technologies may not be usable across a range of conditions and settings, such as across different patient types or technology systems.

Source: GAO. | GAO-22-104629