Universidad de San Andrés

Escuela de Negocios

Master in Business and Technology

Artificial Intelligence applied to the early diagnosis and cure research of certain types of cancer

Student: Eng. Ariel Star

Thesis Advisor: Eng. Enrique Hofman

MASTER IN BUSINESS AND TECHNOLOGY

Thesis

Artificial Intelligence applied to the early diagnosis and cure research of certain types of cancer.

Student: Eng. Ariel Star
Thesis Advisor: Eng. Enrique Hofman

Contents

Abstract

The artificial intelligence has advanced in the last years up to arrive to the daily activities of everybody. Chatbots that attends clients and consumers questions in any web page, virtual assistants in cell phones and home devices are totally common today, self-driving vehicles are not a surprise now. In the medicine field, the artificial intelligence is taken a more relevant presence each day helping to medical centers and specialists to achieve more precise diagnosis using the most powerful information processing techniques in conjunction with big platforms and the availability of quantities of data as never before, the Big Data.

In this work I will present the use of these technological advances in the cancer early diagnosis and as the way to a potential cure in the future. I will introduce to the specific technology under the artificial intelligence term called “Deep Learning” as the key factor by we can show then real cases in the early diagnosis of different types of cancer. The quick speed in that the technology, communications, and huge data availability are evolving is setting to the world, as never before, facing the possibility to arrive in the next years to a cure for one the most mortal diseases. Follow the initial introduction to the technological state of the art and a group of statistics that present the situation of cancer in the world, I’ll present a group of real cases of IA application in the early diagnoses of six different types of cancer with very hopeful results. Then, I describe how the AI is helping in the other side of the fight against cancer, the development of new drugs and the design of new clinical trials where the standard times to release a new drug to the market are around 10 to 15 years.

In the last chapter I’ll propose a global platform with the objective of collect diagnosis images and laboratory data from medical centers of any country with the objective to provide data to an AI’s engine that “learn” from them and help to diagnose early and precisely a wide range of types of cancer. This proposal is not a business case, is a high-level proposal based on the technological current possibilities and some real cases that are been testing right now.

Chapter I: Introduction, problem and objectives

Cancer is the second leading cause of death worldwide and the number of deaths from this disease increases year after year, the World Health Organization (WHO) warned that in the coming decades the new cases would increase to become almost a 50% higher in 2040. Science has made progress in the last three decades in the search for new therapies and drugs to treat the disease, although with disparate results depending on the type of cancer, age range, social position and gender of the patient; however, a widely accepted concept has been reached in the medical community: *the sooner the disease is diagnosed in the patient, the chances of applying a successful treatment increase considerably.*

But reaching an early and accurate diagnosis requires a series of elements that a large part of the world's population is not in a position to access, they are, access to the health system as a whole, which includes specialized health professionals, diagnosis and treatment centers and the corresponding medication. Additionally, as occurs with very few diseases, the mere mention of the word cancer causes a totally demoralizing psychological effect on the patient, so the psychological aspect also adds to the needs to be taken into account in this complex scenario.

Technology is changing many aspects of our daily lives in recent years, the convergence and massification of technologies such as big data, massive Internet access, high-speed communications, cloud computing, the economy of platforms, among others, have served to leverage the development of Artificial Intelligence (AI) coming today to be present in search and recommendation engines that suggest us which movies or songs we may like on entertainment platforms, personal assistants on our cell phones or in the main module of the autonomous vehicles that not only open up new markets and business opportunities of great value, and whose future potential is enormous given that technology giants such as IBM, Microsoft, Amazon, Google and Facebook invest millions of dollars in research and development in the AI kingpin to position himself in this market. But AI as such is a broad term that

encompasses a set of disciplines ranging from natural language processing (NLP), expert systems, autonomous robots, machine learning (ML) that specializes in making machines imitate human reasoning through “learning” techniques; and within this, a relatively new discipline called Deep Learning (DL) that takes the concept further based on techniques known as neural networks that allow learning of the model without assistance or "supervision" and is especially useful in the analysis of unstructured information as images, video, and data that do not have a predefined structure in its format such as information sent by sensors, cameras and mobile devices.

Health is not the exception to the rest of the areas of human activity and has also seen a growing technification, the appearance of new diagnostic imaging devices that have become widespread in the last 30 years such as magnetic resonance imaging, tomographs, ultrasound scanners and modern X-rays have provided health professionals with tools that allow them to be much more assertive in diagnosing patients than decades ago without the need to apply invasive techniques to them.

All these changes and transformations have occurred with greater or lesser speed in the last 25/30 years, with strong acceleration in the last 10 years.

Investigation question

Given the scenario of the reality of cancer in the world and the technological progress in general experienced by the humanity, but in particular that of AI, this paper aims to answer the following research question:

Can the application of AI, in particular deep learning models, help reduce deaths from certain types of cancer?

Objetives

The main objective of this research work is to analyze the possible contributions that AI can make to the early diagnosis of cancer in order to increase the patient's chances of being cured.

Based on the main objective, this research work has the following secondary objectives:

- Present the world situation of cancer, both in terms of deaths caused by the disease, its distribution by different types of segments, and in the economic aspect of those involved in the fight against cancer.
- Introduce the concepts of Artificial Intelligence (AI) as a technological concept and its multiple branches of application.
- Introduce the concept of Deep Learning, its principles and its area of applicability, since it will be the key technological factor of this thesis.
- To present a series of real cases where AI has been successfully applied to diagnose six different types of cancer early, in order to demonstrate that the research question is not a theoretical proposition but a reality that is already happening today.
- To present the advances that AI can bring to the research process for new cancer drugs, shortening the currently extensive timeframes involved in the process from initial laboratory tests to government approval for marketing.
- Propose a platform where the concepts developed previously can be developed as an AI product offered in the cloud where doctors from anywhere in the world can validate early the diagnostic results of their patients to initiate early treatments when necessary.

Chapter II: Investigation methodology

Roberto Hernández Sampieri in his book "Research Methodology" (2014) defines "Research is a set of systematic, critical and empirical processes that are applied to the study of a phenomenon or problem" and states that since the last century the currents of research have been polarized into two main approaches: the qualitative approach and the quantitative approach.

He defines the two approaches for us as follows:

The quantitative approach “is sequential and probative, each stage precedes the next and steps cannot be avoided. The order is rigorous, although of course, we can redefine some phase. It starts from an idea that is delimited and, once delimited, objectives and research questions are derived, the literature is reviewed and a theoretical framework or perspective is built. Hypotheses are established from the questions and variables are determined; a plan is drawn up to test them (design); variables are measured in a given context; the measurements obtained are analyzed using statistical methods, and a series of conclusions are drawn”.

The qualitative approach: “It is also guided by significant areas or themes of research. However, instead of clarity about research questions and hypotheses preceding data collection and analysis (as in most quantitative studies), qualitative studies may develop questions and hypotheses before, during, or after the data collection and analysis. of data collection and analysis. These activities often serve, first, to discover what the most important research questions are; and then to refine and answer them. The investigative action moves dynamically in both directions: between the facts and their interpretation, and it is a rather “circular” process in which the sequence is not always the same, since it varies with each study”.

Research Approach

Under these concepts mentioned, this research will be carried out under a qualitative approach because it will be based on data analysis, information, and publications, and generalizations will be made trying to discover patterns.

The knowledge generated by research under a qualitative approach is mainly formulated in terms of inductive reasoning. In this reasoning, the property observed in a finite number of cases is generalized for all the elements of a set. It must be taken into account that the observations favorable to the conclusions reached do not make the conclusion true, since there could be exceptions. Hence, the conclusion of an inductive reasoning can only be considered probable and, in fact, the information obtained through this mode of reasoning is always uncertain and debatable information. This paradigm establishes a subjective approach to the matter under study. The objective is more associated with the understanding than with the prediction of the phenomenon under study.

Research Type

According to the theory, there are four possible types of research: exploratory, descriptive, correlational, and explanatory.

In particular, the type of exploratory research is defined as research used to investigate a problem that is not clearly defined. It is done to have a better understanding of the existing problem, but it will not provide conclusive results. For such research, a researcher starts with a general idea and uses this research as a means of identifying problems, which can be the focus of future research. An important aspect here is that the researcher must be willing to change the direction of it subject to the revelation of new data or knowledge. This investigation is usually carried out when the problem is at a preliminary stage. It is often referred to as a grounded theory approach or interpretive research as it is used to answer questions such as what, why and how.

To carry out this thesis, an exploratory research approach will be adopted, and from the possible sources of information described by Zikmund (2000) we will select the secondary analysis of data and case studies.

Instruments

The topic addressed is extremely recent, and therefore the information on the theory of Deep Learning, the description of successful cases of Deep Laerning in the early diagnosis of cancer and in the investigation of new drugs was obtained from papers and digital magazines published in different scientific dissemination sites.

Additionally, for the justification and description of figures and statistics on which the statements and courses of action indicated in the following chapters of this work are justified, the data was obtained through market research sites and global statistics.

List of scientific dissemination sites used as sources for cases of AI applied to the early diagnosis of cancer in this research:

www.cancer.gov	Site dedicated to provide scientific cancer information.
www.cancercenter.ai	Site dedicated to provide scientific cancer information
www.aiononcology.org	Site about divulgation of information about the use of AI in oncology
www.frontiersin.org	Site about scientific information.
www.nature.com	Site about scientific information.
www.ncbi.nlm.nih.gov	Site about scientific information.
www.royalmarsden.nhs.uk	Site of medical center in UK.
www.researchgate.net	Site about scientific information.
www.cell.com	Site about scientific information.
www.jlgh.org	Site of Lancaster General Hospital
www.febs.onlinelibrary.wiley.com	Site about scientific information.
www.arxiv.org	Publication of Cornell University.

List of sites on global statistics and world organizations related to health:

www.who.int	World Health Organization web site.
www.paho.org	Pan-American health organization web site.
www.ourworldindata.org	Site about global statistics.

www.statista.com Site about global statistics.

www.idataresearch.com Site about global statistics.

List of sites on technology, AI and technology market research:

www.gartner.com Gartner Group, technology market investigation org. web site.

www.towardsdatascience.com Data Science discipline publication web site.

www.ibm.com IBM's web site.

www.enterprisersproject.com Technology investigation web site.

www.medium.com Technology investigation web site.

List of sites, books and study cases related to business and economy:

Book "INTERNATIONALIZATION HANDBOOK FOR THE SOFTWARE BUSINESS"

Harvard Business School.

www.weforum.com Economy investigation web site.

Thesis structure

The text of this thesis is divided into six chapters, of which the first two are dedicated to describing the methodological aspects of the work, the problem statement, the objectives, the question and the type of research, the instruments used and the structure of the research. the thesis.

Chapter III introduces the reality of cancer in the world with various statistical tables that allow a clear vision of the magnitude of the problem both at a human and economic level. An introduction is also made to artificial intelligence and in particular to Deep Learning technology as a key factor in the problems and solutions proposed in the following sections and chapters.

Chapter IV shows us six real cases of AI successfully applied to the early diagnosis of different types of cancer, also to histopathological diagnosis as a complement to cancer cases and finally a reference to the potential of AI applied to the case of COVID research. 19.

Chapter V complements the previous chapter by presenting how AI can help improve and reduce the extensive timeframes that today have the cycle of research and development of new drugs for cancer treatment, thus allowing not

only to be able to diagnose early the disease, but also accelerate the arrival of new drugs on the market.

The last chapter proposes how a platform implemented with the technologies mentioned throughout the work can allow the implementation of an AI engine that is trained with diagnostic images, laboratory results and medical literature from around the world. The platform with this AI engine would allow any doctor to validate the results of a patient's exams and quickly discover if there is a possibility of developing cancer in the patient.

Finally, I would like to mention that I have written this work in English since all the documents, papers, publications and reference sites are written in English. Attempting to translate into Spanish could distort technical terms that should be interpreted in the idiomatic context in which they were written.

Chapter III: Theoretical framework

In this chapter I'll introduce some concepts about the problem of cancer as the second global cause of death and the introduction to artificial intelligence (AI) and Deep Learning as a technology that is addressed the problem, obtaining currently results as never before in the disease early diagnosis and treatment. First, I'll introduce to the definition of cancer, I'll present statistics about the impact of the disease in the world and finally which is the economic cost that currently the society is paying today to manage all the aspects of the medical's treatments related. The second part of this chapter is dedicated to introduce the concepts of AI and how it is taking a more relevant aspect all the daily aspects of our lives, then I'll continue with the presentation of deep learning, one of the AI subsets of technologies and its use in the cancer early diagnosis.

Some words about cancer

It is known that the word cancer in itself carries a negative emotional connotation that activates all the alert systems of those who receive this difficult diagnosis, when the doctor says the word "cancer" it is a situation that people often fear, they often say that they were stunned when they heard the news and unable to process what was the doctor said afterwards.

According to the "World Health Organization" (WHO) cancer is a generic term for a large group of diseases that can affect any part of the body. Other terms used are malignant tumours and neoplasms. One defining feature of cancer is the rapid creation of abnormal cells that grow beyond their usual boundaries, and which can then invade adjoining parts of the body and spread to other organs; the latter process is referred to as metastasis. Widespread metastases are the primary cause of death from cancer. (WHO, 2022) (0).

Also, it is defined by the National Cancer Institute "as a collection of diseases in which abnormal cells can divide and spread to nearby tissue". Cancers can arise in many parts of the body – leading to a range of cancer types - and in some cases spread to other parts of the body through the blood and lymph systems. (Ritchie, 2019) (1)

Cancer arises from the transformation of normal cells into tumour cells in a multi-stage process that generally progresses from a pre-cancerous lesion to a malignant tumour. These changes are the result of the interaction between a person's genetic factors and three categories of external agents, including:

- Physical carcinogens, such as ultraviolet and ionizing radiation.
- Chemical carcinogens, such as asbestos, components of tobacco smoke, alcohol, aflatoxin (a food contaminant), and arsenic (a drinking water contaminant).
- Biological carcinogens, such as infections from certain viruses, bacteria, or parasites. (WHO, 2022) (0).

Progress against many other causes of deaths and demographic drivers of increasing population size, life expectancy and — particularly in higher-income countries — aging populations mean that the total number of cancer deaths continues to increase. This is a very personal topic to many: nearly everyone knows or has lost someone dear to them from this collection of diseases. (Ritchie, 2019) (1).

The incidence of cancer rises dramatically with age, most likely due to a build-up of risks for specific cancers that increase with age. The overall risk accumulation is combined with the tendency for cellular repair mechanisms to be less effective as a person grows older.

Between 30% and 50% of cancers can currently be prevented by avoiding risk factors and implementing existing evidence-based prevention strategies. **The cancer burden can also be reduced through early detection** and appropriate treatment and care of patients who develop cancer. **Many cancers have a high chance of cure if diagnosed early** and treated appropriately.

There are two components of early detection: **early diagnosis and screening**.

Early diagnosis

When identified early, cancer is more likely to respond to treatment and can result in a greater probability of survival with less morbidity, as well as less expensive treatment. Significant improvements can be made in the lives of cancer patients by detecting cancer early and avoiding delays in care.

Early diagnosis consists of three components:

- Being aware of the symptoms of different forms of cancer and of the importance of seeking medical advice when abnormal findings are observed;
- Access to clinical evaluation and diagnostic services; and
- Timely referral to treatment services.

Early diagnosis of symptomatic cancers is relevant in all settings and the majority of cancers. Cancer programs should be designed to reduce delays in, and barriers to, diagnosis, treatment and supportive care.

Screening

Screening aims to identify individuals with findings suggestive of a specific cancer or pre-cancer before they have developed symptoms. When abnormalities are identified during screening, further tests to establish a definitive diagnosis should follow, as should referral for treatment if cancer is proven to be present.

Screening programs are effective for some but not all cancer types and in general are far more complex and resource-intensive than early diagnosis as they require special equipment and dedicated personnel. Even when screening programs are established, early diagnosis programs are still necessary to identify those cancer cases occurring in people who do not meet the age or risk factor criteria for screening. (World, 2022) (0)

Some numbers about the impact of cancer in the world

Cancer is a leading cause of death worldwide, accounting for nearly 10 million deaths in 2020. (WHO, 2022) (0). Every sixth death in the world is due to cancer, making it the second leading cause of death – second only to cardiovascular diseases. (Ritchie, 2019) (1)

The most common in 2020 (in terms of new cases of cancer) were:

- Breast (2.26 million cases)
- Lung (2.21 million cases)
- Colon and rectum (1.93 million cases)
- Prostate (1.41 million cases)
- Skin (non-melanoma) (1.20 million cases)
- Stomach (1.09 million cases)

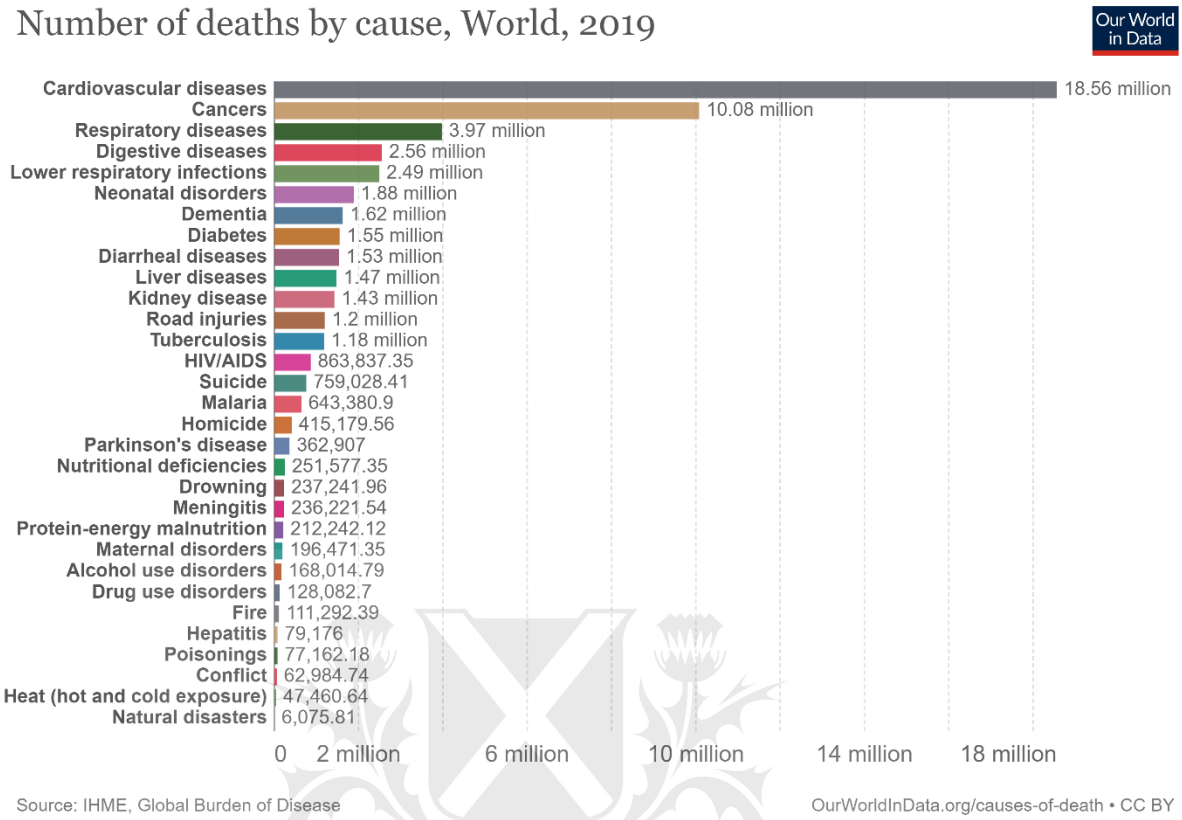
The most common causes of cancer death in 2020 were:

- Lung (1.80 million deaths);
- Colon and rectum (916.000 deaths);
- Liver (830.000 deaths);
- Stomach (769.000 deaths)
- Breast (685.000 deaths)

Each year, approximately 400.000 children develop cancer. The most common cancers vary between countries. Cervical cancer is the most common in 23 countries. (WHO, 2022) (0)

The *Global Burden of Disease* (a major global study on the causes and risk factors for death and disease published in the medical journal *The Lancet*) estimates that 9.56 million people died prematurely as a result of cancer in 2017. Every sixth death in the world is due to cancer.

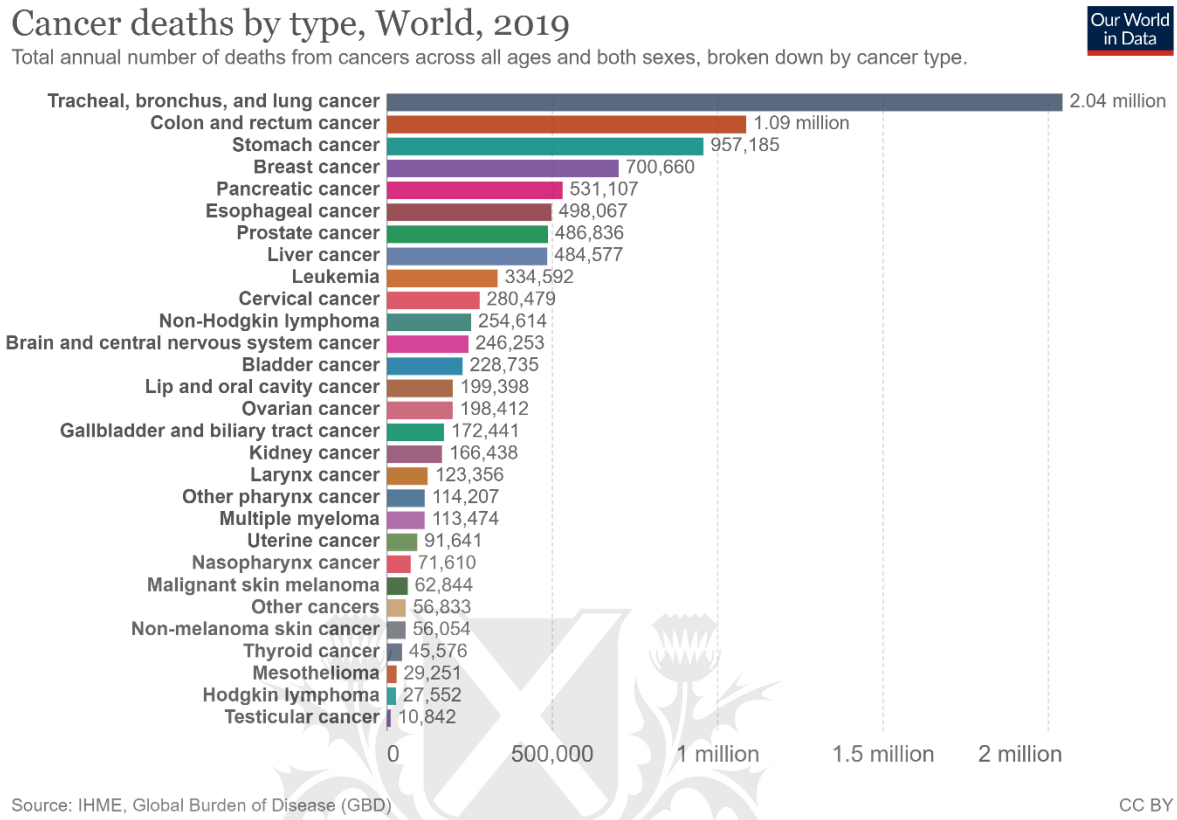
Cancer is a particularly common cause of death in richer countries where people are less likely to die of infectious diseases and causes of deaths that lead to very early deaths for people in poverty. Because cancer is one of the leading causes of death, it is one of the world's most pressing problems to make progress against this disease. Next chart shows the main causes of death in the world, with cancer as a second only behind of cardiovascular diseases:



With a clear picture of the problem's magnitude, we can go in details in order to know which are the different types of cancer that affect the people, and other distributions by geographic and age range. In the next chart we see the total number of deaths in 2019 attributed to the range of different cancers.

The group of tracheal, bronchus, and lung cancers claimed the largest number of lives – 1.9 million in 2019. Next follow colon and rectum, stomach and liver cancer, all claiming between 800,000 and 900,000 globally in 2019.

In exploring patterns across various countries, we see that tracheal, bronchus, and lung cancer is the leading form of cancer deaths across most high and middle-income countries. However, the leading form in lower income countries varies: colon and rectum; liver; cervical; stomach; breast and prostate all top the list in several countries.



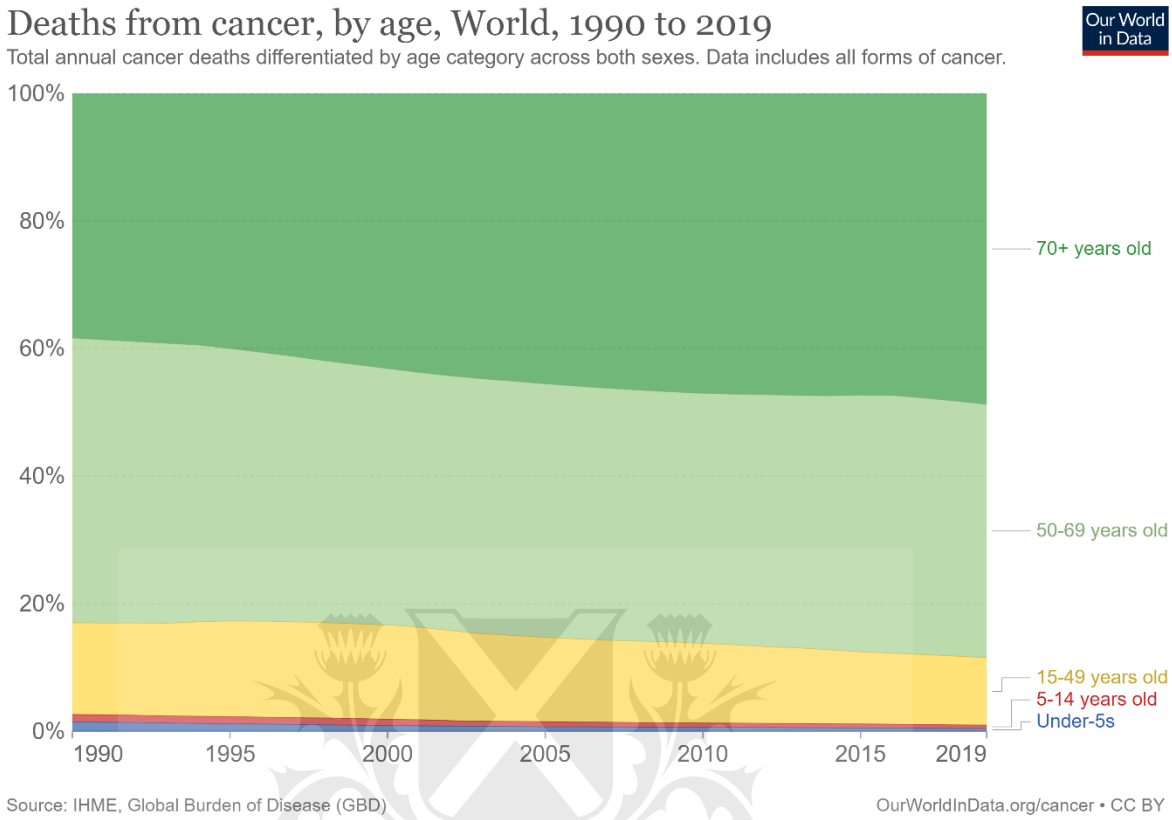
How are cancer deaths distributed across age groups? And how did this change over time?

