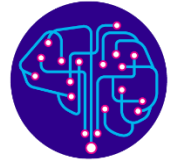


ARTIFICIAL INTELLIGENCE

Detect cancer earlier by interrogating medical and non-medical data sets using machine and deep-learning



CONTEXT

We know that, for almost all cancer types, patient outcomes are improved if the disease can be diagnosed at an early stage. Although we have made significant advances in diagnosing certain cancers early, unfortunately, late diagnosis remains an important problem; in England for example almost half of patients are diagnosed when their cancer is already advanced.

There are numerous barriers to the early detection of cancer. Some cancer types such as pancreatic cancer may display very few (or non-specific) symptoms until they are at an advanced stage. For other cancer types delays may be due to patients not being aware of or not reporting cancer symptoms, or health practitioners having difficulty identifying cancers in people presenting with vague or non-red flag symptoms.

There may, however, be patterns of symptoms and behaviours within accessible data sets that could be used to indicate the presence of a cancer. These data sets may be medical (e.g. GP presentation patterns, prescription records, health insurance claims) or non-medical (e.g. social media activity, shopping history, online search history). There is an opportunity to employ deep-learning approaches to combine these data sets with other cancer risk factors, and devise methods to drive diagnostic investigation at an earlier stage and facilitate the early detection of cancer.

OPPORTUNITIES AND BARRIERS

This Grand Challenge requires the collation and interrogation of both medical and non-medical data sets; this may include data sources that have not previously been explored for the purpose of early detection of cancer.

- It is envisaged that a number of factors may need to be considered:
- Collation of anonymised data sets from health records
- Collation of anonymised data sets relating to online search or social media activity
- Use of pattern recognition algorithms to decipher a set of patterns that can then be prospectively validated
- Development of machine learning approaches to optimise algorithms over time as new data is collected

Teams taking on this challenge will need to consider the ethical considerations surrounding the use of publically available and patient data.

VISION AND IMPACT

The goal of this Grand Challenge is to deliver interventions on a personal level that facilitate the early detection of cancer, particularly for those cancer types that currently present late and/or have poor survival rates. The collation of diverse data sets is also likely to result in more informed health information and more accurate cancer diagnoses.