Executive summary

Artificial intelligence for genomic medicine

Artificial intelligence (AI) techniques offer great potential for advancing genomic medicine. Focusing on the intersection between these two technologies, *Artificial intelligence for genomic medicine* investigates the drivers behind the recent rise of AI techniques for genomics, existing and emerging applications, the limitations of AI for genomic medicine, and the challenges to realising its full potential for health.

Achieving this potential will necessitate meeting the level of current enthusiasm for these technologies with the impetus, resources and collective commitment to tackle the serious issues ahead.

The report sets out practical recommendations for policy makers wanting to make the most of the opportunities AI presents for genomic medicine, minimise harms and speed up its effective delivery into healthcare.

The data challenge for genomic medicine

Genomic medicine has made significant strides in recent years, but the clinical application of genomics continues to evolve as new knowledge and technologies emerge. One major challenge is the ability to make sense of extremely large volumes of genomic sequence data, and effectively integrate and examine it with other relevant information, for example other molecular or clinical data.

The rise of AI

The AI techniques machine learning and deep learning (a type of machine learning) offer new computational approaches to streamlining key analytical problems in genomic medicine. Although some machine learning methods have been applied to key problems in genomic analysis for many years, activity of this kind has been increasing recently, driven by:

- Advances in high-performance computing
- Resurgence of deep learning
- Growing availability of resources for building machine learning models
- Growth of large genomic and biomedical datasets

UNIVERSITY OF CAMBRIDGE

Applications of AI in genomic medicine

Most aspects of genomic analysis have been touched in some way by machine learning and deep learning. These methods are being developed and applied across different elements of the genomic data pipeline, and to a whole spectrum of analyses, from single cell resolution to studies in large populations.

These efforts offer a significant range of potential benefits that could help advance the clinical application of genomics by:

- **Directly facilitating the steps involved in clinical genome analysis**

Examples of current activity include:

- Algorithms for better identification of genetic variants, including those that are currently difficult to accurately detect, e.g. somatic and copy-number variants
- Tools for extracting phenotype data (patient characteristics) from electronic health records, or analysing it e.g. deep-learning driven facial analysis to help inform the diagnosis of congenital conditions
- Tools for predicting the effect of genetic variants, such as their downstream impact on proteins or important molecular processes, e.g. gene expression

- **Improving understanding of genomic variation in relation to health and disease and accelerating discovery in genomic medicine**

We are still far from a complete understanding of the relationship between genomic variation and many known diseases; AI techniques applied to complex or very large datasets can provide valuable insight, and improve the underlying knowledge base upon which clinical genomic analysis relies. Research underway includes:

- Studies to examine how cancers evolve and determine which genetic changes could be drivers for tumour growth
- Algorithms to improve the efficiency and accuracy of CRISPR, a genome editing technique widely used to investigate the role of genes and other DNA sequences
- Methods to integrate and analyse genomic data together with other types of data

Current limitations

The application of AI has yet to generate clearly improved outcomes in genomic medicine, and the discovery potential within genomic datasets remains largely untapped. To make progress, multiple interconnected issues must be addressed:

- **Data quality and accessibility**

The performance of AI algorithms is affected by the volume and quality of data used to initially 'train' (i.e. develop) them, so streamlined access to high quality genomic and healthcare data is essential

Bias

Some populations are under-represented in the databases and datasets used for training AI algorithms. This has the potential to exacerbate existing health disparities for groups that are already underserved. Algorithmic bias can also arise as a result of the availability of data, how those data are prepared and combined, how questions are framed, and because of prexisting prejudices within society

Expectations

Replicating the methods and results of AI studies and tools can be difficult. The increasing number of AI based tools for various steps of genomic data analysis will only make this more challenging

Skills and infrastructure

AI in genomics is a multidisciplinary endeavour - no single sector has a monopoly on all the necessary skills, expertise, data and resources needed to deliver all the potential benefits of AI used at scale in genomic medicine, so a focus across multiple sectors is needed

Privacy and security

Concerns around security, confidentiality, and the ethical use of data must be navigated and addressed effectively, or there is a serious risk of impeding the use and implementation of these technologies

Regulation and clinical governance

The regulatory status of many AI algorithms used within clinical genomics remains unclear. This is influenced by whether or not the algorithm qualifies as a medical device or meets an unmet need. For adaptive algorithms, questions arise about the nature of the regulatory pathway, how they should be certified and who should be liable if their use results in harm

Uncertainty

Another area of uncertainty is how algorithms used for healthcare should meet the regulatory requirements for transparency and explanation within the EU General Data Protection Regulation. These requirements could impact on how algorithms are used for clinical decision making and patient management, particularly when using black box algorithms

Considering the significant financial investment and policy work already underway to deliver AI in health and care, it is vital to address the above priorities early as part of wider efforts to accelerate the adoption of proven AI technologies. In doing so the application of AI, when experts in health, genomics, regulation and ethics are working in concert, presents a significant opportunity to unravel the complexity encoded in our genomes for health benefit.

Priorities for policy

The initial priorities for creating an environment that facilitates the application of AI in genomic medicine and realises its near-term value are to:

Establish the right conditions for facilitating AI in genomic medicine which includes improved digital infrastructure, data acquisition and management, access to specific technical skills, and cross-disciplinary collaborations

Prioritise the development of constructive AI tools that address well-defined, focused, and clinically relevant problems in genomics analysis and clinical genomics service delivery

Mitigate against AI bias in genomics by promoting a workforce and research environment that is representative of societal diversity, as well as monitoring and addressing sources of bias within training datasets

Facilitate research efforts to apply machine learning (including deep learning) to well-curated, high-quality genomics and biomedical datasets, and bridging the gap between knowledge discovery and clinical practice

Support efforts driven by the clinical genomics community to benchmark, review, and determine the most effective use and integration of emerging new algorithms for clinical genome analysis

Establish sector-specific strategies to address the complex challenges and limitations of AI in genomic medicine and research

Establish the clinical governance arrangements for the use of specific AI applications in the practice of clinical genomics

Contact: intelligence@phgfoundation.org

PHG Foundation is a health policy think tank with a special focus on how genomics and other emerging health technologies can provide more effective, personalised healthcare

phg foundation
making science work for health