In the next chart we see the breakdown of total cancer deaths by broad age category, ranging from under-5s to those over 70 years old.

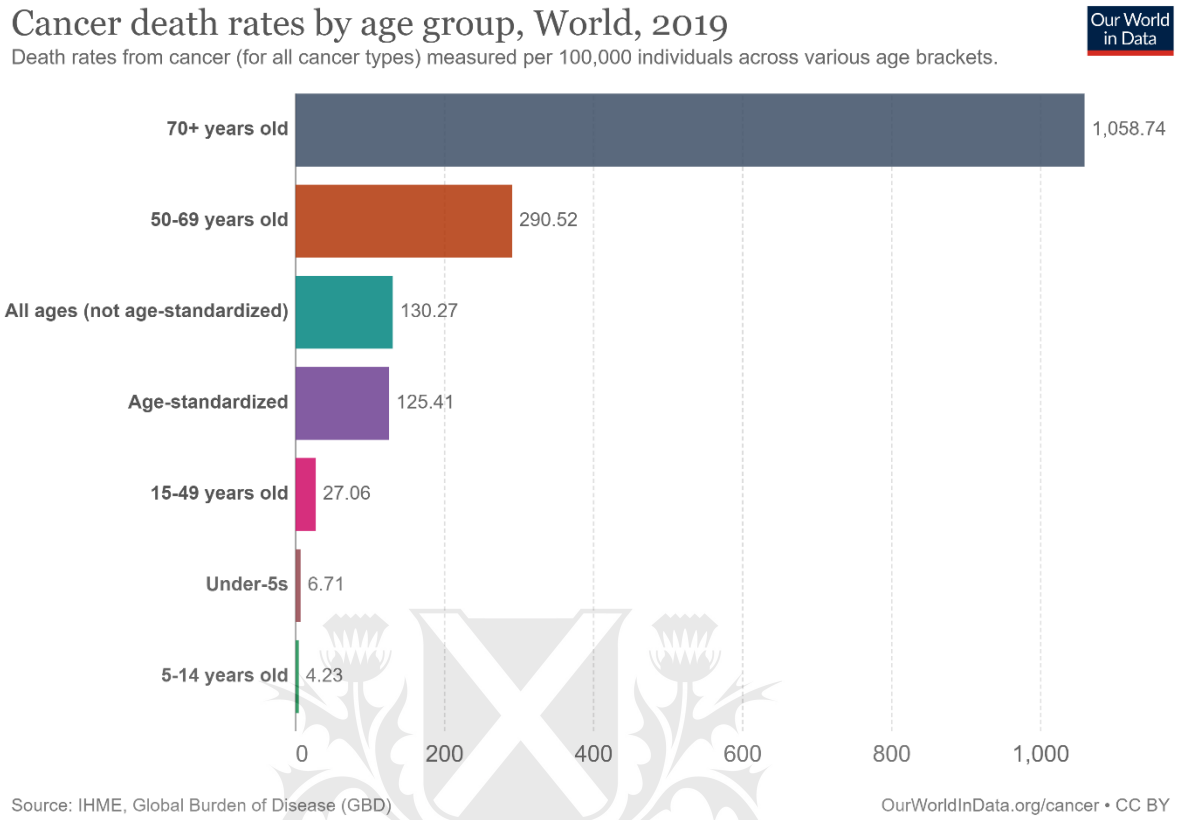
Almost half – 46% in 2017 – of all people who die from cancer are 70 or older. Another 41 percent are between 50 and 69 years old – so that 87% of all cancer victims are older than 50 years.

The distribution of deaths across the age spectrum has changed notably since 1990. The share of deaths which occur in those aged over 70 has increased by 8 percentage points, whilst the share in those aged 50-69 and 15-49 has fallen.

Collectively, children and adolescents under 14 years old account for around one percent of cancer deaths — this equates to around 110,000 children per year.



Next chart shows the death rate – the number of cancer deaths of people in a certain age group per 100,000 people of the same age group. Of the people who are 70 years and older more than 1% (i.e., more than 1000 per 100,000) die from cancer each year.



The global disease burden from cancer:

Death rates only capture the mortality of cancer. However, the impact of cancer on people's lives is more than that. Many live with cancer for long periods and it is important to also capture the morbidity caused by cancer.

The Disability-Adjusted Life Year (DALY) is a metric that captures the total burden of disease – both from years of life lost due to premature death and from years lived with the disease. One DALY equals one lost year of healthy life.

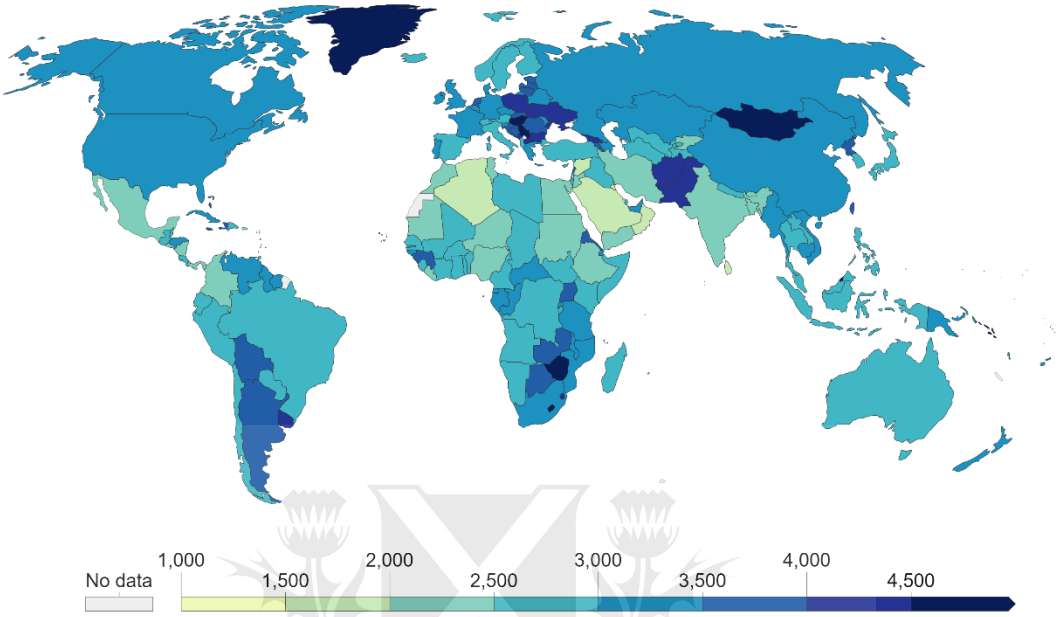
The map shows DALYs from cancers, measured per 100,000 individuals. This is age-standardized to allow comparisons between countries and over time. This is measured across all cancer types.

Also shown are disease burden rates broken down by cancer types. We see that at a global level, the largest burden results from tracheal, bronchus and lung cancer, followed by liver, stomach, colon & rectum, and breast cancer.

Disease burden rates from cancers, 2019

Our World in Data

Disability-Adjusted Life Years (DALYs) per 100,000 individuals from all cancer types.
DALYs measure the total burden of disease – both from years of life lost due to premature death and years lived with a disability. One DALY equals one lost year of healthy life.

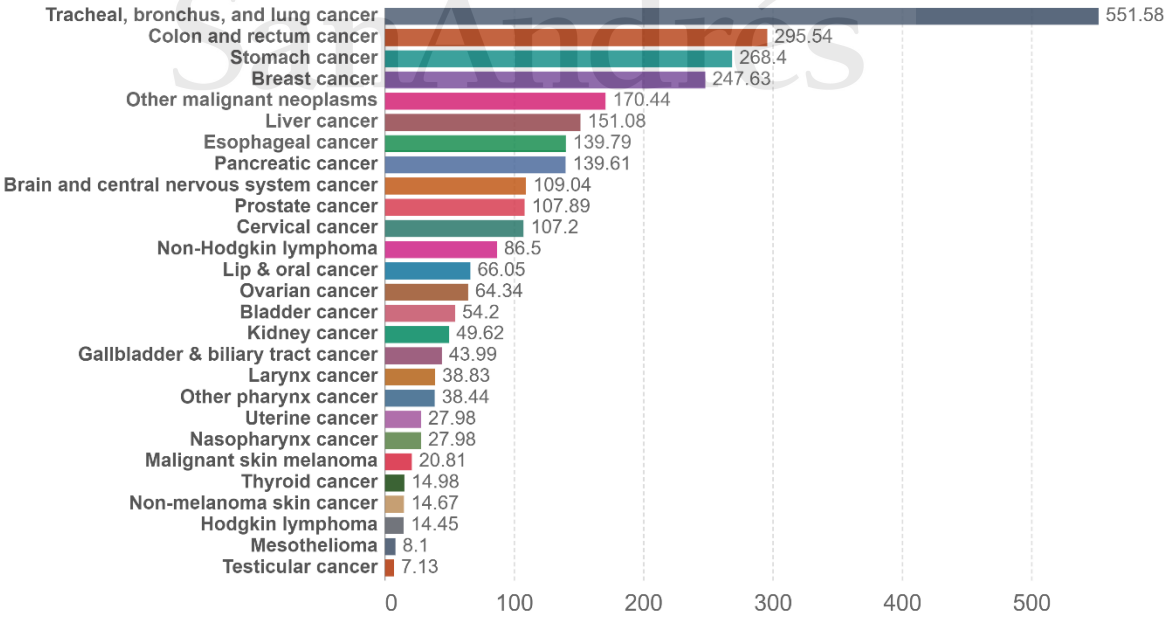


Source: IHME, Global Burden of Disease
Note: To allow comparisons between countries and over time this metric is age-standardized.

Disease burden rates by cancer types, World, 2019

Our World in Data

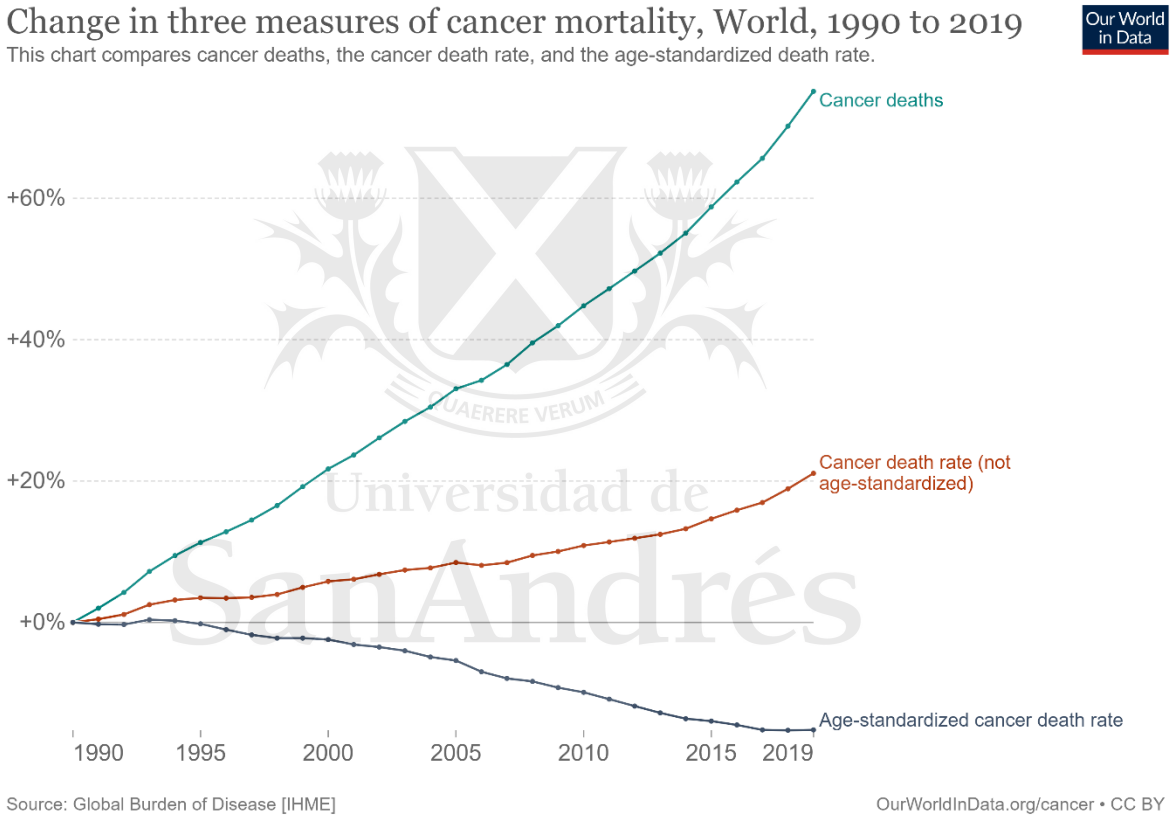
Disability-Adjusted Life Years (DALYs) per 100,000 individuals from all cancer types.
DALYs measure the total burden of disease – both from years of life lost due to premature death and years lived with a disability. One DALY equals one lost year of healthy life.



Source: IHME, Global Burden of Disease
Note: To allow comparisons between countries and over time this metric is age-standardized.

Is the world making progress against cancer? We saw with the previous graphs that cancer is one of the largest health problems in the world. How is cancer mortality changing over time?

Three different measures allow us to understand how the mortality of cancer has changed: the number of deaths, the death rate, and the age-standardized death rate. A comparison of how these three measures have changed is shown in the visualization:



Let's look at what we can learn from each of these:

The number of cancer deaths increased by 66%

More people than ever before die from cancer – 9.6 million in the latest data from 2017.

In 1990, 5.7 million people died from cancer. This means we have seen a 66% increase in the global number of cancer deaths. This increase is what the green line in the visualization shows.

The death rate from cancer increased by 17%

But in a world with more people, we would expect more people to die. As the world population is growing the total number of deaths is rising – since 1990, the number of deaths increased from 46 million to 56 million per year.

This of course means that the number of people who did not die of cancer has also increased. To assess whether we are making progress against cancer we cannot rely on the absolute number of deaths alone. It does not account for the increase of the world population.

This is why health statisticians study the number of deaths relative to the size of the population – the death rate. It is measured as the number of cancer deaths per 100,000 people.

The red line in the chart shows that the death rate from cancer has increased by 17% since 1990. This tells us that if the world population had not increased, then instead of the number of cancer deaths increasing by 66% (as we saw above), they would only have increased by 17%. Only a quarter as much.

The difference between the steep rise in the number of deaths and the slower rise of the death rate is due to the increase of the global population.

The age-standardized death rate from cancer declined by 15%

Cancer kills mostly older people – as the death rate by age shows, of those who are 70 years and older, 1% die from cancer every year. For people who are younger than 50, the cancer death rate is more than 40-times lower (more detail here).

We would therefore expect that many more people die from cancer in an old population than in a young population. Because health is improving and fertility rates are falling, the world is aging rapidly. This impacts the change over time that we are interested in: historically, fewer died from cancer because larger

parts of the population died before they reached the age when cancer becomes a common cause of death.

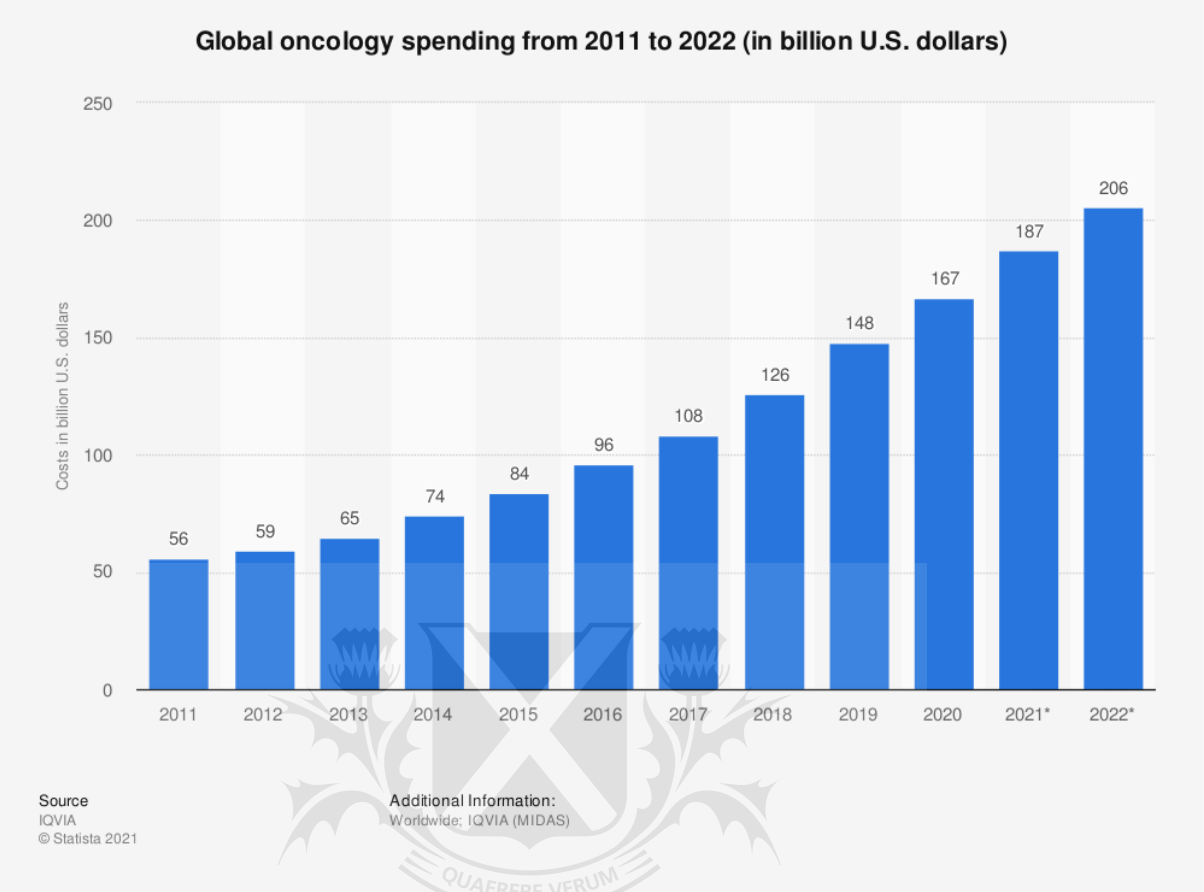
Epidemiologists correct for changes in age-profile over time by relying on the so called ‘age-standardized death rate’. This metric tells us what the death rate would be if the age structure of the population had stayed the same over time and would be the same across countries.

Once we correct for both population changes and aging, we get the blue line in the visualization: the age-standardized death rate from cancer. Globally this has fallen by 15% since 1990. (Ritchie, 2019) (1)

Economic aspects of cancer

With a clear vision of the impact of cancer in the world’s health, a second disease related effect is the economic, the cost that preventions campaigns, diagnostics and treatments has for the public health and private health systems of the world. This section provides a number of statistics that has the aim of quantify the money invested under all aspects related to cancer.

The first chart shows total oncology spending worldwide, including spending for supportive care, from 2010 to 2022. In 2020, global oncology spending totaled 167 billion U.S. dollars. In comparison, costs stood at 74 billion dollars six years earlier.



(Stefan Harrer, 2019) (2)

The cost of comprehensive cancer management has not only been rising steadily as seen in the previous graph, but it is projected that it will continue to increase in the coming years, the following statistic shows a projection of the global spending and growth in the oncology market between 2021 and 2025. In 2021, oncology spending is expected to be around 187 billion U.S. dollars worldwide. Spending in this market is expected to increase to about 273 billion dollars until 2025. The maximum projection for the 2021-2025 CAGR is 12 percent.

Spending and growth worldwide in oncology market 2021-2025

Projected spending and growth in the global oncology market between 2021 and 2025

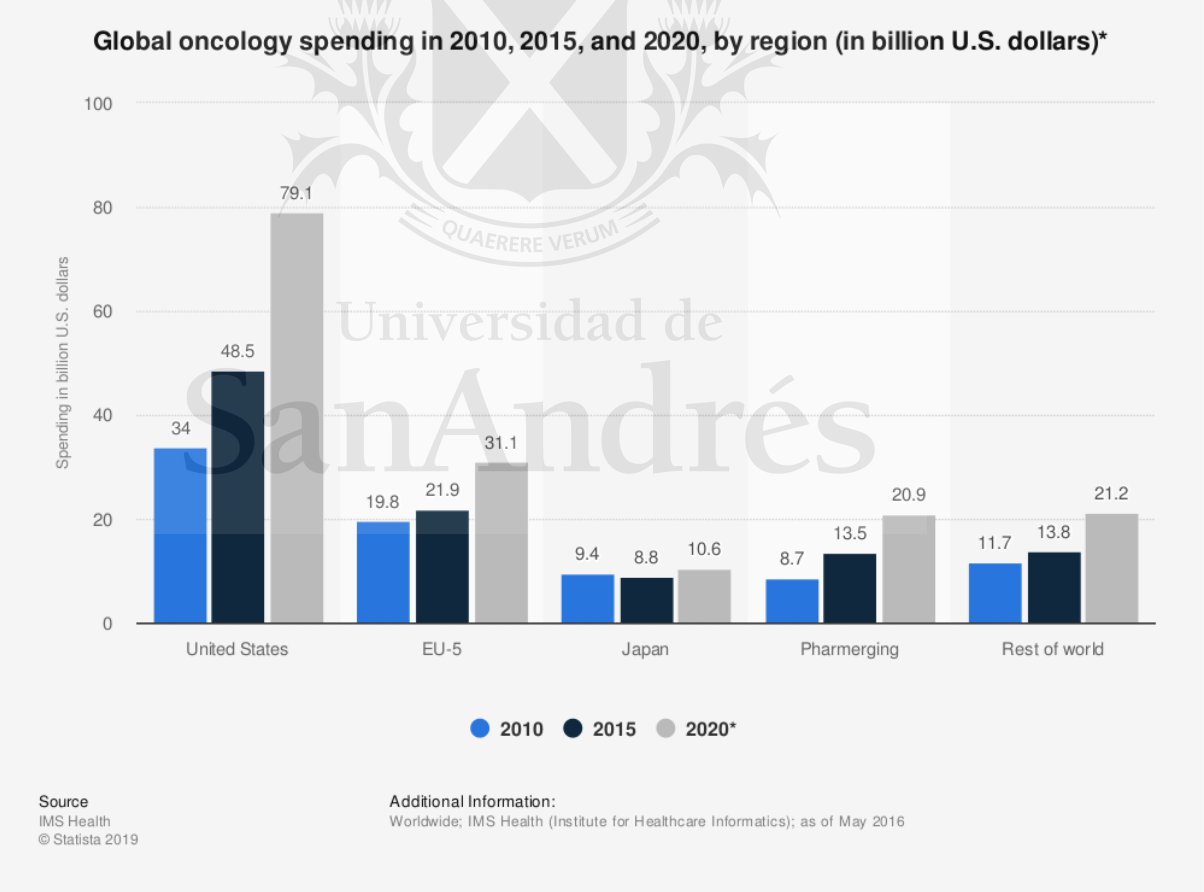
2021 (in billion USD)

187

2025 (in billion USD)	273
CAGR 2021 to 2025 (%)*	12

(Stefan Harrer, 2019) (3)

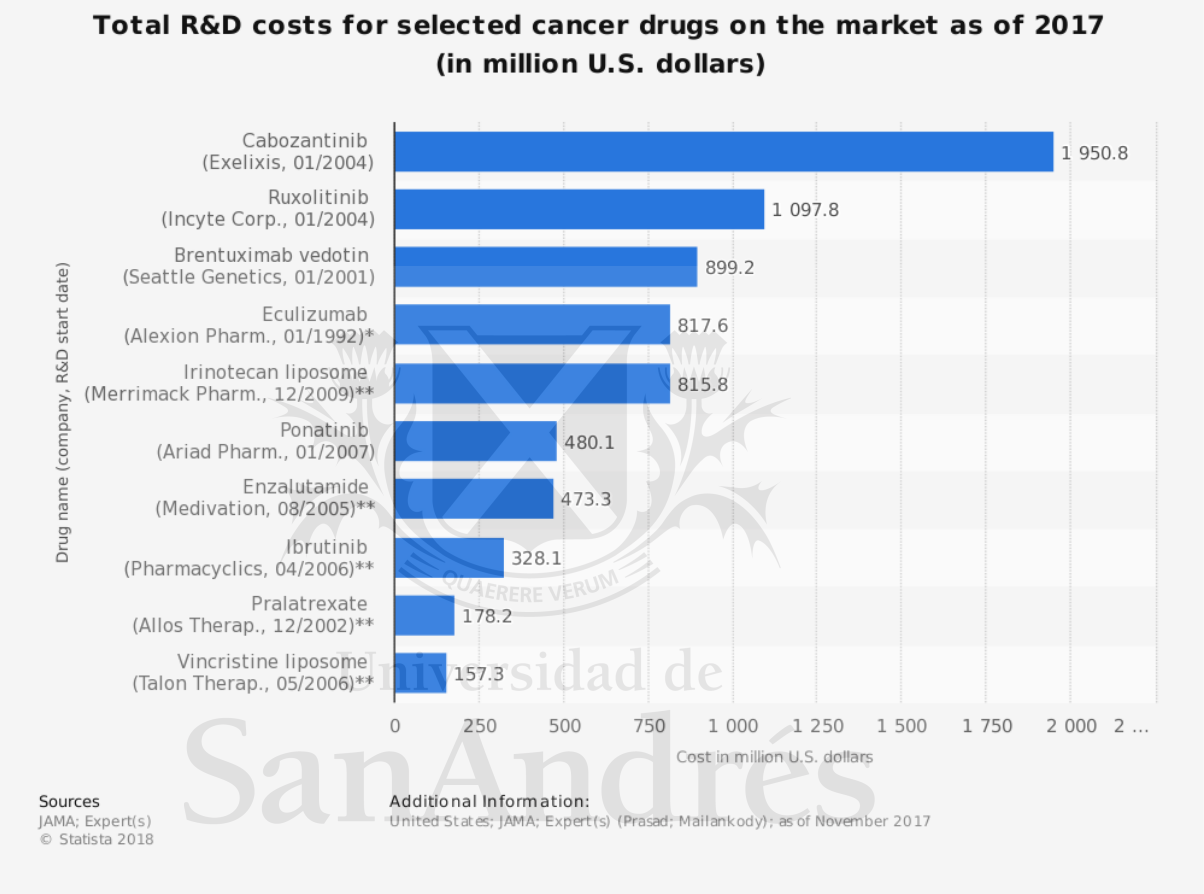
Segmenting the expenditure mentioned by economic zones, it is seen that the developed countries lead this trend, this statistic shows oncology spending in the global market by region, in 2010, 2015, and gives a forecast for 2020. In 2015, the so-called pharmerging countries spent around 13.5 billion U.S. dollars on oncology, including supportive care.



(Statista, 2016) (4)

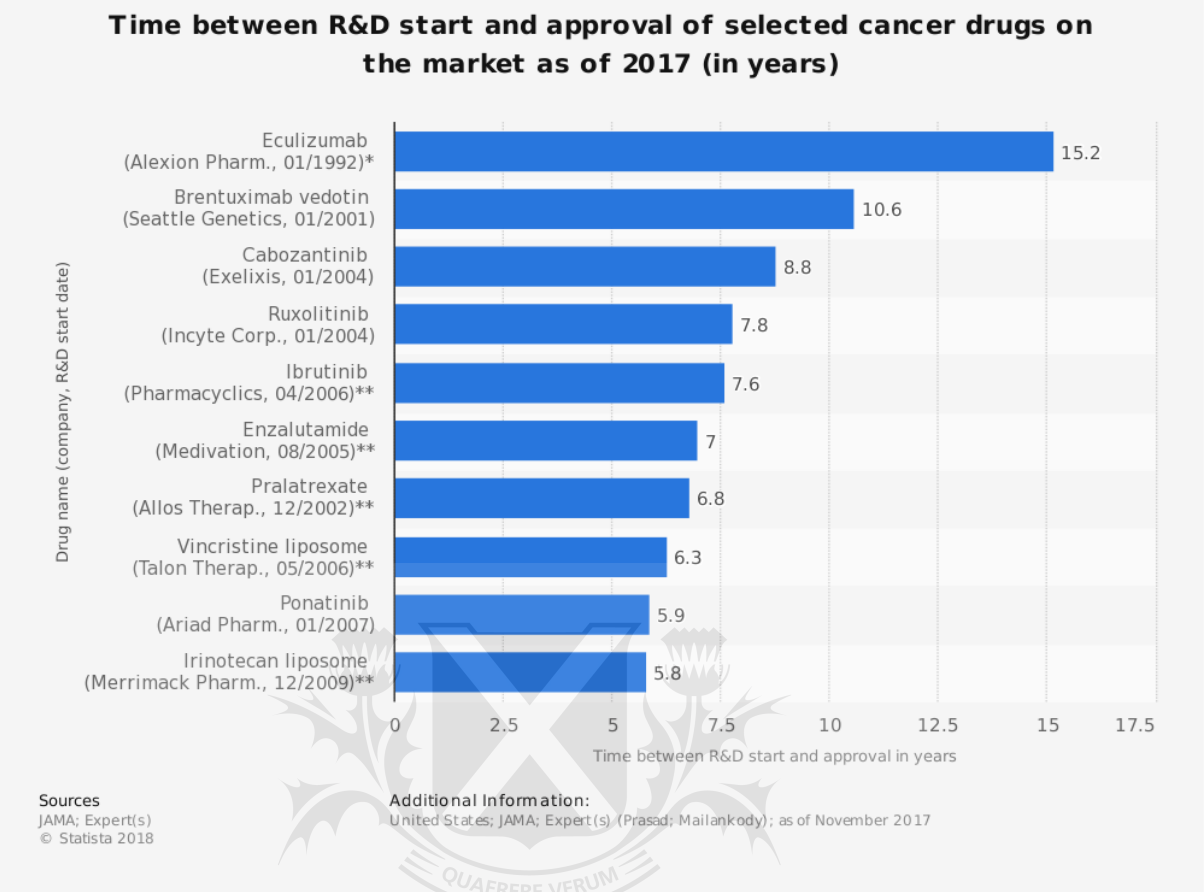
The development of medicines is a long and costly process that is generally concentrated in the most important economic areas of the world, presented in the previous graph, where the main pharmaceutical laboratories operate with

large areas of R&D, in the case of drugs for treatment of cancer this is much more marked, next statistic shows the total cost of research and development (R&D) for select cancer drugs that were on the market as of 2017. Alexion Pharmaceuticals' drug Eculizumab had a total R&D cost of 817.6 million U.S. dollars since R&D start in January 1992.



(Mikulic, Statista, 2017) (5)

And in terms of time elapsed, this statistic shows the time between the R&D start and the approval of selected cancer drugs which were on the market as of 2017. Alexion Pharmaceuticals' drug Eculizumab had a research and development phase of over 15 years until its approval in March 2007.



(Mikulic, Statista, 2017) (6)

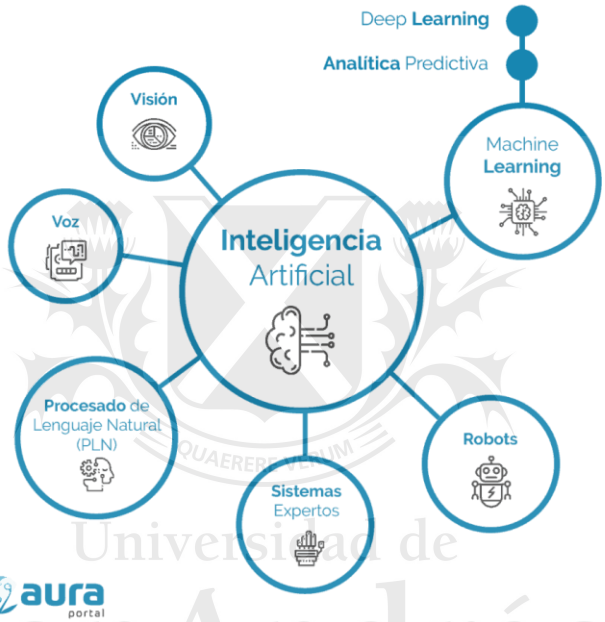
In this section I've presented a series of studies and statistics in order to quantify the impact of cancer in the world, both in terms of the impact on the loss of human lives as well as the economic impact of dealing with this disease in all its aspects. As has been seen, the tendency is for the problem to worsen in the coming years, maintaining the traditional medical and pharmaceutical approach.

Artificial intelligence – Technologies and current maturity level

In the last five years the technologies included under the title of "artificial intelligence" (AI) have reached a point of maturity in which they are already widely used in current applications, in all kinds of areas such as voice assistant, image processing in real time, deep learning, analysis of large amounts of

unstructured information in search of patterns, automation of repetitive tasks in companies.

When we talk about the technologies encompassed under the term artificial intelligence, we refer to technologies such as machine learning, natural language recognition and processing, visual recognition, text recognition, cognitive intelligence, among others.



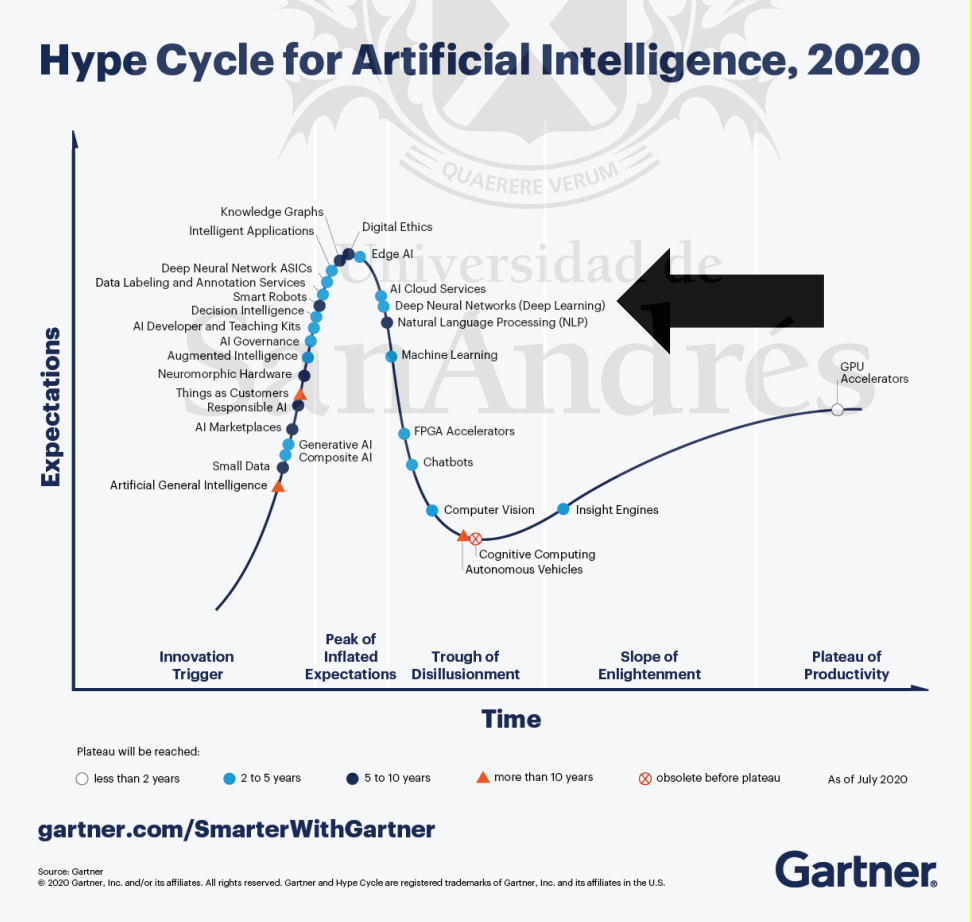
The current state of maturity has led to the large technology providers in the market such as IBM, Google, Microsoft or Amazon, as well as niche providers born to develop and commercialize this type of solutions, are today offering AI services as a platform in the cloud, to which companies can subscribe under different contracting modalities.

The deep learning technologies, among others, appears as the more matures now with potential of massive adoption by the market.

The services available on the platforms (cloud) of the main providers are presented to the client in the form of APIs (Application Program Interface) that encompass artificial intelligence functionalities, allowing the client to use them without having a detailed knowledge of the technology in question, in general, under models of payment by number of calls to said APIs.

This combination of maturity in technology and an accessible business model of the same, as software as a service (SAAS), has produced a rapid adoption in various business areas, scientific and educational communities, and for five years to date there have been developed artificial intelligence applied models that seek to provide superior capabilities to the field of medical research, both in the search for definitive cures for diseases such as cancer and various neurological disorders, as well as in the search for more precise and earlier diagnoses that produce greater chances of early treatment, aiming to reduce mortality in patients and more efficient use of the treatment capacities of health centers.

The next chart shows the maturity state of the vast number of technologies under the AI umbrella term:



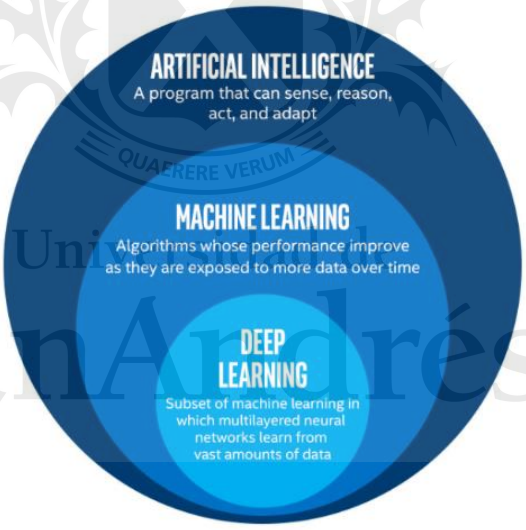
(Goasduff, 2020) (7)

With this currently available technological framework, this work presents the advances made in the early diagnosis of certain types of cancer as a consequence of the use of artificial intelligence to carry out an early diagnosis based on the massive processing of data and diagnostic images.

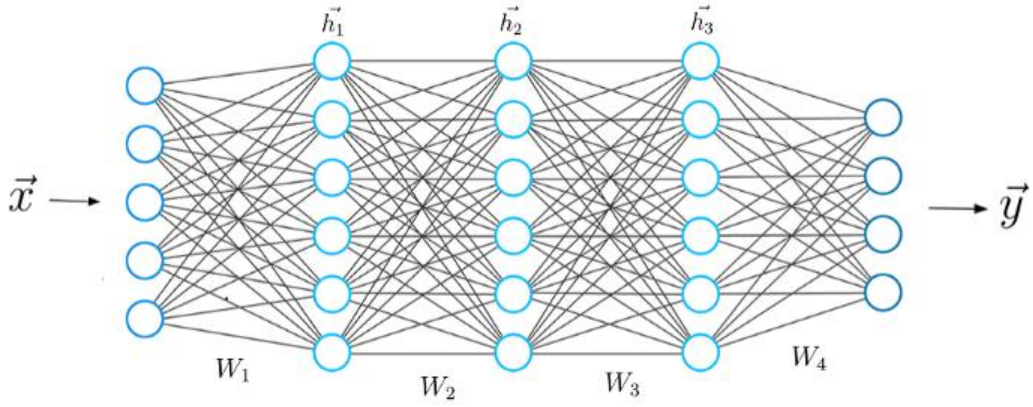
Deep Learning, the key technology in the development of an image's predictive analysis platform. What is?

Deep Learning is a subset of Machine Learning, which on the other hand is a subset of Artificial Intelligence. Artificial Intelligence is a general term that refers to techniques that enable computers to mimic human behavior.

Machine Learning represents a set of algorithms trained on data that make all of this possible.



Deep Learning, on the other hand, is just a type of Machine Learning, inspired by the structure of a human brain. Deep learning algorithms attempt to draw similar conclusions as humans would by continually analyzing data with a given logical structure. To achieve this, deep learning uses a multi-layered structure of algorithms called neural networks.



The design of the neural network is based on the structure of the human brain. Just as we use our brains to identify patterns and classify different types of information, neural networks can be taught to perform the same tasks on data.

The individual layers of neural networks can also be thought of as a sort of filter that works from gross to subtle, increasing the likelihood of detecting and outputting a correct result.

The human brain works similarly. Whenever we receive new information, the brain tries to compare it with known objects. The same concept is also used by deep neural networks.

Neural networks enable us to perform many tasks, such as clustering, classification or regression. With neural networks, we can group or sort unlabeled data according to similarities among the samples in this data. Or in the case of classification, we can train the network on a labeled dataset in order to classify the samples in this dataset into different categories.

In general, neural networks can perform the same tasks as classical algorithms of machine learning. However, it is not the other way around.

Artificial neural networks have unique capabilities that enable deep learning models to solve tasks that machine learning models can never solve.

All recent advances in artificial intelligence in recent years are due to deep learning. Without deep learning, we would not have self-driving cars, chatbots or personal assistants like Alexa and Siri. The Google Translate app would continue to be as primitive as 10 years ago (before Google switched to neural networks for this App), and Netflix or Youtube would have no idea which movies

or TV series we like or dislike. Behind all these technologies are neural networks.

We can even go so far as to say that today a new industrial revolution is taking place, driven by artificial neural networks and deep learning.

At the end of the day, deep learning is the best and most obvious approach to real machine intelligence we've had so far.

Why is deep learning and artificial neural networks so powerful and unique in today's industry? And above all, why are deep learning models more powerful than machine learning models? Let me explain it to you.

The first advantage of deep learning over machine learning is the needlessness of the so-called feature extraction.

Long before deep learning was used, traditional machine learning methods were mainly used. Such as Decision Trees, SVM, Naïve Bayes Classifier and Logistic Regression.

These algorithms are also called flat algorithms. Flat here means that these algorithms cannot normally be applied directly to the raw data (such as .csv, images, text, etc.). We need a preprocessing step called Feature Extraction.

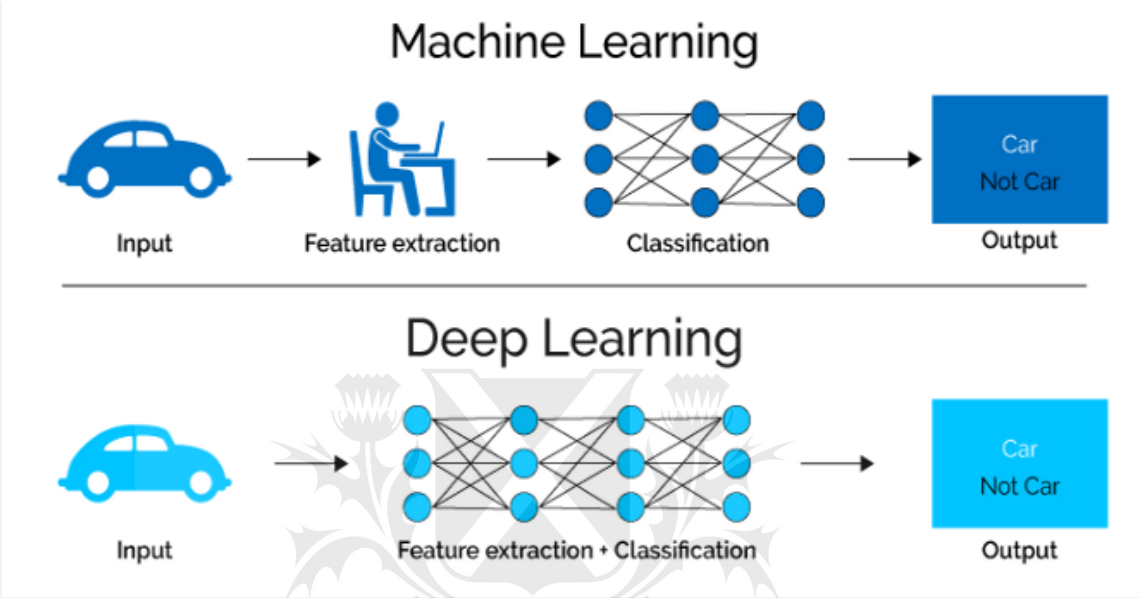
The result of Feature Extraction is a representation of the given raw data that can now be used by these classic machine learning algorithms to perform a task. For example, the classification of the data into several categories or classes.

Feature Extraction is usually quite complex and requires detailed knowledge of the problem domain. This preprocessing layer must be adapted, tested and refined over several iterations for optimal results.

On the other side are the artificial neural networks of Deep Learning. These do not need the Feature Extraction step.

The layers are able to learn an implicit representation of the raw data directly and on their own. Here, a more and more abstract and compressed representation of the raw data is produced over several layers of an artificial neural-nets. This compressed representation of the input data is then used to

produce the result. The result can be, for example, the classification of the input data into different classes.



Feature Extraction is only required for ML Algorithms.

In other words, we can also say that the feature extraction step is already part of the process that takes place in an artificial neural network.

During the training process, this step is also optimized by the neural network to obtain the best possible abstract representation of the input data. **This means that the models of deep learning thus require little to no manual effort to perform and optimize the feature extraction process.**

Let us look at a concrete example, if you want to use a machine learning model to determine if a particular image is showing a car or not, we humans first need to identify the unique feature or features of a car (shape, size, windows, wheels, etc.) extract the feature and give them to the algorithm as input data.
In this way, the algorithm would perform a classification of the images. That is, in machine learning, a programmer must intervene directly in the action for the model to come to a conclusion.

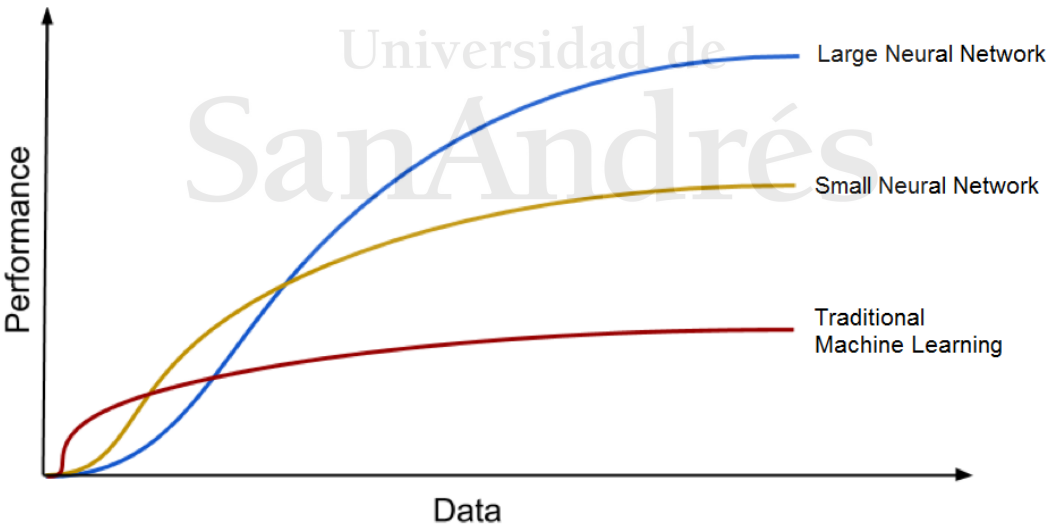
In the case of a deep learning model, the feature extraction step is completely unnecessary. The model would recognize these unique characteristics of a car and make correct predictions.

That completely without the help of a human.

In fact, refraining from extracting the characteristics of data applies to every other task you'll ever do with neural networks. Just give the raw data to the neural network, the rest is done by the model.

The second huge advantage of Deep Learning and a key part in understanding why it's becoming so popular is that it's powered by massive amounts of data. The "Big Data Era" of technology will provide huge amounts of opportunities for new innovations in deep learning. As per Andrew Ng, the chief scientist of China's major search engine Baidu and one of the leaders of the Google Brain Project,

"*The analogy to deep learning is that the rocket engine is the deep learning models and the fuel is the huge amounts of data we can feed to these algorithms.*"



Deep Learning Algorithms get better with the increasing amount of data.

Deep Learning models tend to increase their accuracy with the increasing amount of training data, where's traditional machine learning models such as SVM and Naive Bayes classifier stop improving after a saturation point.

(Oppermann, 2019) (8)

Artificial Intelligence (Deep Learning) applied to early cancer diagnosis

Based on the characteristics and benefits of AI in general and deep learning in particular explained in previous sections, I'll start to link those concepts with cancer research and investigation in order to present the advances that the AI is achieving in different areas of cancer detection centers.

The inherent complexity of human malignancies calls for the development of cutting-edge technologies, concepts and methods that will eventually be applied in the diagnosis and treatment of cancer patients.

AI excels at recognizing patterns in large volumes of data, extracting relationships between complex features in the data, and identifying characteristics in data (including images) that cannot be perceived by the human brain. It has already produced results in radiology, where clinicians use computers to process images rapidly, thus allowing radiologists to focus their time on aspects for which their technical judgment is critical. **For example, last year, the Food and Drug Administration approved the first AI-based software to process images rapidly and assist radiologists in detecting breast cancer in screening mammograms.**

Integration of AI technology in cancer care could improve the accuracy and speed of diagnosis, aid clinical decision-making, and lead to better health outcomes. AI-guided clinical care has the potential to play an important role in reducing health disparities, particularly in low-resource settings. Some examples:

- **National Cancer Institute (NCI) will invest in supporting research, developing infrastructure, and training the workforce to help achieve these goals and more.**

- The journal of AIO (www.aioncology.org) is established to provide the unmet academic requirements in publication, evaluation, and quality control of AI's applications in oncology. To ensure high-quality research paper publications from global talents with or without fundings, **the journal is FREE of charge for submission, peer-review and publication online**.

- The effective utilization of cancer big data entails all the steps from data processing and storage to data mining, analysis, and final applications, such as the identification of patient-specific oncogenic processes and biomarkers. Moreover, the continuous improvement of data quality through standardization procedures that ensure responsible molecular and clinical data sharing, interoperability, and security is a key aspect for cancer research that is strongly catalyzed by initiatives such as the Global Alliance for Genomics and Health

Summarizing, the goal is to use AI's ability to recognize patterns that are too subtle for the human eye to detect to guide physicians towards better-targeted therapies and to improve outcomes for patients. Some scientists are even applying AI to screening tests in the hope of identifying people with an increased cancer risk or catching the disease sooner.
(L, 2019) (9).

Finishing this chapter, I would like to present the main concepts about the cancer situation in the world and the AI-Deep Learning state of the art that that would guide this work in the next chapters:

- The cancer burden can be reduced through early detection.
- Many cancers have a high chance of cure if diagnosed early.
- There are two components of early detection: early diagnosis and screening.
- Deep Learning is powered by massive amounts of data. Deep Learning models tend to increase their accuracy with the increasing amount of training data.

Chapter IV: Real cases of AI use in images processing for premature cancer detection

Going to real cases that are using AI to process images and detect signs of cancer in an earlier stage, here are some examples of this technology applied to the real world:

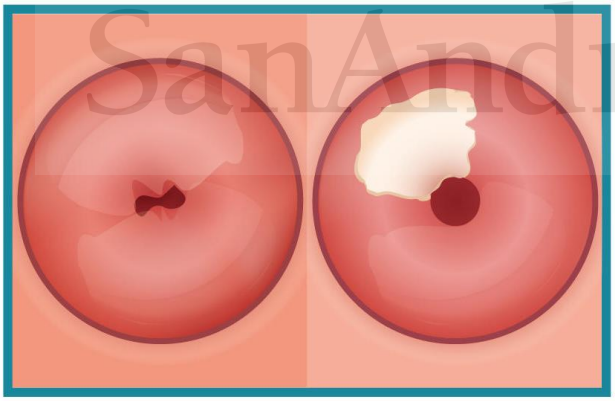
1) Leveraging AI to improve detection of cervical precancer

Scientists on NCI's intramural research program helped to develop an AI approach using deep learning for the automated detection of precancerous lesions using cervical images. The goal was to develop a more-accurate and cost-effective screening method that could be used easily in low and middle resource settings. They testes the approach on more than 60.000 cervical images from an NCI study.

Results:

Cervical Cancer Screening Methods

The project compared multiple cervical screening methods, used at the beginning of the study in women who were followed for 7 years, including:

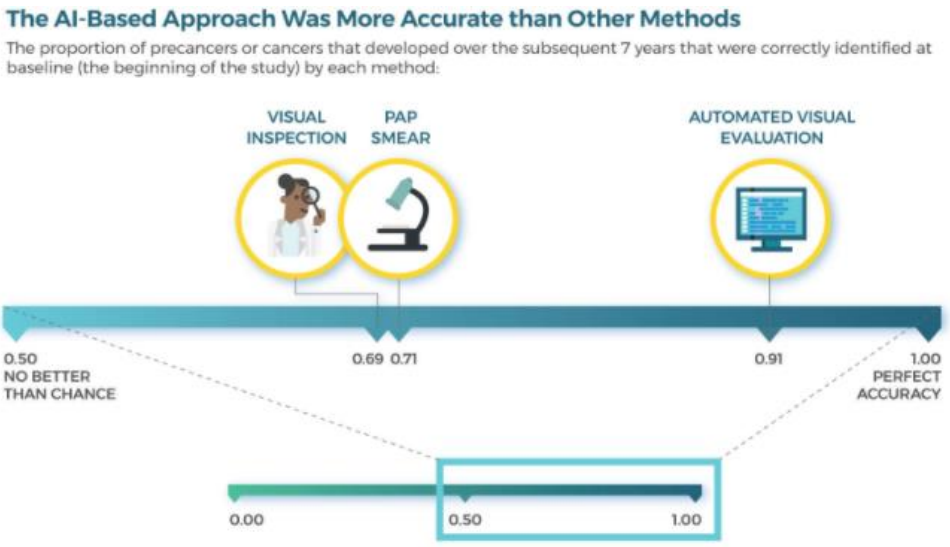


HEALTHY CERVIX (L) AND CERVIX WITH TISSUE CHANGES (IN WHITE) CAUSED BY HPV INFECTION (R)

Visual Appearance: Photographs were taken after each study participant's cervix had been rinsed with vinegar. Vinegar highlights changes to normal tissue caused by HPV infection, including precancer or cancer, by turning the tissue white. A gynecologist evaluated the pictures to identify precancerous or cancerous lesions. The sensitivity (identification of true positives) of this approach was 69%.

Pap Smear: Cervical cells were collected, affixed to a slide, and analyzed by a pathologist for the presence of precancerous or cancerous cells. The sensitivity of this approach was 71%.

Automated Visual Evaluation: A deep-learning, artificial intelligence approach was used to evaluate digitized images of the cervix in an automated process that predicted the probability that the image represented a case of precancer or cancer. The sensitivity of this approach was 91%.



(L, 2019) [(10)]

2) Brain and prostate cancer correct diagnosis

When a Dr. saw the results of some test of a girl who had brain cancer few years ago, she saw that the medulloblastoma for which she had been treated a few years earlier had returned. The girl's recurrent cancer was found in the same part of brain as before, and the biopsy seemed to confirm medulloblastoma.

With this diagnosis, the girl would begin a specific course of radiotherapy and chemotherapy. But just as neuropathologist Matija Snuderl was about to sign off on the diagnosis and set her on that treatment path, he hesitated. The biopsy was slightly unusual, he thought, and he remembered a previous case in which what was thought to be medulloblastoma turned out to be something else. So, to help him make up his mind, Snuderl turned to a computer. He arranged for the girl to have a full-genome methylation analysis, which checks for small hydrocarbon molecules attached to DNA. The addition of such methyl groups is one of the mechanisms behind epigenetics — when the activity of genes is altered without any mutation to the underlying genetic code — and different types of cancer show different patterns of methylation. Snuderl fed the results to an artificial-intelligence (AI) system developed

by a consortium including researchers at the German Cancer Research Center in Heidelberg, and let the computer classify the tumor.

“The tumour came back as a glioblastoma, which is a completely different type,” Snuderl says. The new tumour seemed to be the result of radiation used to destroy the first cancer, and called for a different drug and radiation treatment plan. Treatment for the wrong cancer could have ill effects without actually destroying the cancer. “If I had finalized the case just on pathology, I would have been terribly wrong,” Snuderl says.

The system Snuderl used is an early example of AI as a tool to diagnose cancer. NYU Langone’s Perlmutter Cancer Center received state approval to use its AI classifier as a diagnostic test in October 2019, and researchers around the world are developing similar systems to help pathologists diagnose cancer more accurately. The goal is to use AI’s ability to recognize patterns that are too subtle for the human eye to detect to guide physicians towards better-targeted therapies and to improve outcomes for patients. Some scientists are even applying AI to screening tests in the hope of identifying people with an increased cancer risk or catching the disease sooner.

The methylation method

The methylation-based classifier, developed by a consortium of dozens of researchers, was originally trained to sort medulloblastomas into subtypes. The German-led team eventually expanded the effort to cover all of the 100 or so known cancers of the central nervous system. When the initial results were published in March 2018, the researchers made the classifier available online. Other researchers can upload methylation profiles and, in a few minutes, learn which subtype the cancer fits into. They also receive a confidence score that says how likely the result is to be correct. About 1,000 such profiles are uploaded each month, says

Andreas von Deimling, a neuropathologist at the German Cancer Research Centre who was one of the project's leaders.

Although Langone's use of the test has been approved by New York state, the website notes that the classifier is still a research tool that has not been clinically validated. The classifier was originally trained using around 2,800 tumour samples, but since the website has been operating, that number has grown to around 60,000. "This is much more than a single pathologist sees in an entire lifetime," von Deimling says. "By the sheer number of tumours we can now examine with this system, we find novel entities no pathologist has ever been able to define previously." The system compares data to its reference list of tumours and places the profile into a group, but if it doesn't quite match, the cancer gets a low confidence score. Pathologists examine the low-scoring samples, and if there are at least seven with the same methylation profile, they assign them to a new group and retrain the classifier. The classifier now recognizes about 150 different cancer entities.

The computer's ability to spot those cancer types could cut hospitals' error rates. In the initial study, the algorithm found that 12% of brain tumours had been misdiagnosed by pathologists. Snuderl says that NYU has similar error rates of 12–14% among its patients. "That's not an insignificant number of people that could benefit simply from having the right diagnosis," Snuderl says.

Methylation profiling is expensive — typically, only large cancer research centers can afford it. So, the scientists hope to find simpler biomarkers to identify the subtypes. If, for instance, they can discover differences that are visible by looking at stained tissue under a microscope, they can make the same level of diagnostic sorting available to the many hospitals that don't have the resources for methylation profiling. "You can develop

these markers only if you have the grouping correct in the first place," von Deimling says.

Prostate cancer right diagnosis.

Correctly diagnosing cancers in other parts of the body can also be difficult. Working out if a person has prostate cancer and whether that cancer is aggressive enough to need treatment or merely needs to be watched can be tricky.

Most prostate cancers are diagnosed by taking biopsies from a standard set of locations on the prostate, but this can mean the actual cancer is missed. A newer approach uses multiparametric magnetic resonance imaging (MRI), in which different types of MRI scan are combined. But highly trained radiologists don't always agree on what they're seeing in the images, and those with less experience do even less well at identification. "To reach a certain level of expertise in radiology, particularly in this prostate-cancer MRI diagnosis, requires a lot of training," says Kyung Hyun Sung, a radiologist at the University of Los Angeles, California. As a major prostate-cancer treatment center, the university has a program to train radiologists to read such images and boasts specialists with ten or more years of experience. But that is not the norm. "Community hospitals don't have that training period or expertise in their ranks," says Sung.

With those hospitals in mind, Sung is building an AI-based system called FocalNet to help physicians to better classify prostate cancer. To train the program, Sung and his colleagues collected around 400 pre-operative MRI scans of people who were going to have surgery to remove their prostate. The researchers fed FocalNet a subset of the scans, along with the tumour's Gleason score — a rating of malignancy, defined by pathologists who analyzed the tissue after the prostate was

removed. The system then looked for and learnt to spot patterns in the MRI scans that matched the pathology-based score.

The researchers then challenged FocalNet to provide a Gleason score for a new set of scans. The computer found 79.2% of the clinically significant cancer lesions, as determined by pathology. A group of radiologists, each with at least 10 years of experience of reading more than 1,000 images annually, managed 80.7% — a difference deemed statistically insignificant.

Currently, the value of a Gleason score derived from an MRI is limited because it is dependent on the skill of the radiologist interpreting the image. But that, says Sung, is when machine learning can help. “The machine will be consistent. It’s not going to have inter-reader variability.” With the help of a system like FocalNet, multiparametric MRI could be used even without experienced radiologists, leading to clearer diagnoses that can guide people to the right treatment.

(Charity, 2020) (11)

3) Breast and Lung cancer preventive diagnosis

Although getting the diagnosis right is important, catching cancer early can also lead to higher survival rates. Many women in the United States have annual mammograms starting in their forties or fifties. That produces a lot of imaging data. Regina Barzilay, a computer scientist at the Massachusetts Institute of Technology (MIT) in Cambridge, wanted to see if a machine could use those data to draw a more accurate picture of a person’s risk of developing breast cancer.

Barzilay collected almost 89,000 mammograms from nearly 40,000 women who had been screened over a 4-year period, and checked the images against a national tumour registry to determine which women

were eventually diagnosed with breast cancer. She then trained a machine-learning algorithm with a subset of those images and outcomes, before testing the system to see how well it predicted cancer risk. The computer put 31% of the women who eventually developed breast cancer into the highest risk group. But the standard Tyrer–Cuzick model that physicians use to estimate risk — based on factors such as age, family history of cancer, and age at first menstrual period and at menopause — placed only 18% in that group, even when physicians added measurements of breast density from mammograms to the model.

The researchers are continuing to improve the model, says Adam Yala, a PhD student at MIT who works with Barzilay on the project. The researchers hope that their work can lead to more personalized breast-cancer screening. Specialists currently disagree about how often women should get mammograms — too frequently and it drives up health-care costs with no benefit, not often enough and some early cancers might be missed. If the MIT system can learn to differentiate between people who will develop cancer within five years and those who won't, Yala says, it might allow physicians to personalize screening schedules and offer frequent mammograms only to those whose early scans show they are at high risk.

Researchers at Google are also trying to improve cancer screening. Medical groups in the United States and Canada already recommend screening certain people who are at high risk of lung cancer using computed tomography (CT) scans based on low-dose X-rays, and the same screening protocol is under consideration in the European Union. Computer scientists at Google wanted to see whether they could predict which people would go on to develop lung cancer by using AI to analyze low-dose CT scans of the lungs.

They collected about 43,000 scans from almost 15,000 people that had been amassed by the National Lung Screening Trial (NLST), a study run by the US National Cancer Institute. Of those, 638 people did not have cancer at the time of the initial scan, but were diagnosed within one year, with the cancer confirmed by biopsy. “Our main goal was to try and predict whether someone ends up with lung cancer a year from when they got screened, or two years in some cases,” says Shravya Shetty, a software engineer at Google in San Francisco, California.

In people with only one scan available, the AI outperformed all of the six radiologists who also examined the CT scans to assess risk of lung cancer. The AI reduced the number of false positives by 11% and false negative by 5%. When there were two scans, the radiologists did about as well as the computer. Researchers hope that more accurate screening will lead to more effective treatment. “Ultimately what we want is patients to get their cancers caught earlier,” says Daniel Tse, a medical doctor at Google Health who led the project.

The Google model is still very new, Tse says, and AI systems under development have a way to go before reaching widespread clinical use. “It does show great promise,” he says, “but we’re going to be doing further studies to see how the models interact in larger scales of data, new environments, things like that.” The goal, he says, is to blend computer technology with the knowledge and skills of doctors, “and hopefully produce even better results than any one of the two could produce on its own”.

(Charity, 2020) (11)

4) Using AI to Benefit cancer patients during COVID-19 pandemic

Researchers led by The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London, (ICR) and funded by The Royal Marsden Cancer Charity, are investigating how they can utilize the latest technology in artificial intelligence (AI) to provide better care and treatments for cancer patients during the COVID-19 pandemic.

COVID-19 is a serious viral infection that can affect people of all ages but is likely to impact cancer patients differently who are at increased risk of developing the virus and more vulnerable due to their weakened immune systems. Lung cancer is one of the most common cancers in the UK and worldwide, with around 47,000 people diagnosed in the UK every year. This group of patients are particularly vulnerable to the virus as it is known to affect and severely impact the respiratory system.

Dr Richard Lee, Consultant Physician in Respiratory Medicine and Early Diagnosis, who is funded by The Royal Marsden Cancer Charity is leading a team at The Royal Marsden in collaboration with The ICR and Imperial College London on this research who have already been carrying out work in using artificial intelligence as a tool to help early diagnosis of lung cancer.

Researchers are aiming to use AI tools to analyze scans from cancer patients with the results then helping clinicians to balance and prioritize between targeting treatment for the infection and targeting cancer treatment. Some cancers are often treated by immunotherapy, which has been established as a key therapy for this condition. However, a key challenge which researchers have identified during the COVID-19 pandemic is that it can be difficult to distinguish symptoms of immunotherapy side effects from the virus and other infections such as infective pneumonia as these can all often present in a similar way. These side effects might include cough, breathlessness and changes of CT scan

imaging. An additional part of this study will be to diagnose subtle changes of lung cancer recurrence earlier so that it might be treated more effectively.

Dr Richard Lee said:

“A lot of the work we do is using AI approaches to CT scan images to identify subtle changes in patients which might help us to understand why one group of patients behave in a certain way. By using AI technology in this new research during the COVID-19 era, we are aiming to identify to what extent these changes are due to Coronavirus or if they are being caused by side effects from treatment. Being able to distinguish between infections and side effects will give us crucial information which is needed to help clinicians treat patients in the best possible way and improve patient outcomes.

“Unlike a traditional trial, we’ll be using CT scan images that already exist of cancer patients who have and haven’t tested positive for COVID-19. The advantage to having the data ready to use is that we should be able to fast track our research, enabling us to gather results quickly and effectively which will directly benefit cancer patients and improve their quality of life at this challenging time.”

The Royal Marsden and ICR have launched several critical research studies such as this at unprecedented speed, with The Royal Marsden Cancer Charity needing to raise over £500,000 over the coming weeks to ensure support for the research studies can continue.

Professor David Cunningham, Consultant Medical Oncologist at The Royal Marsden and Director of the NIHR Biomedical Research Centre at The Royal Marsden and The Institute of Cancer Research, London, said:

“We are uniquely placed to look at COVID-19 in a cancer setting, investigating the pandemic’s impact across a wide range of patients. These trials call upon our multidisciplinary expertise in areas such as

systemic therapies, radiotherapy, circulating tumor DNA which is detectable in blood tests, surgery and holistic care."

"Teams have been working at pace to establish studies that adhere to our usual rigorous protocol; each will have varying durations, with a focus on immediate impact through to longer term understanding of this novel virus. Importantly, with commercial, NHS and academic partners across the country, and thanks to fundraising from The Royal Marsden Cancer Charity and support from The NIHR Biomedical Research Centre we hope this research will have a national and international impact."

(Charity, 2020) (12)

5) Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis

Histopathology refers to the microscopic examination of tissue to study manifestations of disease. Specifically, in clinical medicine, histopathology refers to the examination of a biopsy or surgical specimen by a pathologist, after the specimen has been processed and histological sections have been placed on glass slides.

Pathologists face a substantial increase in workload and complexity of histopathologic cancer diagnosis due to the advent of personalized medicine. Therefore, diagnostic protocols have to focus equally on efficiency and accuracy. In this section we introduce 'deep learning' as a technique to improve the objectivity and efficiency of histopathologic slide analysis. Through two examples, prostate cancer identification in biopsy specimens and breast cancer metastasis detection in sentinel lymph nodes, we show the potential of this new methodology to reduce the workload for pathologists, while at the same time increasing objectivity of diagnoses

Microscopic analysis of hematoxylin and eosin (H&E) stained sections has been the basis for cancer diagnosis and grading for the past century1.

Protocols for the complete workup of biopsies or resected tissue specimens, including microscopic analysis, exist for many of the most common cancer types (e.g. lung, breast, prostate). Use of these protocols has led to strong prognostic and widely used grading strategies (e.g., the Gleason grading system).

Due to the rise in cancer incidence and patient-specific treatment options, diagnosis and grading of cancer has become increasingly complex. Pathologists nowadays have to go over a large number of slides, often including additional immunohistochemical stains, to come to a complete diagnosis. Moreover, there is an increase in the number of quantitative parameters pathologists have to extract for commonly used grading systems (e.g., lengths, surface areas, mitotic counts). Due to these difficulties, analysis protocols have been adapted and fine-tuned to offer the best balance between prognostic power and feasibility in daily clinical routine.

The recent introduction of whole-slide scanning systems offers an opportunity to quantify and improve histopathologic procedures. These systems digitize glass slides with stained tissue sections at high resolution. Digital whole slide images (WSI) allow the application of image analysis techniques to aid pathologists in the examination and quantification of slides. **One such technique which has gained prominence in the last five years in other fields is 'deep learning'. While 'deep learning' cannot be considered a single technique, it can roughly be described as the application of multi-layered artificial neural networks to a wide range of problems, from speech recognition to image analysis**. **In recent years, 'deep learning' techniques have quickly become the state of the art in computer vision. A specific neural network subtype (convolutional neural networks; CNN has become the de facto standard in image recognition and is approaching human performance in a number of**

tasks. These systems function by learning relevant features directly from huge image databases (typically millions of images). This is in contrast to more traditional pattern recognition techniques, which strongly rely on manually crafted quantitative feature extractors.

In spite of these huge successes, 'deep learning' techniques have not yet made a big impact on the field of medical imaging. One of the main reasons is that for the traditional imaging-based specialties (e.g. radiology) the large numbers of images that are needed to train complex 'deep learning' systems are not readily available. In digital histopathology this is easier: one WSI typically contains trillions of pixels from which hundreds of examples of cancerous glands (in the case of prostate or breast cancer) can be extracted.

Some initial work has been published over the last five years discussing the application of 'deep learning' techniques to microscopic and histopathologic images. Ciresan et al. were the first to apply convolutional neural networks to the task of mitosis counting for primary breast cancer grading. Furthermore, in a different publication, they showed the applicability of patch-driven convolutional neural networks to segmentation tasks. Wang et al. later expanded the work on mitosis detection by combining hand-crafted features and convolutional neural networks. Other applications of convolutional networks include primary breast cancer detection, glioma grading and epithelium and stroma segmentation. Last, Su et al. used another 'deep learning' technique, called stacked denoising auto-encoders to perform cell detection and segmentation in lung cancer and brain tumors.

The number of prostate biopsy sections has strongly increased in the past decades due to the advent of prostate specific antigen (PSA) testing. Because of the nature of the standard biopsy procedure (eight to twelve random biopsies under ultrasound-guidance), each procedure results in

several slides. The majority of these slides typically do not contain cancer. The histopathological analysis could be streamlined significantly if these normal slides could automatically be excluded without expelling any slides containing cancer. We collected consecutive single-center biopsy specimens of 254 patients who underwent MR-guided biopsy procedures for prostate cancer at our institution. These specimens were prepared according to standard histopathologic protocol and subsequently digitized using an Olympus VS120-S5 system (Olympus, Tokyo, Japan).

After digitization of the H&E-stained slides cancer and metastases were manually delineated using a computer mouse by a resident of pathology (I.K., prostate cancer experiment) and a lab technician (M.H., sentinel lymph node experiment), under the supervision of experienced pathologists (C. A. H. K., P. B.). From these annotated areas small prototype image regions (‘patches’) were extracted to train CNNs to detect cancer areas in validation data sets (schematic overview in Fig. 1). These validation data sets were used to optimize the network parameters. After training, the CNN was converted to a fully convolutional network which gave per-pixel predictions on the presence of cancer and metastases in separate, not previously used, test data sets. For prostate cancer detection the CNNs were evaluated on a per-slide level using receiver-operator curve (ROC)-analysis. We also investigated how well the system could exclude slides without cancer from further diagnostic processing. For the sentinel lymph node procedure, we assessed how well the system was capable of identifying individual micro- and macro-metastases using free-response ROC (FROC) analysis and if it is capable of excluding slides which do not contain any metastases using ROC analysis.

Results

Subjects

Prostate cancer

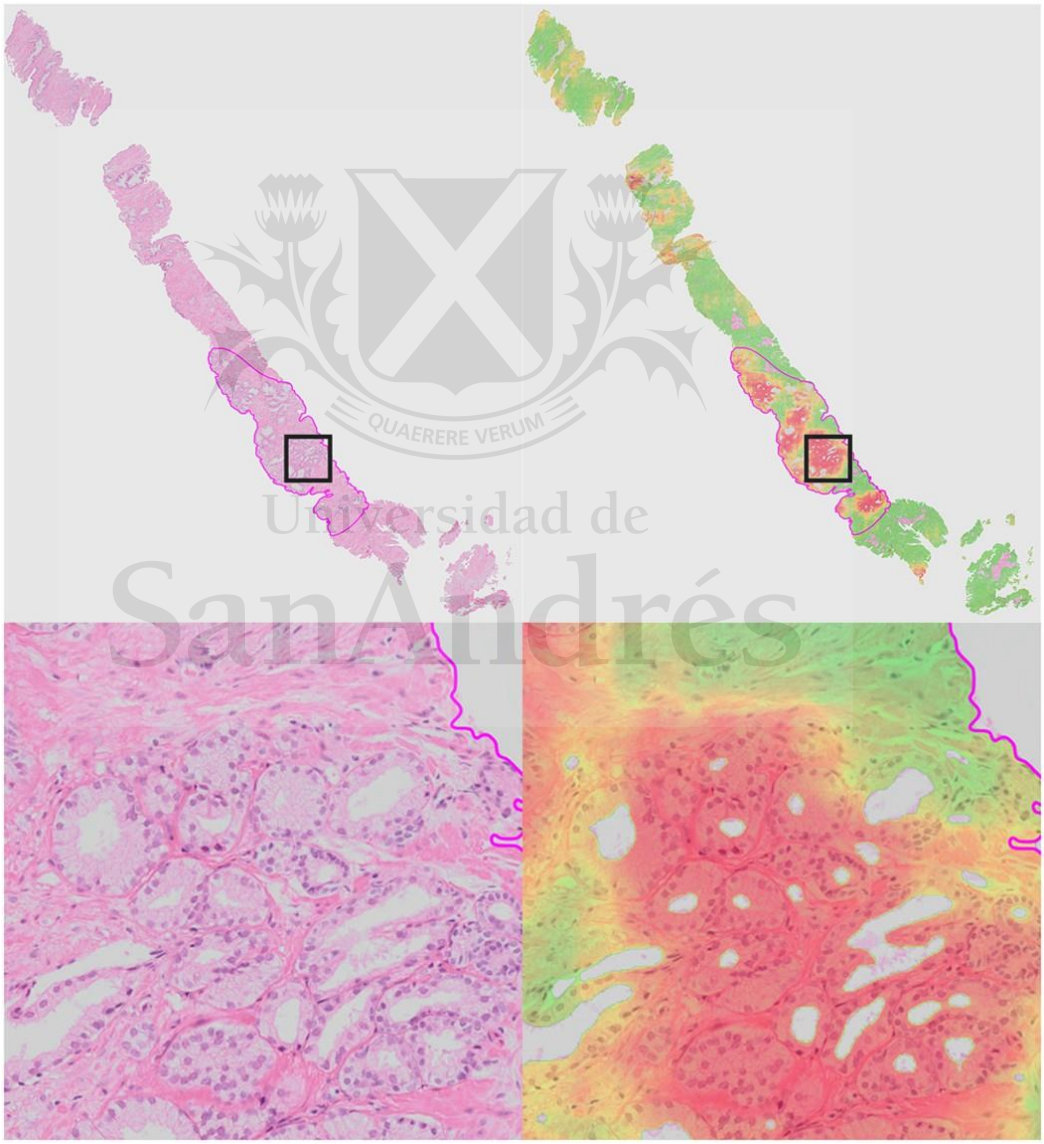
From the initial set of 254 patients, eleven were excluded because the glass slides were not available. Four were excluded because no biopsy was taken during the procedure and one was excluded as the tissue sample was too small for pathologic analysis. Out of the remaining 238, we randomly selected 225 glass slides for digitization, of which 100 were assigned to the training set, 50 to the validation set and 75 to the test set. The training set sampling was stratified such that a near-50/50-distribution between slides containing cancer and slides not containing cancer was obtained. All slides were successfully digitized and annotated.

Breast cancer sentinel lymph nodes

Data collection for the sentinel lymph node experiments was performed in two batches. The first batch was obtained by including 173 slides from the case files of an experienced breast pathologist (P.B). These initial slides were split into a training (98), validation (33) and test (42) set. These slides were subsequently digitized and every metastasis was annotated. To make sure our results were not biased to a single pathologist's case selection, we acquired a second set of data by including all the consecutive sentinel lymph node cases for breast cancer from October 2014 to April 2015, resulting in an additional 98 whole-slide images. For the second batch no on-slide annotations were available, only the per-case outcome (presence of macro- and/or micro-metastases and isolated tumor cells (ITC)).

Prostate cancer detection

A cancer likelihood map (CLM), the output of the CNN indicating cancer likelihood per pixel, for a representative WSI from the test set with cancer covering 30% of the tissue area is shown in next figure. The cancerous glands indicated by the pathologist's outline (in magenta) are correctly identified with high likelihood. The stroma within the annotation areas is correctly identified as a low cancer likelihood region (in green, most easily identifiable in the high-resolution sub-images).



Representative example of a whole slide prostate biopsy specimens with 30% cancer. The top row shows the complete field of view, the bottom row a close up (close-up area indicated by the square rectangle). The second column shows the cancer likelihood map

as an overlay on the original image. Red indicates a high likelihood of cancer, whereas transparent/green indicates a low likelihood.

Although deep learning is an active research field, the application of deep learning to histopathology is relatively new. Most already published work has focused on the detection of mitotic figures9,11 or identification and segmentation of individual cells15,19. One paper used a convolutional auto-encoder to segment basal-cell carcinoma in H&E-images of breast cancer20.However, this model is only evaluated on images from pre-selected regions of interest and not on whole slides, making it difficult to assess its practical value.

The two papers most closely related to our work have focused on different entities. Cruz-Roa et al. used a CNN to detect and segment primary breast cancer12 and Ertosun et al. investigated the grading of gliomas13. We explored the applicability of CNNs to digitized histopathology through two different experiments: prostate cancer detection in H&E-stained biopsy specimens and identification of metastases in sentinel lymph nodes obtained from breast cancer patients. In contrast to these two papers, which perform patch-by-patch classification, we use fully convolutional networks to obtain per-pixel cancer likelihood maps and segmentations in whole-slide images. Furthermore, we are the first to report slide-level accuracies for cancer detection.

In both experiments we were able to successfully train convolutional neural networks, although the amount of case data was less than what is generally typical in 'deep learning' experiments. The fact that we performed extensive data augmentation and boosting in combination with the relatively limited domain (i.e., H&E-stained histopathologic images compared to natural images) made this possible.

In both applications we investigated whether it was possible to identify slides not containing disease without overlooking any slides containing

disease. In the prostate cancer slides, up to 32% of the slides not containing disease could be excluded. For the sentinel lymph nodes, specificity was even higher at 44%, without missing any slide containing micro- or macro-metastases. **This indicates that substantial gains in efficiency are possible by using CNNs to exclude tumor-negative slides from further human analysis.**

Next to the performance of the CNN at high sensitivity, area under the ROC curve was also high in both cases, with an AUC of 0.99 for the prostate cancer experiment (median analysis) and 0.88 for the sentinel lymph node experiment (consecutive set). Furthermore, localization accuracy was high for micro- and macro-metastases in the sentinel lymph node experiment (90% sensitivity at 1 false positive per normal image).

There are some limitations to the application of the CNNs, especially for the sentinel lymph nodes. Although the accuracy of detecting micro- and macro-metastases is high, adding the requirement of having to identify all clusters of isolated tumor cells lowers performance significantly (0.74 AUC for the consecutive set). However, the importance of ITCs is debated. Some have found no prognostic implication of ITCs at all or when the ITCs are visible through immunohistochemistry only. Others did find ITCs having a negative prognostic impact, albeit effect sizes differ. However, **for the clinical application of CNNs this is of limited importance. If the application of the CNNs can detect the micro- and macro-metastases with high accuracy and we have shown this, the ITCs can be detected by immunohistochemistry, without having a pathologist looking at the H&E stained slides**. In The Netherlands, according to the national guideline for breast cancer, immunohistochemistry is mandatory when no tumor is found in the H&E-stained slides.

In the prostate cancer experiment, some detection errors (i.e., false positive detections) still occur at the boundaries of the tissue, mostly due

to tearing and tissue deformation. **These are expected artifacts that occur during histopathologic processing and ideally our CNN would be robust to this.** However, due to the fact that these artifacts can have a wide range of appearances and only occur sporadically, this is not yet the case with the size of the training set used in this study. For the slide level analysis, these spurious detections are not problematic; they occur equally in slides containing and not containing cancer, making their separation still possible.

One further limitation is the fact that we only investigated data from a single center. Although we included data from distinctly different tissue types and used digitization equipment from two different vendors, it is important that these results are confirmed in future, multi-center studies.

As far as the authors are aware, this is the first paper describing the general applicability of a 'deep learning' technique to the diagnostic analysis of whole slide images of sentinel lymph nodes and prostate biopsies. **We have shown that this technique is potentially highly suitable to improve the efficiency of the diagnostic process in histopathology**. **This could in turn lead to adapted protocols, where pathologists perform a more detailed analysis on the difficult samples, as the easy samples are already handled by a computer system.**

Although we specifically looked at clinical diagnosis in this study, the potential of these 'deep learning' techniques reaches further. They could also be used to quickly analyze huge clinical trial databases to extract relevant cases, or automatically annotate areas of disease to allow fast quantification (e.g., area, diameter). Furthermore, the technique is not limited to H&E-stained images and could readily be applied to immunohistochemistry, which might be of

interest when researching the efficacy of drugs or the expression of genes. Both are worthwhile avenues for future research.

(Kinnor Das, 2021) (13)

6) Machine Learning and Its Application in Skin Cancer

According to the US Skin Cancer Foundation, skin cancer affects more people in the United States each year than all other cancers combined, skin cancer including both malignant melanoma and non-melanoma skin cancer (NMSC), are common cancers and their incidence is on the rise.
AI can be of use for the early detection of skin cancer. For example, the use of deep convolutional neural networks can help to develop a system to evaluate images of the skin to diagnose skin cancer. Early detection is key for the effective treatment and better outcomes of skin cancer

Melanoma is the skin cancer with the worst prognosis. If diagnosed early, it can be treated successfully with surgical procedures. However, once there is metastasis, rates of survival are reduced significantly. Diagnosis of melanoma depends on the clinical examination and classic findings on the lesion biopsy. Examples of NMSC include basal cell carcinoma (NMSC) and squamous cell carcinoma. The success of skin cancer depends on early diagnosis and appropriate treatment. Visual inspection may not be sufficient to differentiate benign lesions from malignant tumors. The gold standard procedure is histopathology examination of the skin biopsy. The invasive nature of the procedure, Int. J. Environ. Res. Public Health associated pain, and the need for repeated samples in suspected lesions with varied presentations are some of the limitations for skin biopsy. Non-invasive tools can also assist in clinical diagnosis. Expertise, cost, and availability are the challenges for the widespread use of these tools. Several advancements in science and technology have resulted in the availability of different non-invasive imaging methods to detect melanoma.

Overall, early detection is key for the effective treatment and better outcomes of skin cancers. Specialists can accurately diagnose the cancer, however, considering their limited numbers, there is a need to develop automated systems, which can diagnose the disease efficiently to save lives and reduce health and financial burdens on the patients. Skin tumors can be difficult to recognize from common benign skin lesions, and melanoma has a particularly varied look. AI can aid in the early detection of skin cancer, lowering the burden of morbidity and mortality associated with the disease. In addition to reducing the workload, AI-based systems can also help by improving skin lesion diagnostics.

There is rising optimism regarding applications of AI in healthcare, ranging from assistance in medical diagnostics, treatment and administrative support to reduce timelines of new drug development. It may also be of benefit as an adjuvant in clinical decision making. Dermatology, as a visually intensive field, is at the precipice of an AI revolution. Because skin disease diagnosis is mostly based on visual perception, computer vision algorithms may be able to recognize skin lesions based on their morphology. By September 2018, the US Food and Drug Administration (FDA) had authorized AI approaches for clinical usage, including devices to detect skin cancer from clinical photos obtained via a smartphone app.

Skin Cancer and Deep Learning

Codella used the International Skin Imaging Collaboration (ISIC)-2016 dataset to create a conglomeration of deep learning algorithms and compared them against the performance of eight dermatologists to comprehend 100 skin lesions as either benign or malignant. Their conglomeration outmatched dermatologists, with a precision of 76% and specificity of 62%, compared to a precision of 70.5% and specificity of 59% for dermatologists.

Haenssle used a large dermoscopic dataset with more than hundred thousand benign lesions and melanoma captures to train a deep learning algorithm called InceptionV4, and compared its performance with 58

dermatologists. The level of diagnosis was divided into two categories. In the first level, only dermoscopy was used, while in the second category, dermoscopy was used in addition to clinical information and patient images. In the first level, dermatologists reported a median sensitivity of 86.6% and specificity of 71.3%. The sensitivity and specificity in level II increased to 88.9% and 75.7%, respectively. The improvement in the specificity was statistically significant ($p< 0.05$).

However, the improvement in the sensitivity was statistically non-significant ($p=0.19$). The deep learning CNN receiver operating characteristics curve showed a significantly higher specificity than for the dermatologists in level I ($p< 0.01$) and level II ($p< 0.01$). In this study, CNN outperformed most dermatologists, suggesting a promising role in the detection of melanoma using dermoscopic images.

Another study by Brinker et al. showed similar results. In this study, investigators used a convolutional neural network (ResNet50) to compare the efficacy of 157 dermatologists on hundred dermoscopic images (MClass-D). The dermatologists had an overall sensitivity of 74.1% and a specificity of 60.0%, whereas the deep learning method had a specificity of 69.2% and a sensitivity of 84.2%. In a head-to-head comparison, the performance of CNN was better than 86.6% of dermatologists in the study. The performance was better across subgroups of dermatologists based on experience in the classification of dermoscopv melanoma images. Thus, CNN has significant potential to assist dermatologists in the accurate diagnosis of melanoma.

Tschandl et al. used convolutional neural networks such as InceptionV3 and ResNet50 to diagnose non-pigmented skin malignancies using a mixed dataset of 7895 dermoscopic and 5829 close-uplesion photos. The results were compared to those of 95 dermatologists, separated into three groups based on experience. With beginning and intermediate raters, the deep learning algorithms attained an accuracy like humans and outmatched the human groups. The area under the ROC curve of the trained combined CNN was significantly higher than for the

dermatologists. It showed correct diagnoses in a higher percentage of cases than for overall dermatologists, but not compared to the experts, i.e., more than 10 years of experience.

Maron et al. tested the sensitivity and specificity of a ResNet50 deep learning system for multiclass categorization of skin lesions, along with 112 German dermatologists. The sensitivity and specificity of primary end-point of correct classification of skin lesions for dermatologists was 74.4% and 59.8%, respectively. At a similar level of sensitivity, the specificity of the algorithm was 91.3%. For the secondary end point of correctly classifying a given image into one of the five diagnostic classes, dermatologists had a sensitivity and specificity of 56.5% and 89.2%, respectively. At a similar sensitivity level, the algorithm provided 98.8% specificity. Overall, for the primary end point, dermatologists were significantly outmatched by the deep learning algorithm ($p< 0.001$). The comparison for the secondary end point also showed an outperformance of the algorithm over dermatologists in all categories, except basal cell carcinoma, for which the algorithm had similar performance as that of dermatologists.

On a dermoscopic test set of 100 instances, Haenssle et al. weighed up InceptionV4-based deep learning architecture with dermatologists. This study had two levels: level I was a dermoscopic image, and level II had a clinical close-up image, a dermoscopic image, and clinical information. The deep learning system had a sensitivity of 95% and specificity of 76.7%; however, the dermatologists in level I had a mean sensitivity of 89% and specificity of 80.7%, respectively. The dermatologists' mean sensitivity reached 94.1% with extra information in level II, while their mean specificity remained same. Tschandl et al. conducted an open, web-based study to diagnose the dermatoscoping images. The investigators juxtaposed the average potential of the AI algorithms (139 in total) and 511 human readers on an experimental 1511 set of photos in the ISIC 2018 competition. The diagnoses (seven predefined categories) provided

by the humans were compared with those from the algorithms prepared from machine learning.

The differences in the percentage of correct diagnoses were compared. Out of the human participants, 55.4%, 23.1%, and 16.2% were board-certified dermatologists, residents of dermatology, and general practitioners, respectively. The results showed a mean of 2.01 for more correct diagnoses by the algorithms than the humans. The difference was statistically significant ($p < 0.0001$). As a result, the AI algorithms were able to make more accurate diagnoses than the human readers.

Algorithms for Machine Learning in Skin Cancer

Because of the high prevalence of skin malignancies, an increasing number of people require prompt diagnosis and ongoing monitoring. This places a huge strain on specialist medical services, which may be allayed by better patient self-surveillance techniques as well as the use of decision support systems for less experienced physicians. Machine diagnosis is not subjective, and is not impacted by external factors. However, human diagnosis is associated with subjective variations and may be impacted by some external factors. If implemented with the necessary regulations, the use of AI for the detection and progression of skin cancer may result in fewer biopsies. Following a training intervention, patients with skin cancer and their guardians can perform self-skin examination (SSE). This also boosts teledermoscopy, leading to fewer medical consultations.

The inclusion of AI in smartphone applications can teach people to perform skin examination and forward the information to the physician. Each form of skin lesion is assigned a class, such as “benign” and “malignant”, or “naevi” and “melanoma”, in order to construct a new ML skin cancer algorithm. Deep learning algorithms are taught on a large number of photos in each class before being evaluated on a new image. There are three basic parts to the procedure. In the first stage, the algorithm is fed digitized macroscopic or dermoscopic images labelled

with the “ground truth” in the first stage (in this case, the ground truth is the lesion diagnosis, which is determined by an experienced dermatologist or by histological study). In stage 2, convolutional layers extract the feature map from the images. A feature map is a visual representation of the data, which has several degrees of abstraction. Low-level features such as edges, corners, and forms are extracted by the first convolutional layers. To recognize the type of skin lesion, later convolutional layers extract high-level data. The machine learning classifier uses the feature maps in stage 3 to recognize distinct kinds of skin lesion patterns. A fresh image can now be classified using the deep learning method.

Skin Cancer Datasets

Particularly in dermatology, clinical and dermatoscopic images are often generated to track changes in skin conditions. New applications will make the gigantic amounts of data that already exist and will be created in the future, e.g., in hospitals, accessible to algorithms and lead to an improvement of CNNs. There are already data sets accessible for research. ISIC archive gallery contains numerous of clinical and dermoscopic skin lesion datasets, including the ISIC Challenges datasets, HAM10000, and BCN20000. Interactive Atlas of Dermoscopy has 1000 clinical examples including 270 melanomas and 49 seborrheic keratoses. Each case has a minimum of two images—dermoscopic and close-up. Its price is €250 and is available for research purposes. Dermofit Image Library has 1300 high-resolution photographs of skin lesions divided into 10 categories. A licensing agreement is required, with a one-time license charge with the availability of academic license. PH2 Dataset contains 200 dermoscopic images, including 40 melanoma and 160 nevi cases. It is free for downloading after the completion of an online registration form. MED-NODE Dataset contains 170 clinical photos, including 70 melanoma and 100 nevi cases. This dataset can be downloaded without any cost for research purposes. Asan Dataset contains 17,125 clinical photos of 12 different forms of skin illnesses that affect Asians. It is available to

download for research purposes. The Hallym Dataset has 125 clinical photos of BCC cases (34Han JID). SD-198 dataset contains 6584 clinical photos of 198 skin illnesses. The 25 SD-260 dataset is more balanced than SD-198 dataset, since it manages the class size distribution while preserving 10–60 photos for each category. There are 20,600 photos in all, representing 260 skin illnesses. Dermnet NZ is the source of one of the most comprehensive and diverse collections of clinical, dermoscopic, and histology photographs. Additional high-resolution pictures are available for purchase. Derm7pt contains 1011 dermoscopic images including 252 melanoma and 759 nevi cases based on a seven-point checklist. The Cancer Genome Atlas has 2871 pathological skin lesion slides, making it one of the largest collections of its kind. It is openly available for usage by the research community.

Deep Learning and Clinical Images

Clinical photos of various skin lesions are routinely captured using cell phone cameras for remote assessment and assimilation into patient medical records. On the SD-198 dataset, Yang et al. achieved clinically observed skin lesion identification utilizing the well-known ABCD rule. They compared the performance of deep learning methods with dermatologist outputs. It received 57.62% accuracy compared to the 53.35% accuracy for the best performing deep learning system (ResNet). Only senior experienced clinicians had an average accuracy of 83.29% when compared to the rest of the clinicians. Han et al. used a MED-NODE dataset and atlas site images for training a deep learning architecture (ResNet-152) to differentiate clinical photos of 12 skin illnesses, and then examined it on an Asan testing set and an Edinburgh Dataset (Dermofit). Upon taking 480 random photos merged from the Asan test dataset (260 images) and the Edinburgh dataset (220 images), the algorithm's performance was equivalent to a team of 16 dermatologists, but the AI system outclassed dermatologists while diagnosing basal cell carcinoma (BCC). Fujisawa et al. trained deep CNN with 4867 clinical images from

14 skin diseases from 1842 patients with different diagnoses, including malignant and benign diseases, to evaluate a deep learning algorithm. The results of the algorithm were compared with those of the dermatologists. The deep learning algorithm produced a diagnostic accuracy, sensitivity, and specificity of 76.5%, 96.3%, and 89.5%, respectively. The accuracy of classifying images into the benign or malignant category by the dermatologist's board certified dermatologists (n= 13), dermatology trainees (n= 9), and deep CNN was 85.3%, 74.4%, and 92.4%, respectively. The performance by the board-certified dermatologists was significantly better than for the trainees. However, the accuracy of deep CNN was higher than for both human raters.

In a test case of 100 clinical skin lesion photographs (MClass-ND), Brinker et al. evaluated 145 dermatologist performances and a deep learning approach (ResNet50) for the 80 nevi cases and 20 histopathologically proven melanoma cases. The dermatologists had sensitivity of 89.4% and a specificity of 64.44%, while a deep learning technique at the same sensitivity had a mean specificity score of 68.2%. Overall, the CNN performance was on par with dermatologists in terms of the classification of clinical images. Variance with CNN was smaller, suggesting a greater robustness of AI than humans for the classification of images. Only 19 (13.1%) dermatologists had a higher sensitivity than the CNN. Out of these 19 dermatologists, 16 (84.2%) achieved a sensitivity of more than 95%.

Deep Learning and Histopathology Images

Dermatopathologists confirm the diagnosis of skin cancer through histopathologic examination of a tissue biopsy under a microscope. One of the important challenges in the confirmatory diagnosis of skin cancer is the high rates of discordance between different pathologists. In the case of the diagnosis of melanoma, there can be discordance in classifying whether it is a benign or malignant lesion. With whole-slide imaging, deep learning methodologies have been successful for digital pathology. These

methods are used to classify biopsy tissue specimens in order to diagnose malignancies. Different investigators have performed studies to compare the performance of an expert versus that of AI system. Heckler et al. compared pathologists' performance for identifying melanoma and nevi using a deep learning approach. The study included 695 lesions classified by an expert, of which 595 were used for training the CNN. The remaining 100 were used to test the results of CNN with those of the 11 experts. In this study, the investigators digitalized the entire slides. The image sections with magnification were randomly cropped. The sensitivity, specificity, and accuracy of the CNN was compared with that of the pathologists. In a recently published study, Brinker et al. reported comparative results of the ability of CNN to differentiate melanomas from nevi using hematoxylin − eosine stained whole slide images (WSI). In this study involving whole slide images of 50 melanomas and 50 nevi, the performance of CNN was on par with the experts. Jiang et al. came up with a deep learning method for diagnosing BCC using smartphone-captured histopathology images. They found that the algorithm's performance on smartphone-captured images and WSI was comparable, accompanied by an AUC of 0.95. For an in-depth analysis of the difficult cases, they used a deep segmentation network, which resulted in a score of 0.987 (AUC), 0.97 (sensitivity), and 0.94 (specificity). The work of Jiang and colleagues suggests the usefulness of deep learning methods for the diagnosis of BCC, with a high sensitivity and specificity.

On 1417 images from 308 regions of interest (ROI) of skin histopathology images, Cruz-Roa et al. employed deep learning architecture to identify between BCC and normal tissue patterns. They compared deep learning to classical ML using feature descriptors such as the bag of features, canonical wavelet transforms, and Haar-based wavelet transform.

The deep learning architecture outperformed previous approaches with an F-Measure of 89.4% and a balanced accuracy of 91.4%. From 2008 to 2018, Xie et al. published a humongous dataset of 2241 histopathology pictures from 1321 individuals. They tested the categorization of

melanoma and nevi on various magnification scales by two deep learning architectures, viz. VGG19 and ResNet50, by making use of the 9.95 million patches created on 2241 histopath images. With a mean F1 (0.89), specificity (0.94), sensitivity (0.92), and AUC (0.94), they were able to identify melanoma from nevi with a good accuracy (0.98). It should be noted that different results from different studies suggest that the amount of data presented to the AI system, the methodology used for the study, and the complexity of disease may affect the level of difficulty for a given task and thus the performance of both AI algorithms and human observers.

Overall, it seems that CNNs can be of valuable assistance to humans for the diagnosis of skin cancers such as melanoma. Similarly, the diagnosis of BCC needs intensive work because of the need to examine a large number of images. Deep learning methods can be of use to assist the diagnosis of BCC. WSIs and microscopic ocular images with use of smartphone cameras can be useful for developing neural network models for the diagnosis of BCC. A reduced time for the diagnosis and cost benefit are some of the advantages of CNN in the diagnosis of skin cancers.

(Kinnor Das, 2021) (14)

Chapter V: AI applied to new drugs investigation

Having reviewed the current use that is being given to deep learning technology in the early detection of different types of cancer and in the handling of histopathological samples, a second use of the same technology refers to its application in the R&D process of new drugs for the treatment of the disease. The historical model used in pharmacological research is at a point of exhaustion due to its long times and high costs, as well as the low percentage of projects that reach the market compared to those that fail to pass any of the testing phases. Next, I will present two real cases that show, as in the case of premature diagnosis, that AI can contribute a great step towards the goal of obtaining better drugs for the treatment of the disease in less time.

Deep Learning for Drug Discovery and Cancer Research: Automated Analysis of Vascularization Image

Likely drug candidates which are identified in traditional pre-clinical drug screens often fail in patient trials, increasing the societal burden of drug discovery. A major contributing factor to this phenomenon is the failure of traditional *in vitro* models of drug response to accurately mimic many of the more complex properties of human biology.

The total cost of bringing a new drug from discovery to approval has exhibited a steady, exponential rise over the past five decades. One contributing factor to this phenomenon, dubbed Eroom's law (Moores law backwards), appears to be the failure of traditional, pre-clinical models to accurately simulate many of the more complex features of their clinical successors. These pre-clinical, in vitro studies serve to quickly and cheaply identify compounds that exhibit promising effects for further study in vivo. However, traditional 2D monolayer culture systems (i.e., petri dishes) lack many features that are present in vivo, such as 3D cellular structure, heterogeneous cellularity, cell-cell interactions, the presence of a complex extracellular matrix (ECM), biomechanical forces (e.g., shear forces generated by fluid flow), and the presence of perfused vasculature. Animal studies, on the other hand, are too complex to analyze and

expensive to substitute for invitro pre-screening, and often fail to identify potential human toxicity due to physiological differences between humans and the animal model. In short, a compound that appears effective in traditional, pre-clinical studies may fail spectacularly in the human body, further contributing to the costly societal burden of failed clinical trials.

Microphysiological systems (MPSs), or "organ-on-a-chip" platforms, promise to help close the gap between in vitro and in vivo drug screens, and have seen rapid, recent development, supported in part through private-public partnerships fostered under the auspices of the National Center for Advancing Translation Science. These MPSs make significant strides toward more accurately modeling the pertinent properties of in vivo biological environments for drug discovery, however many remain in a proof-of-concept stage and require complex peripheral equipment and accessories to operate and maintain.

An MPS for growing vascularized, perfused microtissues produces highly robust and uniform vascular networks which are suitable for screening anti-tumor compounds and in large-scale drug discovery studies, all while requiring little additional training for the user and no added equipment beyond a standard incubator. The survival of these miniature tissues is dependent on nutrients delivered through living vasculature. Importantly, by accurately identifying drugs that target tumor cells, the vascular networks that supply them, or both, the system has proven much better at mimicking human drug responses than previous models. In the studies using FDA-approved or clinical trial compounds to target the vasculature, we have found that antiangiogenic compounds such as sorafenib and axitinib induce regression on sprouting vessels, but do not have profound effect on mature, interconnected vascular networks. Therefore, they often show a milder effect on the vasculature. On the other hand, non-specific, anti-vascular compounds such as bortezomib and vincristine aggressively fragment the vascular network. In brief, this system exhibits exceptional potential for developing more targeted, effective anti-vascular and antiangiogenic compounds to target the tumor vasculature without adverse effects on normal tissue.

A remaining obstacle to deploying this system for truly large-scale anti-angiogenic and anti-vascular drug screening is the need to have human experts determine whether each compound is effective in targeting the vasculature network.

Effects are categorized as:

- No-hits (i.e., the compound had no effect on the vasculature network).
- Soft-hits (i.e., the compound moderately disrupted the vasculature network or induced vascular regression).
- Hard-hits (i.e., the compound had a devastating effect on the vasculature network).

From a primary screening soft-hit and hard-hit compounds can be further analyzed in a dose-response screen to identify the half maximal inhibitory concentration (IC50), optimized for molecular structure, and subsequently characterized for their pharmacokinetics in vivo. Soft-hit compounds are treated as anti-angiogenic while hard-hit compounds are treated as anti-vascular. In the past, human raters have made this determination by manually analyzing each pair of before- and after-drug-application images and quantifying their total vessel length difference using “AngioTool”. However, this workflow is imprecise—e.g., in its insensitivity to anti-angiogenic compounds that do not significantly affect total vessel length of a fully mature vascular network and its reliance on subjective human judgment—and low throughput—for its need to carefully tune several dataset-specific parameters in the software and the time it takes a human to look at each image.

Automatic classification of these images via machine learning could provide an attractive replacement to the slow and error-prone process of requiring human ratings. In this paradigm, a set of carefully hand-labeled images would be fed to a classifier which could "learn" to distinguish between classes. A convolutional neural network is a type of machine learning model that is particularly suited to applications in computer vision. Not only do they offer state-of-the-art performance in general image classification tasks, they have also proven effective for biological

applications, with past work demonstrating convolutional networks capable of detecting cardiovascular disease, spinal metastasis, and skin cancer from medical images.

A model of a convolutional neural network to automatically classify images of vasculature networks formed in a MPS into no-hit, soft-hit, and hard-hit categories. The accuracy of the best model is significantly better than our minimally-trained human raters and requires no human intervention to operate. This model is a first step toward automation of data analysis for high-throughput drug screening and applications of machine learning in drug discovery can predict drug-related properties of small molecules such as binding affinity, toxicity, and solubility.

METHODS

A. Data Collection

Drug studies were performed in the MPS as previously described. Briefly, the cell-ECM suspension was loaded into the platform and cultured for 7 days to allow the vascular network to develop inside the tissue chambers. Each tissue unit was exposed to various compounds obtained from the National Cancer Institute (NCI) Approved Oncology Compound Plate or purchased from Selleck Chemicals. Time course images of vascular network before and after drug treatment were taken using a Nikon Ti-E Eclipse epifluorescent microscope with a 4Œ Plan Apochromat Lambda objective. For close-up imaging of the tissue chambers, a 1.5Œ intermediate magnification setting was used.

B. Preprocessing

Each image in our dataset was between 1000 and 1300 pixels wide. Images of this size contain far more information than is needed for deep image classification, so we down sampled images to create 4 separate constant-size datasets: one each of 128×128px, 192×192px, 256×256px, and 320×320px. Next, we z-normalized each image, subtracting the mean pixel intensity and dividing by the standard deviation of the pixel intensities within that image to obtain images with 0-centered pixel values and unitary standard deviation. This normalization helps our models to converge more quickly and uniformly across

random initializations. After all this, we concatenated the pre-drug-application and post-drug-application images to obtain a single, 2-channel image.

1) Image Alignment: We would like the pre-drug-application and post-drug-application images to spatially align as closely as possible. If they do not, then our model would be required to learn an extra invariance: that the channel images need not be aligned. Because the pre- and post-drug-application images were captured three days apart, it is not in general possible to ensure that the two images will be perfectly aligned (e.g., the later image might be shifted or rotated slightly compared to the original). To combat this effect, we implemented a rigid alignment preprocessing step to align the post-drug image to the pre-drug image using the warpAffine method in OpenCV3. For each image, we tried three sets of transformations:

1) A single Euclidean (translation + rotation) transformation on the full-resolution image.

2) A Euclidean transformation on a smaller (32x32px) copy of the image followed by a Euclidean transformation on the full-resolution image.

3) A translation-only transformation on a smaller (32x32px) copy of the image followed by a Euclidean transformation on the full-resolution image.

From these three, we selected the transformed version which yielded the highest possible correlation coefficient between the pre- and transformed post-drug image. See figure 2 for two examples of this alignment process in action.

C. Human Ratings

Two human experts rated each of the 277 images, comparing disparate ratings where necessary to come to a consistent set of gold-standard ratings. 164 images were labeled as 0 or no-hit (59.2%), 52 were labeled as 1 or soft-hit (18.8%), while 61 were labeled as 2 or hard-hit (22.0%). These ratings are used throughout the remainder of this paper.

We also obtained ratings from 4 additional humans: undergraduate research assistants who were trained to recognize each image class and who had been assigned this task in the past. Raters were presented with the full set of 277

images in randomized order and were asked to provide an integer class assignment for each using the following instructions: “How much of an effect did the drug have? (0 for no effect, 1 for solid effect, 2 for devastating effect)”.

D. Loss Weighting

For the purposes of drug discovery, false negatives are potentially much costlier than false positives. A false positive (i.e.: predicting that an image from an ineffective drug was actually effective) will result in secondary screening in which the ineffectiveness of the drug may be confirmed. A false negative (i.e.: predicting that an image taken from an effective drug did not actually have any effect) may result in a potentially useful compound being overlooked in this and any future drug trials.

To help control our model’s false-negative rate, we employed a weighted cross-entropy loss function of the form:

$$\text{loss}(y_i, \hat{y}_i | W) = -\sum_{c=0}^{c=2} W_{c_{itrue},c}\, y_{ic} \log(\hat{y}_{ic})$$

where i indexes over datapoints, cover classes, yic is an indicator variable that takes the value of 1 if the true class of datapoint i is c and 0 otherwise, citrue represents the true label of datapoint i(i.e.: 0, 1, or 2), and the weights Wcitrue, c are drawn from the hand-tuned confusion weighting matrix shown in table I. Note that if all elements of this weight matrix were set to 1.0, then our weighted cross-entropy loss would reduce to standard cross-entropy

TABLE I
LOSS FUNCTION WEIGHT VALUES

	$Y_{ipred} = 0$	$Y_{ipred} = 1$	$Y_{ipred} = 2$
$Y_{itrue} = 0$	0.8	0.8	0.8
$Y_{itrue} = 1$	2.0	1.0	0.8
$Y_{itrue} = 2$	2.0	0.8	1.0

This loss function penalizes false negatives at twice the default value. In addition, it penalizes the treatment of all true no-hit images at 0.8 times the

default value and reduces the penalties for confusing soft- and hard-hits to the same amount.

We arrived at these weights through trial and error and use them for all experiments presented in this paper.

E. Training Procedure

We partitioned the full dataset of 277 images into a test set consisting of 25% of the images (69 images) and a training + validation set consisting of 75% of the images (208 images). We employed 4-fold cross validation on the training + validation set, training on 75% of its datapoints (156 images) and tracking validation loss on the remaining 25% (52 images). Unless otherwise noted, we trained on each fold for a total of 200 epochs. All deep neural networks presented in this paper were built in Keras and trained on NVIDIA GPUs. We selected the model from each fold which attained the lowest validation-set loss value across all training epochs.

We combined the best models from each fold into a 4-model ensemble of models. We averaged the predictions across all 4 models in the ensemble to attain final predictions for each set of hyperparameters on the test dataset.

1) Data Augmentation: Since our training set is rather small, we employed random data augmentation during training. In each pass over the data, each training image was randomly rotated between -5 and 5 degrees clockwise, translated between -5% and 5% vertically and horizontally, zoomed in between 0 and 10%, and possibly flipped horizontally and vertically, with each transformation value selected uniformly at random from the legal range. Empty pixels that resulted from the random rotation and translation were filled with the values from their nearest existing neighbor pixel. This random data augmentation scheme with continuous parameters yields an infinitude of variations for each 156-image training set and helps prevent our models from overfitting to the specific details of our training data.

At inference time, we randomly generated five versions of each validation or test image and averaged the model's predictions for each image over all five of its randomly-generated copies.

F. Baseline Models

We trained a number of increasingly complex machine learning models on the data to use as comparison baseline:

1) Logistic regression on the raw data.

2) Logistic regression on a bag of words (BoW) representation of SIFT or SURF features; and

3) RBF-kernel support vector machine classifiers (SVMs) trained on a BoW representation of SURF or SIFT features. We used the SVM-classifier and logistic regression implementations provided by scikit-learn1.

The logistic regression model was trained on images of sizes 128×128, 192×192, 256×256, and 320×320 with varying L2 regularization using the LBFGS optimizer, treating all concatenated pixels of both the pre- and post-drug application images as single input vector. We used OpenCV2to extract SIFT and SURF features from the images at a resolution of 320×320, which yields a varying number of key-points/features per image. To build a bag of words representation we first clustered all SIFT/SURF descriptors of the training set with k-means. Then we mapped all descriptors to their nearest centroid (as found by k-means) and compute the histogram of these centroid mappings for each image separately. We experimented with two histogram normalization approaches: globally rescaling the bins to the 0−1range or a binary representation that encodes whether at least one SIFT/SURF descriptor from the image was assigned to a given centroid. We treated pre- and post-drug application images separately and concatenated their BoW representations (the histograms) into one feature vector, as computing the BoW representation across SIFT/SURF features of both images together discards crucial discriminatory information and resulted in a reduced performance in

preliminary experiments. We optimize the L2 regularization coefficient as well as the size of the BoW representation (number of clusters) for all models.

G. Convolutional Neural Network Models

Convolutional neural networks are based on a weight-sharing scheme in 'convolutional' layers. These layers learn translation-invariant filters that are applied to e.g., all pixels of an image in the case of computer vision, and have led to models achieving state-of-the-art classification performance on a variety of tasks. Standard convolutional architectures for image classification include a series of convolutional layers followed by one or more fully connected layers. Each convolutional and fully connected layer is followed by a rectified linear unit (ReLU) nonlinearity and max pooling layers are interspersed through some subset of the convolutional layers to repress non-maximal responses and reduce the number of parameters in subsequent layers. Dropout may also be used on some of the convolutional and fully connected layers to help prevent overfitting.

Overall, convolutional neural networks offer a well-established process for performing high-quality image classification.

H. Hyperparameter Search for Convolutional Architectures

Building a convolutional neural network requires specifying a large number of hyperparameters, such as the number of convolutional and fully-connected layers in the network, the size of each layer, dropout probabilities etc. The number of possible hyperparameter combinations grows exponentially with the number of hyperparameters, so a thorough grid search of hyperparameter combinations quickly becomes unwieldy.

Instead, we employ a Gaussian-process-based meta-model which maps from a set of chosen hyperparameters to an estimate of the out-of-sample accuracy attained by a model trained with the given hyperparameters. This meta-model of hyperparameter fitness is used in an outer-loop hyper-parameter optimization process. First, the meta-model proposes a hyperparameter set to try. For each hyperparameter set, we follow the same training procedure as that detailed in section II-E, using 4-fold cross-validation on the training + validation set,

building a 4-model ensemble from the best version of the model for each fold (across epochs and as judged by validation-set accuracy), and averaging each model's validation- and test-set predictions over 5 randomly generated versions of each input image. At the end of training, we report the validation-set accuracy (averaged across all 4 folds) as the objective value attained for the given hyperparameter set. This objective value is used to update the meta-model of hyperparameter quality and the process repeats.

I. Pre-Trained Convolutional Architecture

Given the small size of our training dataset, we next tried a large convolutional architecture that had been pre-trained on a large, general purpose image recognition problem. For this purpose, we picked the InceptionV3 architecture as implemented in Keras with weights that had been pre-trained on the ImageNet classification challenge. The full convolutional portion of the InceptionV3 model contains 21,611,968 parameters and some 216 layers. We instantiated the model without including the final fully-connected layers, opting not to fine-tune its convolutional weights, but to train two fully connected and one 3-class softmax layer anew for our classification problem while using the convolutional portion of the InceptionV3 model as an elaborate, fixed computer vision preprocessing routine. While fixing our convolutional architecture fixed many of the hyperparameters of our model, several still remained. These were: the input image size, the number of neurons in the fully connected layers, dropout probabilities for the dropout layers before and after the fully connected layers, the optimization batch size, the learning rate, and L1 and L2 regularization coefficients. Hyperparameters that control the amount of dropout, or the strength of the L1-, and L2-penalty terms have a regularizing effect and reduce the chances of overfitting the data, whereas the exact effect for other hyperparameters is in general more difficult to estimate. The exact ranges of hyperparameters that we optimized can be found in the supplementary material.

J. Custom Convolutional Architecture

Though the Inception architecture employed in section II-I has proven very useful for general-purpose image classification, the images of microscopic

blood vessel networks used in this task have their own structure that does not necessarily match the constraints of general object recognition3. For this purpose, we also trained a series of custom convolutional architectures specifically for this blood-vessel classification task. We constrained our architecture to contain several convolutional layers followed by two fully connected layers. The hyperparameters that we optimized were: the input image size, the number of convolutional layers, number of convolutional filters, and number of neurons in fully connected layers in the model, the size of the max pooling receptive fields, the optimization batch size, and parameters related to model regularization: dropout probabilities and L1- and L2 penalty terms.

RESULTS

A. Human Rating Results

The four human raters found the vessel rating task difficult compared to the expert raters, matching the gold-standard ratings 72.9%, 76.5%, 69.3% and 83.0% of the time. The rounded average of all four raters' ratings (i.e.: 0, 1, or 2) matched the gold standard ratings 85.9% of the time.

TABLE II
TEST SET CONFUSION MATRIX FOR AVERAGE OF FOUR HUMAN RATERS

	$Y_{ipred} = 0$	$Y_{ipred} = 1$	$Y_{ipred} = 2$
$Y_{itrue} = 0$	86%	14%	0
$Y_{itrue} = 1$	27%	65%	9%
$Y_{itrue} = 2$	0	0	100%

B. Baseline Models Results

The best logistic regression model trained on raw pixels obtained an average validation set accuracy of 79.6% across five repeated five-fold cross-validation experiments, using an input image size of 320px×320px and an L2 regularization strength of 0.05. This model obtained an average three-class test

accuracy of 73.3%, which is a notable improvement over guessing the majority class (62.3%). An even higher accuracy was reached by models using a bag of words (BoW) representation of SIFT or SURF features. The best such model was a support vector machine (SVM) using SURF features that were clustered into a binary feature vector of size 200, obtaining a validation accuracy of 84.4% and test-set accuracy of 78.0%. This is almost 5% better than the logistic regression model that was trained on raw pixels. A summary of best models, as determined by their validation accuracy, for each category is given in Table III; all BoW model results are averages over three repeated full cross-validations runs.

TABLE III
AVERAGE TEST- AND VALIDATION SET ACCURACIES OF BASELINE MODELS

Model	Features	Validation Acc.	Test Acc.
Logistic	raw pixels	79.6%	73.3%
BoW Logistic	SIFT	82.7%	77.7%
BoW Logistic	SURF	81.3%	77.6%
BoW SVM	SIFT	82.0%	76.0%
BoW SVM	SURF	**84.4%**	**78.0%**

C. Pre-Trained Convolutional Neural Network Results

We explored a total of 100 hyperparameter sets for the pretrained convolutional architecture4using the procedure explained in section II-H. The best model, as judged by three-way validation-set accuracy (87.0%), used 320px×320px input images, its first fully connected layer after the InceptionV3 convolutional stack contained 256 neurons, its second fully connected layer contained 1024 neurons, and the final dropout probability before the 3-way softmax layer was 0.27. The optimization was completed with a batch size of 16, log10 of the learning rate of -1.24, a per-epoch learning rate decay factor of 0.98, log10 of L1 shrinkage of -9.0, and log10 of L2 shrinkage of -1.0.

A 4-model ensemble based on this architecture achieved a three-class accuracy value of 87.0% on the hitherto-unseen test (see the confusion matrix in table IV for details).

TABLE IV
TEST SET CONFUSION MATRIX FOR PRE-TRAINED CONVOLUTIONAL ENSEMBLE

	$Y_{ipred} = 0$	$Y_{ipred} = 1$	$Y_{ipred} = 2$
$Y_{itrue} = 0$	98%	2%	0
$Y_{itrue} = 1$	45%	36%	18%
$Y_{itrue} = 2$	0	7%	93%

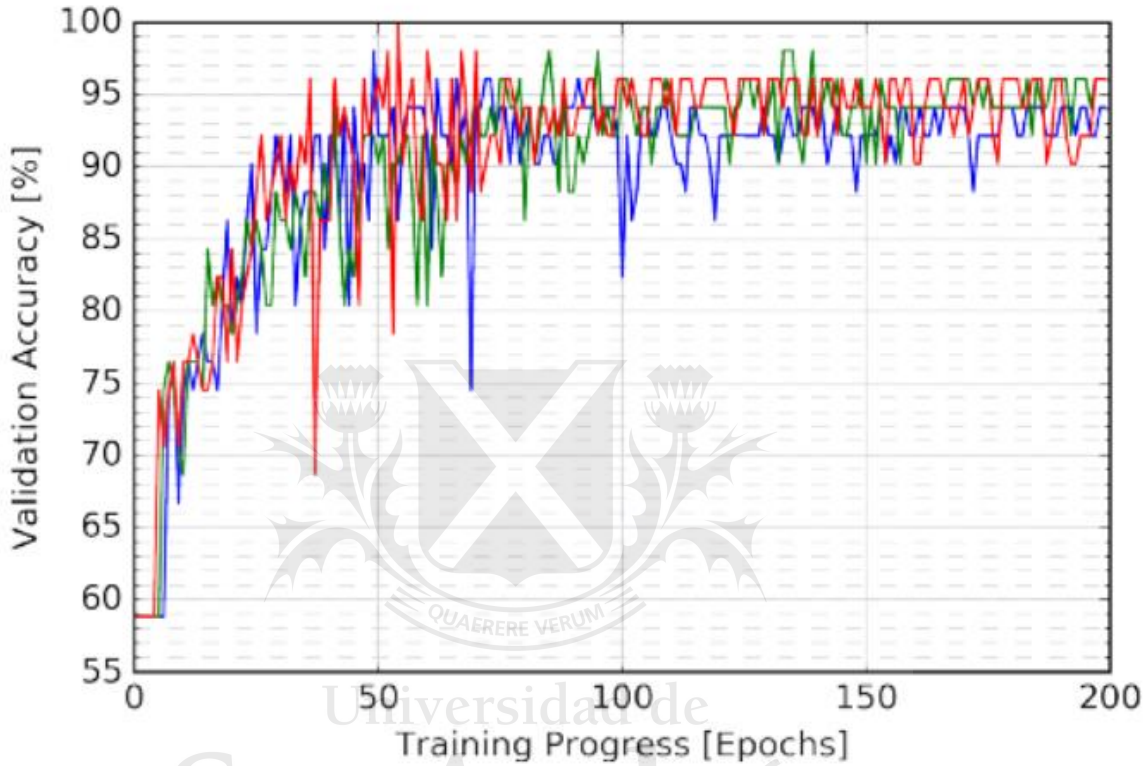
D. Custom Convolutional Neural Network Results

We explored a total of 1000 hyperparameter sets for our custom convolutional architecture, the best of which, as judged by three-class validation-set accuracy (96.6%), is a 21-layer convolutional neural network, the architecture for which is illustrated in figure 4. The optimization was completed with a batch size of 1, a learning rate of 0.012, a per-epoch learning rate decay factor of 0.98, and both L1 and L2 coefficients at a value of 10−9.

A 4-model ensemble based on this architecture achieved a three-class accuracy value of 95.7% on the hitherto-unseen test set with no false negatives (see the confusion matrix in table V for details).

The data set (with 277 datapoints) is small in comparison to typical machine learning data sets, which raises concerns over potential overfitting of deep learning models. To shed light on whether overfitting occurs we plot the evolution of the validation accuracy for three, independently - randomly - initialized and trained, instantiations of our custom CNN model in Figure 5. The curves are not smoothed and thus, as expected, relatively jagged due to the small size of the dataset and the various noise-injecting regularization techniques. Interestingly, we observe no evidence of overfitting within 200 epochs of training. Overfitting would have manifested as a decline in the average validation accuracy towards the end of training, but instead we only

observe a reduction in the variance of validation accuracies. In short, we conclude that the employed model regularization techniques are very effective and that early stopping is, while still beneficial, not as crucial as initially expected.



It is desirable to further estimate the sensitivity of the model to the number of samples in the training set. To this end we artificially and progressively reduce the amount of training data, while keeping all other factors identical (e.g., model architecture, hyperparameters, validation set). Figure 6 presents results from training the custom CNN with ten different training set sizes in 10% increments, repeating the four-fold cross-validation training process four times for each training set size and averaging over these. As expected, decreasing the amount of training data directly reduces the validation accuracy. Interestingly, we also find that the CNN is able to match or outperform the best baseline model (an SVM trained on SURF features with a validation accuracy of 84.4%) when trained on only 40% of the original training data. Further, from extrapolating the graph beyond the 100% point, it seems virtually guaranteed that having access

to more training data would enable us to train better models with accuracies beyond our current best result of 96.6%.

Final Discussion

In this description, we present a new classification problem: to distinguish effective from ineffective drug compounds through automatic analysis of vascularization images.

This problem may appear to be simple in some cases, such as in Figure 1, and solvable by merely counting the number of bright pixels in the pre- and post-treatment images. However, we find that a linear model obtains an overall test accuracy of 73.3% only, providing only a relatively small improvement over guessing the majority class (62.3%). The difficulty appears to be driven by the nuances of the classification problem, which cannot be captured in a simple linear decision boundary in pixel space. For example, the death of a bridge-to-nowhere vessel should be treated as less important than the death of a vessel on a major thoroughfare in the vasculature network.

To further highlight its difficulty, even an ensemble of four trained human raters had some difficulty with this task (three-way accuracy: 85.9%).

Convolutional neural networks significantly outperform the baseline models as well as human raters on this dataset. Where a cadre of four undergraduate raters achieved a three-way accuracy of 85.9% on this dataset, a convolutional ensemble based on the InceptionV3 architecture and pre-trained on ImageNet data achieved three-way accuracy of 87.0% (though it committed more false negatives than the human raters). A custom convolutional architecture, however, achieves a 95.7% three-way accuracy for drug-hit classification, while committing no false negatives. This pattern repeats itself if we reduce our 3-way classification problem to a binary problem by aliasing together the soft-hit and hard-hit categories.

The success of this convolutional model is driven in part by carefully tuning our loss function to discourage false negatives (see section II-D), but also by the steps taken to control overfitting in the model. One regularization strategy was

to augment our limited training dataset to virtually infinite size via randomly transforming images during each training pass (see section II-E1). Heavy use of dropout also contributed to the result. In fact, the hyperparameter optimization scheme that we used automatically picked a model with a large final layer (512 neurons) and a high dropout probability (0.90). Dropout can be interpreted as implicitly performing a geometric average over an ensemble of regularized subnetworks, so this model can be interpreted as implicitly averaging over a large ensemble of diverse sub-networks.

These regularization strategies were important, as our final network contained 2,485,827 learned parameters and 15 optimized hyperparameters, more than enough capacity to memorize the identity of 208 training + validation datapoints.

However, our network still exhibits excellent generalization power, with test accuracy of 95.7% only barely lagging behind the hyperparameter optimized 96.6% validation accuracy which in turn closely follows the training accuracy of 98.1%. This tendency toward strong generalization performance is often seen in deep networks, and cannot yet be fully explained by any known regularization mechanism or learning theory

In this description, we have developed a convolutional neural network to improve the data analysis processes for high-throughput drug screening using our MPS. This network can classify new images near instantaneously and surpasses human accuracy on this task. A larger scale drug screening can be achieved by coupling this classifier and an automated microscope camera system to capture images before and after drug treatment.

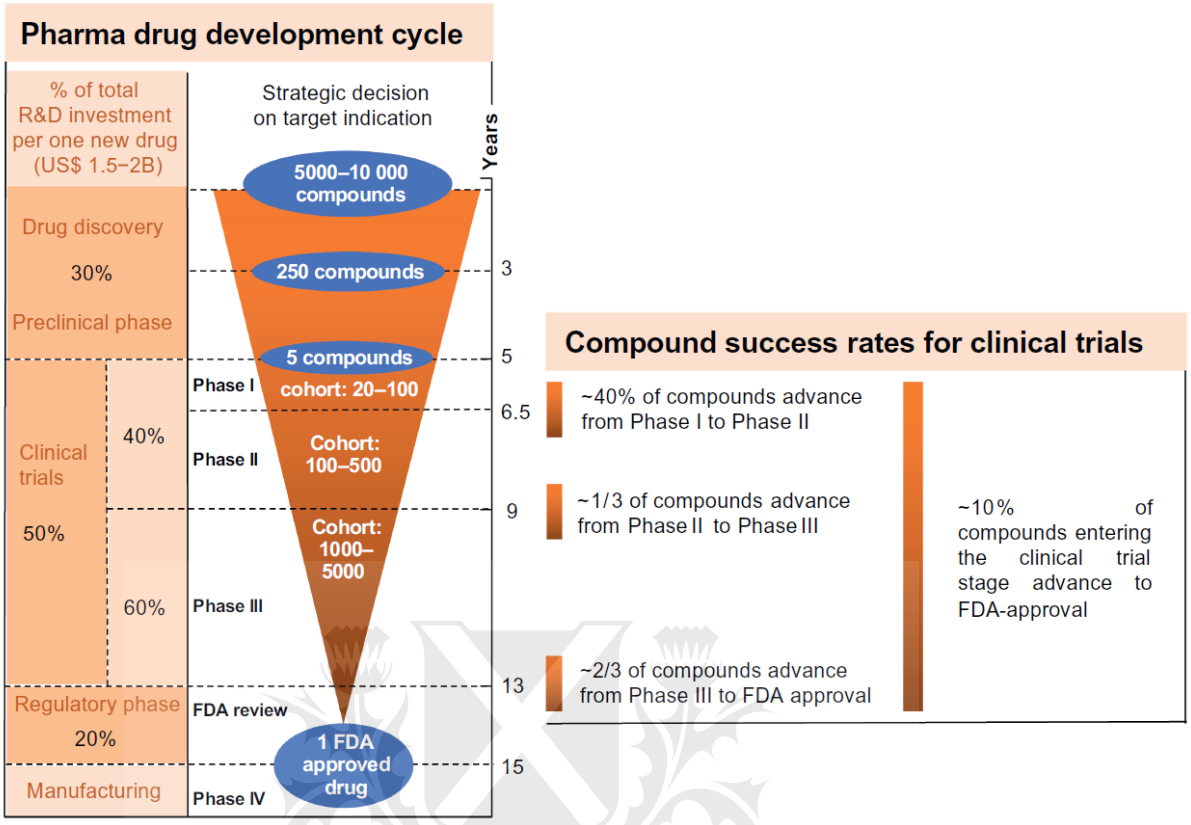
(Stefan Harrer, 2019) (15)

Artificial intelligence for clinical trial design

Clinical trials consume the latter half of the 10 to 15 years, 1.5–2.0 billion USD, development cycle for bringing a single new drug to market. Hence, a failed trial sinks not only the investment into the trial itself but also the preclinical

development costs, rendering the loss per failed clinical trial at 800 million to 1.4 billion USD. Suboptimal patient cohort selection and recruiting techniques, paired with the inability to monitor patients effectively during trials, are two of the main causes for high trial failure rates: only one of 10 compounds entering a clinical trial reaches the market. We explain how recent advances in artificial intelligence (AI) can be used to reshape key steps of clinical trial design towards increasing trial success rates.

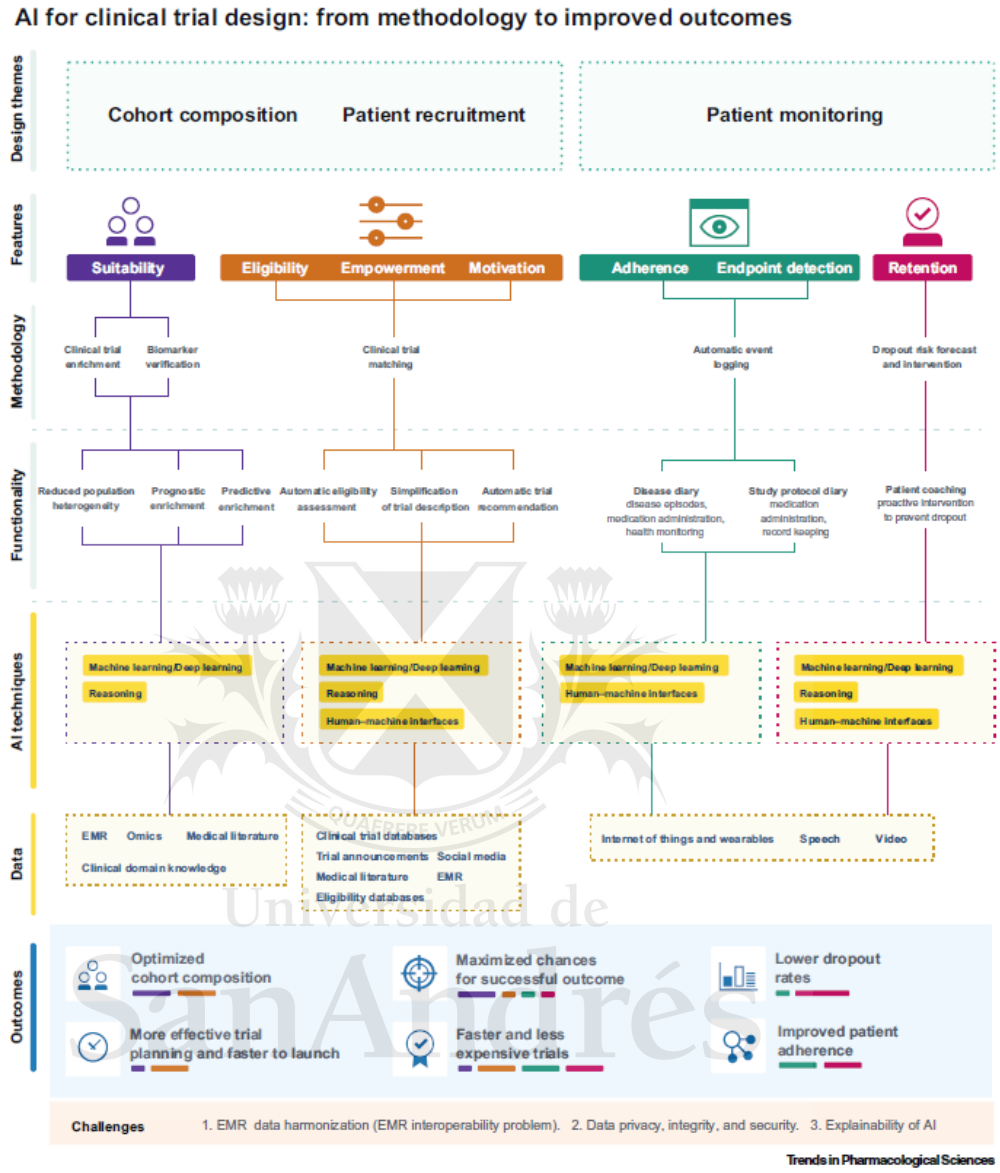
It takes on average 10–15 years and USD 1.5–2.0 billion to bring a new drug to market. Approximately half of this time and investment is consumed during the clinical trial phases of the drug development cycle. The remaining 50% of R&D expenditure covers preclinical compound discovery and testing, as well as regulatory processes. Although pharma and biotechnology companies have continuously increased R&D investment for decades, the number of new drugs gaining regulatory approval per billion USD spent has halved approximately every 9 years. Reversing Moore's law (*in 1965, Gordon Moore postulated that the power of computing would increase while its relative cost would decrease at an exponential pace. This trend held for decades and became known as 'Moore's Law*) from the world of semiconductor technology, this trend has been termed Eroom's Law. It is ongoing and poses a severe threat to the existing clinical development business model: in the post-blockbuster drugs (a drug that creates in excess of $1B in annual sales) era a lack of go to market efficiency of that magnitude is not sustainable. One of the main stumbling blocks in the drug development pipeline is the high failure rate of clinical trials. Less than one third of all Phase II compounds advance to Phase III. More than one third of all Phase III compounds fail to advance to approval. Because these crucial checkpoints do not occur until far into the second half of the R&D cycle with the most complex Phase III trials carrying ~60% of the overall trial costs, the resulting loss per failed clinical trial lies in the order of 0.8–1.4 billion USD, thus constituting a significant write-off of the total R&D investment.



Two of the key factors causing a clinical trial to be unsuccessful are patient cohort selection and recruiting mechanisms which fail to bring the best suited patients to a trial in time, as well as a lack of technical infrastructure to cope with the complexity of running a trial - especially in its later phases – in the absence of reliable and efficient adherence control, patient monitoring, and clinical endpoint detection systems. AI can help to overcome these shortcomings of current clinical trial design. Machine learning (ML), and deep learning (DL) in particular, are able to automatically find patterns of meaning in large datasets such as text, speech, or images. Natural language processing (NLP) can understand and correlate content in written or spoken language, and human–machine interfaces (HMIs) allow natural exchange of information between computers and humans. These capabilities can be used for correlating large and diverse datasets such as electronic health records (EHRs), medical literature, and trial databases for improved patient–trial matching and recruitment before a trial starts, as well as for monitoring patients automatically and continuously during the trial, thereby allowing improved adherence control and yielding more reliable and efficient endpoint assessment.

Patient Selection

Every clinical trial poses individual requirements on participating patients with regards to eligibility, suitability, motivation, and empowerment to enroll. The medical history of a specific patient might render them ineligible. An eligible patient might not be at the stage of the disease, or belong to a specific sub-phenotype, that is targeted by the drug to be tested, thus making that patient unsuitable. Eligible and suitable patients might not be properly incentivized to participate, and, even if they are, they might not be aware of a matching trial or find the recruitment process too complex and cumbersome to navigate. Moving enough patients through these bottlenecks under tight recruitment timelines constitutes a major challenge and is in fact the number one cause for trial delays: 86% of all trials do not meet enrolment timelines, and close to one third of all Phase III trials fail owing to enrolment problems. Patient recruitment takes up one third of the overall trial duration. For example, Phase III trials carry 60% of the total costs for moving a drug through all trial phases because they require the largest patient cohorts. A 32% failure rate because of patient recruitment problems in Phase III trials illustrates one of the most severe shortcomings of state-of-the-art clinical trial design: those trials with the highest patient demand suffer most from inefficient patient recruitment techniques. AI- and ML-driven systems can help to improve patient cohort composition and provide assistance with patient recruitment.



Cohort Composition

Clinical trials are usually not designed to demonstrate the effectiveness of a treatment in a random sample of the general population, but instead aim to prospectively select a subset of the population in which the effect of the drug, if there is one, can more readily be demonstrated, a strategy referred to as ‘clinical trial enrichment’. If a patient is a priori not part of the suitable subset, then their participation in the trial will automatically decrease the observed efficacy of the drug being tested. Suitability may not be confused with the

degree of treatment success or absence thereof during the trial: it denotes a condition that does not render it outright impossible or highly unlikely for participating patients to respond to the tested drug. Recruiting a high number of suitable patients does not guarantee success of a trial, but enrolling unsuitable patients increases the likelihood of its failure.

In an ideal world the assessment of suitability would use patient-specific diagnostic genome-to-exposome profiling to determine whether biomarkers which the drug targets are sufficiently strongly represented in the patient profile or not. Although trials which could benefit from such an approach form a relatively small subset of all trials, they also tend to be the most expensive trials – especially when medical imaging techniques are used. Hence, although in practice there may not be a comprehensive 'omic profile', and effective biomarkers may need to be identified for the majority of therapies under clinical development, biomarker testing should still be considered whenever applicable. Sophisticated analytics methods are necessary to combine omic data with electronic medical record (EMR) and other patient data, scattered among different locations, owners, and formats – from handwritten paper copies to digital medical imagery – to surface biomarkers that lead to endpoints that can be more efficiently measured, and thereby identify and characterize appropriate patient subpopulations. This presents a unique opportunity for NLP and computer vision algorithms such as optical character recognition (OCR) to automate the reading and compiling of this evidence. Moreover, treating data from different sources and formats as a single coherent dataset for the purpose of its comprehensive analysis is especially challenging in the case of EMR data owing to their volume, velocity, veracity, and variety. The data source-agnostic nature of AI models makes them a unique tool for EMR data harmonization which is key to designing tools for clinical trial enrichment and biomarker discovery. However, care must be taken to reduce overfitting of ML models as a result of class-imbalance in the training data.

Preclinical compound discovery, compound-target testing, and defining lead compounds for clinical trials can be assisted by using generative and prediction-based AI, ML, and reasoning techniques. For example, a broader

and more efficient search for correlations between indications and biomarkers than conventional discovery techniques have been reported. This may allow lead candidates to be chosen that have a higher chance of success during clinical trials, and the elimination of those with a higher likelihood of failing before they enter the clinical phase.

AI models and methods can also be used to enhance patient cohort selection through one or more of the following means identified by the Food and Drug administration (FDA):

(i) By reducing population heterogeneity.

(ii) By choosing patients who are more likely to have a measurable clinical endpoint, also called 'prognostic enrichment'.

(iii) By identifying a population more capable of responding to a treatment, also termed 'predictive enrichment'.

Electronic phenotyping is a well-established discipline within health informatics that focuses on reducing population heterogeneity, namely the process of identifying patients with specific characteristics of interest. The characteristics can be as simple as patients with type 2 diabetes, or as complex as patients with stage II prostate cancer and urinary urgency without evidence of urinary tract infection. The task of electronic phenotyping is far more challenging than a simple code search, and requires sophisticated methods to account for heterogeneity among patient records, across multiple data types, and to leverage complex representations of clinical domain knowledge. Although early methods relying on hand-crafted rules were effective for simple cases, they proved to be insufficient for more complex and more nuanced cases. In recent years there have been increasing efforts to design a diverse range of ML methods, ranging from NLP to association rule mining to DL, that have shown great progress towards being able to handle complex real-world situations.

Although electronic phenotyping can be leveraged to reduce patient population heterogeneity, it is not designed to achieve prognostic or predictive enrichment. ML methods are increasingly being deployed for prognostic enrichment for

neurological diseases where key biomarkers, which are typically expensive or invasive to measure, are approximated by non-linear combinations of multiple cheap and non-invasive measures which provide similar prognostic information. Predictive enrichment requires more complex models that are necessary to characterize and assess disease progression. The Coalition Against Major Diseases (CAMD) recently led a process that successfully advanced a clinical trial simulation (CTS) tool for Alzheimer's disease (AD) through the formal regulatory review process at the FDA and the European Medicines Agency. The CTS tool includes computational components for modeling drug, disease, and progression of mild cognitive impairment (MCI) and early AD that can be used for model-based clinical trial design. Expanding on this effort, ML methods for disease progression modeling are being developed to provide increasingly accurate and nuanced understanding and characterization of complexity and heterogeneity of many diseases, particularly those such as AD where disease-modifying drugs are not yet available.

Assistance in Recruitment

The complexity of trial eligibility criteria in terms of number and medical jargon generally makes it challenging for a patient to comprehend and assess their own eligibility. Manually extracting meaningful information from this large and unstructured data source is a significant task that imposes a heavy processing burden on doctors and patients alike. Nonetheless, it is this step that largely defines whether a patient is deemed suitable and eligible to participate in a study, and also whether the recruiting site and the patient become aware of each other. Several AI techniques can offer viable assistance with automatically finding the needles in the EMR haystack: NLP can be used to comprehend written and spoken language from a variety of structured and unstructured data types. Reasoning techniques allow content to be digested into actionable recommendations for the human decisionmaker. ML and in particular deep reinforcement learning empowers systems to learn and integrate feedback on the quality of their analytic output into adapted underlying algorithms. Assistive systems using these AI techniques or subsets thereof can be used to automatically analyze EMR and clinical trial eligibility databases, find matches

between specific patients and recruiting trials, and recommend these matches to doctors and patients. Such AI-based clinical trial matching systems have successfully been demonstrated and have proved their value in real life use cases. Because of the AI nature of these systems, any added future functionality and improved performance predominantly will depend on the quality and amount of data which are accessible for analytical model development and pilot study field validation work.

AI and ML techniques such as NLP and OCR have also been proposed to proactively mine publicly available web content such as, for example, digital trial databases, trial announcements, and social media to automatically identify potential matches between trials of relevance and specific patients. By assisting patients in their conventional manual web search, such a system could make patients aware of trials of interest much faster and allow them to proactively engage with clinicians for further assessment of eligibility and suitability. Indeed, the first enrolment plans employing a social media component have successfully been demonstrated. We expect the integration of AI will improve the reach, efficiency, and thus the impact of such digital enrolment plans substantially in the future.

Challenges

The digitalization and accessibility of EMR data that are used extensively by AI methods are not trivial. Both tasks are challenging for contrary reasons: on the one hand a lack of regulatory frameworks on data collection causes EMR formats to differ widely, to be incompatible with each other or not digital at all, and to reside in a decentralized ecosystem without established data exchange or access gateways. On the other hand, a strongly regulated legal environment strictly limits third party access to patient data and even makes it difficult for patients themselves to access their own data. This so called ‘EMR interoperability dilemma’ is being recognized as major hurdle to making healthcare systems more efficient, and substantial investments are being made by governments and medical institutions towards overcoming this hurdle. In parallel, legal frameworks such as, for example, the US Health Insurance Portability and Accountability Act (HIPAA) and the EU General Data Protection

Regulation (GDPR) continue to evolve as governing and protecting sensitive health data becomes an increasingly complex endeavor in the growing network of devices, data owners, and service providers [24,25]. Further, exactly as with EMR mining, for clinical trial matching the legal aspects of data privacy and security as well as a sufficient degree of explainability of AI models need to be addressed to ensure that AI-based systems are operable and gain regulatory approval.

Patient Monitoring

Recruiting the right patients into a clinical trial is a massive investment of both time and funding. The return on this investment can only be realized through successful completion of the trial. Hence, it is imperative that patients stay in the trial, adhere to trial procedures and rules throughout the trial, and that all data-points for monitoring the impact of the tested drug are collected efficiently and reliably. Only 15% of clinical trials do not experience patient dropout, and the average dropout rate across clinical trials is 30%. Dropouts caused by a lack of adherence to trial protocols require additional recruiting efforts, which lead to trial delays and substantial additional costs. A linear increase of the non-adherence rate in a trial leads to an exponential increase in additional patients required to maintain the statistical power of the outcomes. For example, a study in which half of the patients are non-adherent means an additional 200% of patients need to be recruited to keep the statistical power of the results stable. Improved patient monitoring and coaching methods during ongoing trials can be used to lower the adherence burden, make endpoint detection more efficient, and thus reduce dropout and non-adherence rates. AI techniques in combination with wearable technology offer new approaches to developing such power efficient, mobile, real-time, and personalized patient monitoring systems.

Patient Adherence Control, Endpoint Detection, and Retention

To comply with adherence criteria, patients are required to keep detailed records of their medication intake and of a variety of other data-points related to their bodily functions, response to medication, and daily protocols. This can be an overwhelming and cumbersome task, leading to on average 40% of patients

becoming non-adherent after 150 days into a clinical trial. Wearable sensors and video monitoring can be used to automatically and continuously collect patient data, thereby relieving the patient of this task. ML and particularly DL models can then be used to analyze such data in real-time for detecting and logging events of relevance. This approach allows disease diaries to be generated which – because the underlying analytical DL models are periodically retrained with updated measurement data – evolve to be patient specific and adaptive to any changes in disease expression and patient behavior. Such disease diaries may serve as evidence for adherence or lack thereof and – as minimal or no manual patient input is required – will also collect data-points for endpoint detection more reliably and efficiently than current patient-driven self-monitoring methods. AI also has an important role to play in image-based endpoint detection – a task that is currently addressed manually at reading centers. ML technologies have been proposed – and recently approved – for screening applications for the rapid detection of diseases from medical images. Complementing this with algorithms that quantify pathological conditions will reduce the cost associated with image-based studies by circumventing manual processing.

AI and ML methods may also be used to dynamically predict the risk of dropout for a specific patient, in other words to detect the onset of patient behavior that suggests the patient might be experiencing issues with adhering to the study protocol. One such example described the use of deep reinforcement learning algorithms to determine the fewest, smallest doses that could still shrink brain tumors, while reducing toxicity associated with chemotherapy dosing regimens. Powered by a 'self-learning' ML technique, the system looks at treatment regimens currently in use, and iteratively adjusts the doses. Eventually, it finds an optimal treatment plan, with the lowest possible potency and frequency of doses that should still reduce tumor sizes to a degree comparable to that of traditional regimens. In simulated trials of 50 patients, the ML model designed treatment cycles that reduced the potency to a quarter or half of nearly all the doses while maintaining the same tumor-shrinking potential, and thus promises improvements in patient adherence and reductions in dropouts and censuring.

Picking up early warning signs for non-adherence allows proactive engagement with individual patients and permits the root causes of problematic behavior to be addressed: for example, severe side effects or incompatibility of study and personal routines could be detected and remedied before they lead to dropout. The choice of sensors and analytical models is highly disease-specific and will need to be part of the clinical study design.

Using DL for object recognition in images and video, as well as for analyzing time-series data from wearable sensors, first studies for testing and exploring AI-assisted patient monitoring systems have recently been started or completed successfully. The advent of commercially available wearable devices with medical-grade health-sensing capabilities, as well as complementary software ecosystems for running advanced DL models on such mobile platforms, will allow more diversified sensor combinations to be investigated for a variety of diseases. In a previous study, Shah et al. evaluated the significance and efficacy of clinical evidence generated from advanced technology-enabled non-invasive diagnostic screening (TES) using low-cost smartphones and other point-of-care medical sensors versus conventional vital signs examination. They report that, although routine health screening continues to be important, the emerging techniques of TES can play an important synergistic role in stratifying populations and providing personalized screening and care in support of clinical trial designs and observational studies to generate innovative, new treatment approaches. We expect to see more pilot studies benchmarking the impact of such technologies on trial efficiency alongside ongoing clinical trials in the near future.

Recent advances in custom-developing mobile processors and coding environments allow DL models to be run close to or at the point of sensing. This transforms wearables from pure information storage and transmission devices into information digestion and analytics devices – a novel concept which we call 'cognitive sensing'. Wearables measure biometric parameters through mobile systems attached to the human body, and either store the collected data on the device or send it to the cloud for offline analysis. As the wearable revolution unfolds, a rapidly increasing number of parameter types can be monitored

simultaneously, making storage and transmission of unfiltered sensor data impossible. Algorithms for analyzing, in other words continuously correlating, contextualizing, and filtering raw data in real-time directly at the point of sensing, will be necessary to extract actionable information before the need for data storage or transmission arises. DL models in combination with on-sensor data preprocessing and curation systems allow this task to be accomplished. The architecture of such wearable, autonomously operating, always-on, cognitive sensors consist of the following system components: (i) minimum-footprint biosensors feeding into (ii) low-power mobile processors capable of locally running DL models with (iii) closed-loop interfaces to (iv) an event diary which instantly and proactively logs information on specific disease episodes and interacts with wearer or caregiver for patient support, guidance, and intervention. The event diary can thus utilize a local memory unit, a remote cloud repository, or a hybrid version of both. Various wearable biosensor and actuator platforms in different stages of technical maturity have been demonstrated or are currently under development.

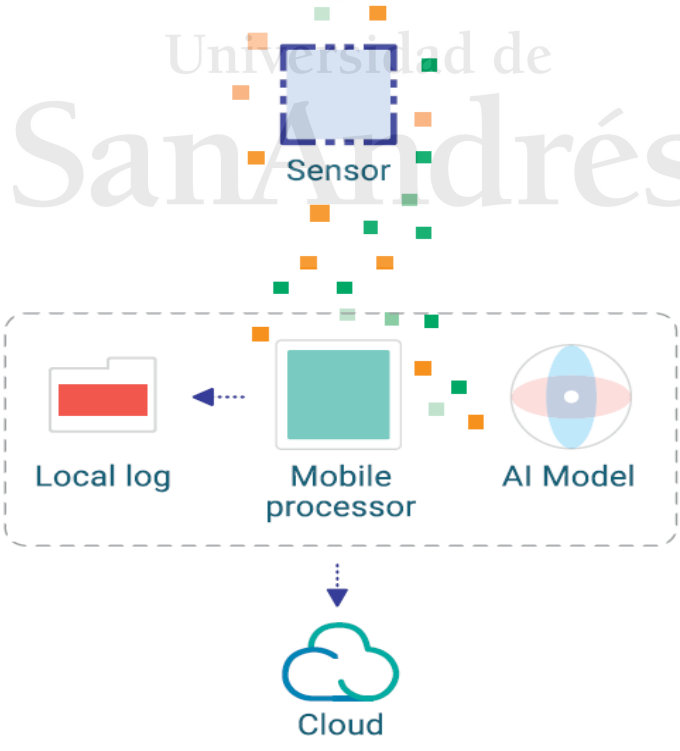


Figure 3. Cognitive Sensors. Data are measured by a wearable sensor and analyzed in real-time at the point of sensing by a mobile processor that runs an artificial intelligence (AI) model. Analysis results are then stored on a local log, in the cloud, or through a combination of both.

The predominant type of data which most of these sensor types measure is time-series data. Although DL has traditionally focused on analyzing imagery data using deep convolutional neural networks, recent work has demonstrated that custom-designed neural network models are also uniquely suitable to analyze complex time-series streams. To run DL algorithms continuously in real-time at the point of sensing, ultra-low-power consumption mobile processors are needed. Advances in developing both novel AI hardware and AI software techniques over the past 3 years have led to several versions of such AI-tailored mobile processing solutions now being available for real-life use. These solutions can be categorized into three general types:

(i) custom-developed hardware requiring custom developed AI coding environments, such as IBM's TrueNorth chip, (ii) custom-developed hardware compatible with standard AI programming tools, such as Qualcomm's Snapdragon chip series and Intel's Movidius processor, and (iii) conventional mobile processors which can be programmed using standard AI coding platforms, such as the Apple Watch, Apple's XS iPhone series carrying the A12 Bionic chip, and also a variety of other smartphones.

As pointed out previously, interoperability and standardization of data and methodology are key challenges for integration of AI into clinical trial design. The same is true for wearable AI technology and devices. Regulatory bodies, in collaboration with academic, medical, and pharma institutions, have started to produce standardization frameworks and best practice recommendations for incorporating wearable technology into clinical trial design.

Ongoing research at the intersection of AI, Internet of things (IoT), and healthcare will produce more medical-grade devices with advanced analytics capabilities for continuous real-time monitoring of patients and disease progression. If an equally strong focus on standardization and interoperability is maintained, these devices might make cognitive sensing an effective tool for improving the performance of trials. It is important to note, however, that data integrity and safety occupy a central role in the conception, implementation, and exploitation of digital disease diaries: patients, doctors, and regulatory bodies

will rely on the integrity and safety of sensitive patient data and of analytical insights derived from it. While HIPAA-compliant environments constitute the data security baseline, advanced generations of AI-based monitoring and data-housing platforms will employ blockchain technology for ensuring trusted and traceable multiparty communication and exchange of monitoring data.

Over the past 5 years modern AI techniques have advanced to a level of maturity that allows them to be employed under real-life conditions to assist human decision-makers in computer vision, navigation, and in some cases of medical and healthcare environments. At the same time, pharma and healthcare are still among the most highly regulated and risk-averse industries. Infusing innovation that changes established processes is a difficult task that needs to be approached and implemented in a stepwise manner. Although AI has the potential to impact numerous steps of clinical trial design from preparation to execution, any AI pitch that aims to tackle all aspects at once is predestined for failure. Instead, data scientists and medical scientists should jointly define achievable use cases where the application of well-understood AI tools to a specific subtask of clinical trial design promises the greatest improvement of overall trial performance. Such AI technology first needs to be tested alongside the existing technology it aims to complement or replace, and the added value must be demonstrated and benchmarked in an explainable, ethical, repeatable, and scalable way – not only to users but also to regulatory bodies. Following this approach AI may be adopted into the clinical trial ecosystem step-by-step, making trials faster, while at the same time hopefully lowering failure rates and R&D costs. Several pharma and AI companies have started to jointly explore this avenue. Regulators have put in place and continue to expand frameworks for assessing AI-based technologies in healthcare.

Further, completed trials have amassed a corpus of data which carries a wealth of information on correlations between trial design features and trial performance. This includes data from failed clinical trials. These large and unstructured datasets are predestined to be analyzed by AI technologies. Insights could be used to educate future improved trial designs and also to investigate the potential relevance of already trialed drugs against comorbidities

for drug repurposing. Nevertheless, failed trial data in particular tend to be a neglected asset that has remained largely untouched on the shelves.

It is important to note that the measurable impact of any such steps on the efficiency of the pharma R&D pipeline – even if implemented successfully now – will not show up in the statistics until after a 5–8-year delay. Moreover, there will be additional R&D costs on top of the ongoing costs; in other words, from a required investment perspective, things will get worse before they will get better.

The AI techniques described in this review offer real-life practicability; however, particularly with respect to explainability, these techniques must mature to allow their broader inclusion in healthcare and life sciences applications. Although these developments are in full swing, we need to acknowledge that the opportunity to transform the drug development cycle through AI comes with a great responsibility for all the disciplines involved and the mandate to qualify the value and reliability of any innovation through rigorous R&D work. This exploratory research pilot phase may not be bypassed for any reason because any breach of research protocol or premature setting of unreasonable expectations will inevitably undermine trust and ultimately the success of AI in the clinical sector.

In the same way as a change of clinical trial design alone will not turn efficiency of the pharma R&D cycle from decay to growth, AI is not a magic bullet that will make the success rates of clinical trials skyrocket overnight. Both reshaping clinical trial design and using AI techniques for doing so are important building blocks of a much-needed overhaul of the drug development cycle.

(Stefan Harrer, 2019) (16)

Chapter VI: Proposal of a global platform based on AI – DL for cancer early diagnosis

Justification and business aspects

Platform Objective:

At the end of chapter II, I have presented a synthesis of key concepts for the development of AI as an early detection mechanism and therefore a high probability of cure of certain types of cancer, one of them referring specifically to deep learning technology mentions: **"**Deep Learning is powered by massive amounts of data. Deep learning models tend to increase their accuracy with the increasing amount of training data", that is why the approach of this chapter is how to articulate a global AI platform that can be fed with images and diagnostic data from all over the world to achieve diagnostic precision that can be a source of consultation for health professionals in early stages treatment of the patient, thus increasing the probability of its cure.

The proposed Platform aims to unite producers of medical diagnostic images, diagnostic lab results and medical literature with health professionals who wish to validate the tests performed on a patient in order to detect possible forms of cancer early and act accordingly to prevent cancer disease development.

Enablers

Which are the factors that allow us to think that today it is possible to implement a platform with these characteristics?

1) Massive internet access, although with differences in speed and quality, is a worldwide reality.
2) Access to technology in the cloud in the form of software as a service (SaaS) and platform as a service (PasS) today has a level of maturity that allows the implementation of projects of all kinds regardless of

variables such as the volume of data, the location of the producers and consumers of information and the type of technology necessary for its realization. Leading cloud computing providers like Amazon and Microsoft have state-of-the-art, end-to-end AI solutions for hire as a service.

3) Diagnostic imaging devices such as computed tomography, magnetic resonance imaging and ultrasound are common in most health centers and diagnostic centers and today all these devices generate digital images by themselves or through a personal computer. These digital images can be uploaded to an internet platform only with a connection in the health / diagnosis center, something totally common today.
4) As mentioned in chapter II of this work, deep learning technology has a level of maturity that makes it possible to use it with a population of trained resources large enough at a global level to design, implement and maintain a learning solution. these characteristics.

Therefore, we can say that the technological conditions, the possibility of generating the necessary information so that the precision of the result delivered by the application is what is sought, and the capacity of health centers to make use of these results online, are given. So that a project of these characteristics can be viable today.

Target Market

The platform's target market is the world. Being that cancer is a disease that affects humanity throughout the planet and that the elements that we have mentioned as part of our solution proposal are also global (being aware of the differences that exist in these issues between developed countries and that are developing) we propose that:

1) The platform must have as its objectives to obtain data in as many as possible and from as many countries as possible.

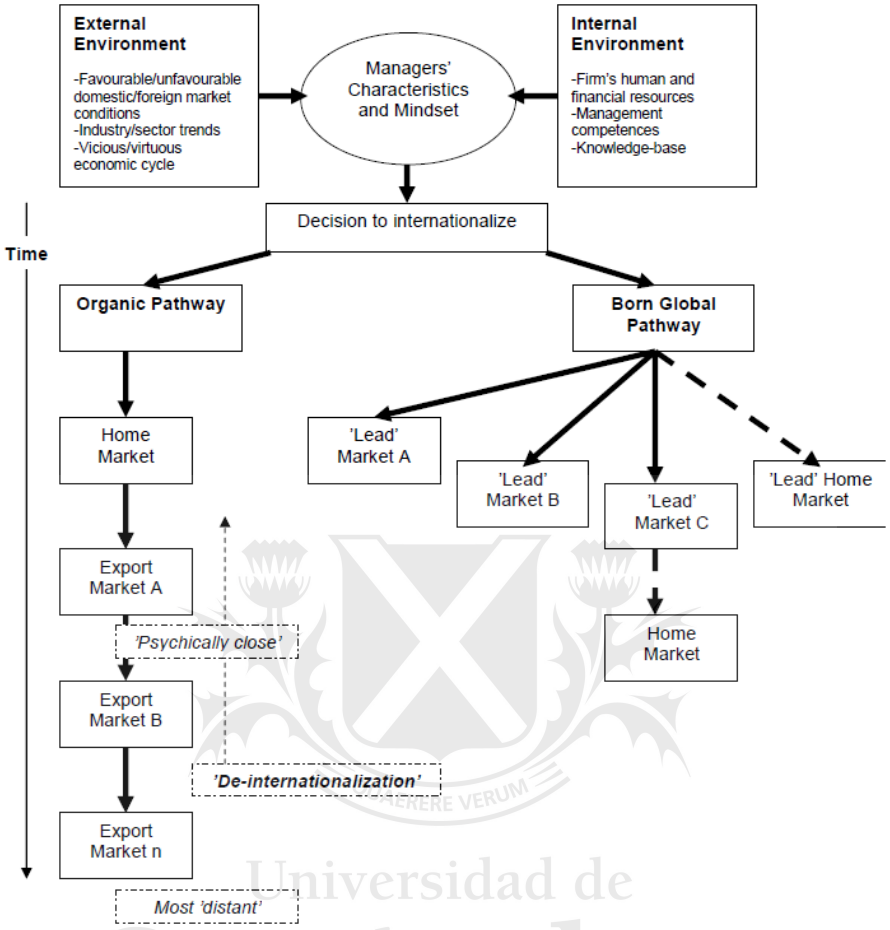
2) Provide your diagnostic validation service to the greatest number of users in the greatest number of countries possible.

Google states that its mission is “Our mission is to organize the world’s information and make it universally accessible and useful.” Without wanting to compete with Google, we state that the mission of our platform is **"Collect diagnostic information and medical literature on cancer, organize it and use it to provide a global service for early validation of patient data to achieve an early diagnosis that increase the possibility of the patient's cure”.**

Platform internalization pathway

According to the book “Internationalization handbook for the software business” (Toivo Äijö, 2005) (17) which explains the paths that a company can take to start the internationalization stage, tells us: “As we proceed further in the analysis of the various choices and context specific nature of internationalization, it can be noted that the real-life multitude of paths to growth and internationalization can be grouped into three distinct and more or less typical pathways. They are **organic, collaborative and born global pathways.** In many ways the slow organic and accelerated born global pathways are the opposites of one another, at the two extreme ends of a

spectrum”



Likewise, the book says in its chapter 2 “Choice of the pathway and internationalization Strategy” “The choice of pathway must be understood within the context os strategic planning”. Therefore, based on the objectives and scope of the platform that we have described, it is proposed that the model to be followed be “born global” given that it is the model in which our proposal can achieve the objectives it sets out in a time frame that makes sense of it.

Graphically, it is seen in the following image:

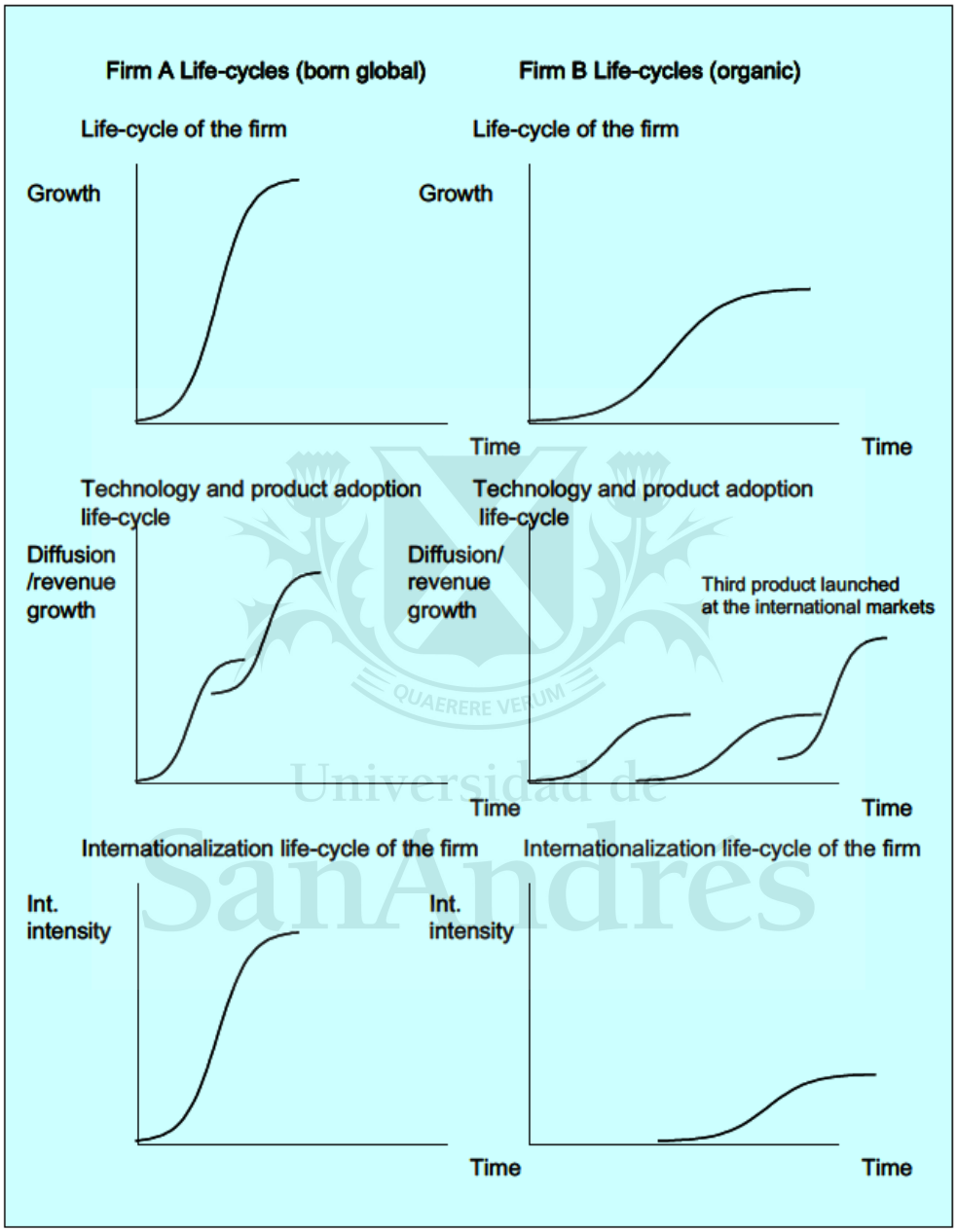


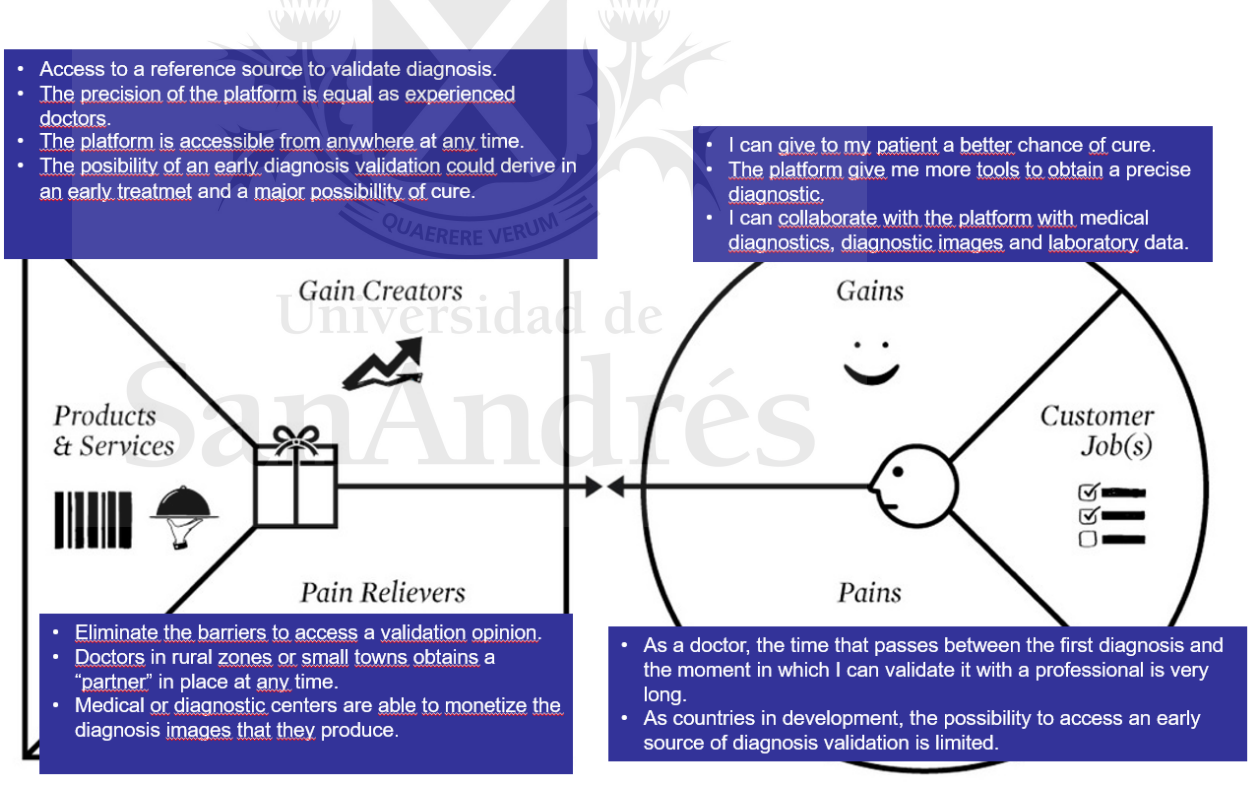
Figure 7: An example of the life cycle differences between born global/accelerated growth and organic pathways

Value Proposition

Summarizing what has already been stated, the platform's value proposition consists of "to unite producers of medical diagnostic images, diagnostic labs

results and medical literature with health professionals who wish to validate the tests performed on a patient in order to detect possible forms of cancer early and act accordingly to prevent cancer disease development". Allowing:

- Quickly validate the initial diagnosis of the patient.
- Allow early treatment of the disease, increasing the chance of a cure for the patient.
- Reduce the development cycle of new drugs for the treatment of different forms of cancer.
- Monetize diagnostic images, laboratory data, and medical literature generated by medical centers, laboratories, and medical research centers.



Marketing Plan

To meet the objectives set out in the platform's value proposition, it is necessary to set out **the strategic objectives** in the marketing plan, of which we have already presented the target market and the value proposition. The

third strategic objective is "data acquisition", being the diagnostic images, laboratory results and medical literature the raw material of our service, the acquisition of the same from its producers and retention of the same is a strategic objective on which concrete and permanent actions must be focused on the part of the business.

In addition to the strategic objectives, the marketing plan has the following pillars:

Communication and media plan: Given the global reach of the platform and the sensitivity of the topic it addresses; the communication plan must be clear and focused on the target audiences. It should contain definitions about:

- **Channels:** Through which media will we advertise the service. Clearly, any publication or scientific dissemination medium in the area of oncology in particular, but of medicine in general, is a candidate to be a channel for disseminating the service. We must consider digital and traditional scientific popularization publications, popularization editions of the pharmaceutical industry, presence in congresses and events.
- **Content:** The content must clearly reflect the value that the platform brings to the oncologist in the early stage of patient treatment, as well as to the health center and the patient's health coverage, since early treatment that increases the probability of cure decreases costs. of treatments in advanced stages of the disease as well as surgeries, hospitalizations and other high-value specialized medications.
- **Target Audience:** The marketing action must be directed to the two groups that we seek to capture, the first are professionals / health centers and health insurance, the second the data generators for the platform, diagnostic centers, medical centers, scientific research centers related to cancer.

Dissemination and conversion plan: To ensure that the content we seek to disseminate reaches the target audience and that the solution offered is of interest to users, the content dissemination plan must guarantee the greatest

impact of the content on the target. Since our target is a highly qualified audience, the content must have the greatest credibility, it must be segmented according to whether we seek to reach the oncologist, or the diagnostic center that generates the data, in the digital channel we must seek the highest positioning in search engines, a specific strategy for social media and a state-of-the-art website focused on showing the value of the service provided by the platform.

Monitoring and analysis: From the moment the marketing actions are launched, both digitally and in traditional media, metrics must be available to measure the performance of the campaign launched or the website. Some metrics are: measure exposure, interest, link and result of the action. After the campaign, the return on investment, retention and brand value must be measured.

Network effect

The main challenge of any platform is to achieve the network effect in its service, the network effect is defined as "The more people that use the service, the more useful it is because there are more people to use it with". In the case of a platform whose service is based on collecting and providing data, achieving the network effect is even more difficult and is known as "data network effect". As described by Alexandre Gonfalonieri in his article "Why is it hard to build AI & data network effects" (Gonfalonieri, 2020) (18) where defines that "A Data Network Effect is a property of a product that improves with the more data it has available, due to emergent relationships between segments of the data". This definition is aligned with the concept analyzed in previous chapters where we saw that the platform of deep learning it increases its precision to the extent that it has more data to "learn". According to the author, the complexity to achieve the data network effect lies in the following list of problems:

1) Data Culture and expectations: In any organization, the absence of a data-centric culture that prioritizes the effective use of data, no relevant data network effects can exist. Data network effects produce little value

before they reach critical mass, and most newly applied algorithms suffer from a “cold start” before acquiring adequate data. In our case, the data network effect will be the core due to the need of imaging data for our AI system’s training that will produce more clients interested in using the system to check the patient’s laboratory results.

2) Relationship with data & Closed Loop System: Actually, the goal should be to improve your models with additional data sources in addition to your own data. In general, the broader data sources you have, the more accurately you can model your environment and make accurate predictions. In our case, as much information we can collect, more precise will be the system response to doctor’s questions.
3) Improvements vs time: Another interesting element to bear in mind is that after the dataset reaches a certain size, the algorithm no longer meaningfully improves as the dataset grows. The reality is that most data network effects struggle with this. Moreover, the point beyond which the data network effects diminish varies by domain. In our case, this point should be analyzed after that the system arrives to a levels of diagnosis precision similar to medical professionals.
4) Automation & Scope: From an operational perspective, data network effects require at least some level of automated productization of the insights. The more automation you add into the loop, the more efficient the virtuous circle of AI becomes. Furthermore, the narrower the scope of a product and the greater the degree to which machine learning drives its value, the easier it is to tie these advantages together to create a competitive advantage through data network effects. In our case this is the approach to achieve the objective that we are looking for.
5) Types of data: The data gathered from new users must be of the right kind and of sufficient volume to enable data network effects. This learning must be optimized effectively enough to create new product value. And this value must be strong enough and productized well enough to attract more customers. Any break in this chain means there

is no self-reinforcing cycle and hence no learning effects. In or case, as the previous point, this is aligned with the porpoise of our platform.

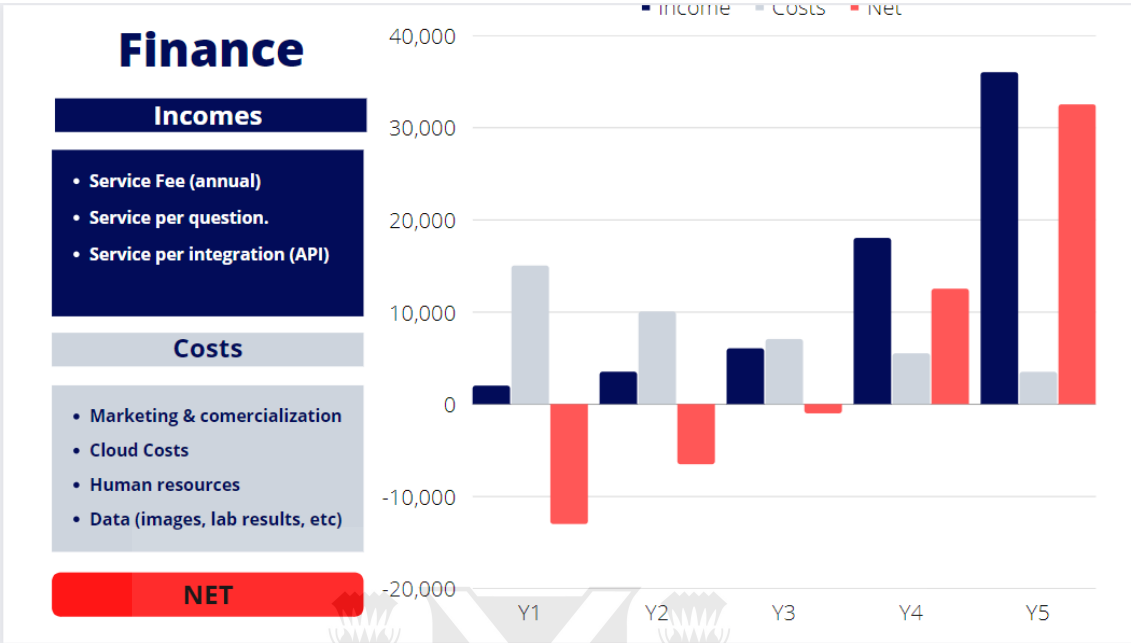
After reviewing these considerations and the characteristics of the platform that we propose to implement the early diagnosis of possible cancer cases through AI, we understand that it meets the objective of achieving the network effect and also the data network effect as result of the design of the service it provides as well as the operational characteristics of its operation.

Finances

The financial flow of the platform requires evaluating the table of income and costs over five years, since an initial period of 2 years is required for the following foundational tasks that will allow us to obtain the result we are looking for from the platform:

- Human resources: Engineers and programmers specialized in AI, C level (CEO, CFO, CTO), marketing and sales staff, staff.
- Cloud infrastructure: Platform as a service, software stack as a service for the development of the AI solution, front ends of data and client loading, back office and digital marketing.
- Advertising: Digital, in specialized media, medical conferences and medical institutions.
- Data acquisition: Hire diagnostic centers and health centers that provide the diagnostic images they produce in exchange for a monetary value for each image received.
- AI's platform developing and training.

The graph of income and initial expenses the first two years with a high cost of starting operations for the concepts mentioned above, these costs are covered by the source of financing that is mentioned in the next section, the income of the first years is low but once the platform is created and trained, delivering the results with the promised precision, the income grows exponentially given the massiveness achieved by network effects.



Funding model

Another point that must be planned with the greatest precision is the financing model of the platform that is being developed. The traditional models of investment rounds and seed capital are a possibility as they are in any digital product development cycle, however, given the characteristics of this platform and its purpose, which seeks to help find a way to lower the numbers of deaths caused by cancer, should be sponsored by international organizations dedicated to health such as the WHO, PAHO or similar.

The reason for this proposal is that the objective of this platform is not primarily commercial, it is not a platform that seeks a return of X times the invested capital, but rather a humanitarian purpose. Funding to launch this platform and keep it in the data acquisition process until the AI engine achieves the accuracy sought in early diagnosis can be provided by one or more global health organizations. Once this point has been passed, at which the platform has "learned" enough from the data to be able to give precise diagnoses, a transition to private capital can begin with the aim of sustaining the necessary growth.

Once the financing has been obtained for the development of the solution and its release to the market to capture data that allows the AI engine to learn up to an acceptable level of precision, investment rounds are opened to maintain and expand the solution in operation.

As an example of funding programs of health associations, we can see the PAHO (Pan American Health Organization) “Health technology assessment”, the PAHO’s web site says:

“Given the growing interest for health technologies, PAHO has launched several initiatives in HTA with member countries to promote and strengthen health technology assessment in the Americas. PAHO's role is important for the development and implementation of HTA in the Americas, and to support the promotion of evidence-based decision-making processes, which contribute to the incorporation of cost-effective technologies.

In 2012 member states adopted the resolution "Health Technology Assessment and incorporation into Health Systems" (CSP28.R9). The resolution proposes linking HTA with the decision-making processes involved in incorporating these technologies into health systems. Since the approval of CSP28.R9, there have been clear advances in the institutionalization of HTA in the Region, both at regional and national levels. Despite the major progress, the implementation of HTA remains at a low level in some countries. PAHO encourages the establishment of an institutional framework for HTA-based decision-making. This framework would establish linkages between HTA and decision-makers, encouraging institutional responsibility and creating linkages between the use of technologies and evaluative data to feed into the decision-making process.

To support the development of HTA in member countries, the Medicines and Health Technologies Unit is involved in many activities, such as:

Human resources development: establishing a regional strategy to assess different regional needs. Promotion of regional meetings, workshops, and training through online courses, and webinars.

Dissemination of information: identifying existing opportunities and disseminating findings amongst stakeholders and decision-makers, through the Regional Platform on Access and Innovation for Health Technologies (PRAIS).

Rational use of health technologies: developing and implementing clinical guidelines to evaluate the use of health technologies in health services.

Promotion of network collaboration: promotion of regional cooperation among member countries, strengthening the HTA Regional Network through the Regional Network of Health Technology Assessments for the Americas (RedETSA). “

(PAHO, s.f.) (19)

Globally growing diagnostic imaging equipment market

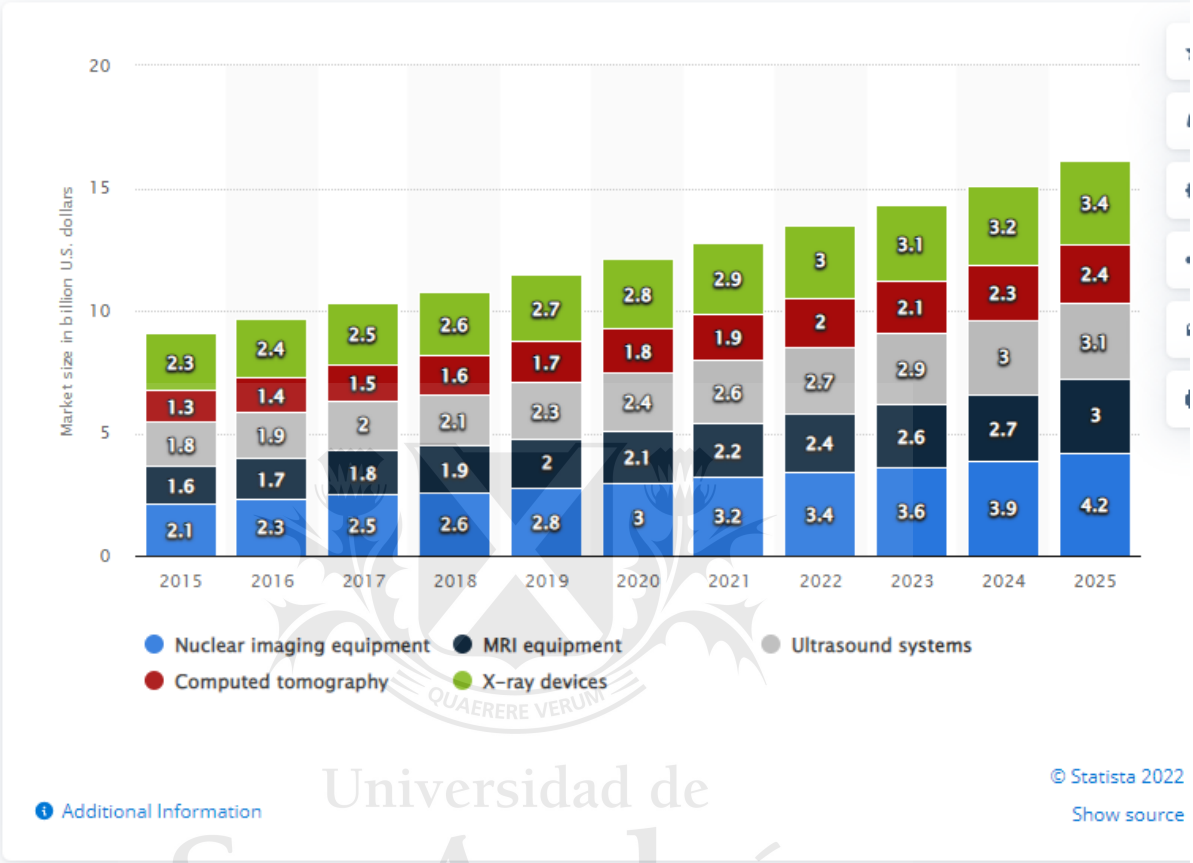
In the case of an AI solution based on data, it is important to deepen the concept mentioned in the previous section on the sustained trend of overcrowding of diagnostic imaging equipment, which manages to reach more diagnostic centers, reduce the cost of their service and therefore increase the amount of images and data available that can be stored on a platform that uses them as an information base.

The following study presents the evolution trend of the diagnostic imaging equipment market in the United States until the year 2025. For the cases of nuclear medicine equipment, magnetic resonance imaging, ultrasound systems, computed tomography and X-ray equipment Taking the installed base of the year 2015, it is seen in all cases the market growth for the year 2025 is close to double.

Health, Pharma & Medtech › Medical Technology

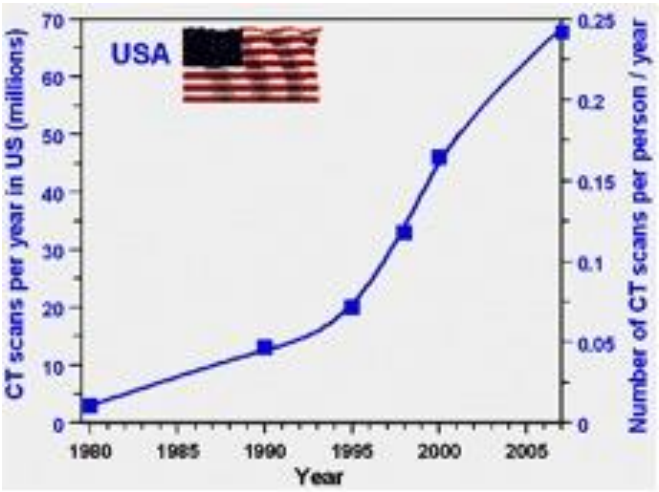
Medical imaging market size in the U.S. from 2015 to 2025

(in billion U.S. dollars)



The statistic shows the size of the U.S. medical imaging market from 2015 forecasted until 2025, by product type. In 2021, the total U.S. market stood at around 13 billion U.S. dollars, of which 3.2 billion dollars were generated by nuclear imaging devices and 2.9 by X-ray devices. (Stewart, 2019) (20)

A second statistic that supports the massification of diagnostic images has to do with the evolution of computed tomography performed in the United States since the first equipment with these characteristics arrived on the market in the early 1980s until the year 2005, the growth curve is completely clear, showing that the adoption of these devices and the results they generate is massive and they are widely used by the medical community today.



(Montagnese, 2012) **(21)**

According to the latest results from "iData Research's medical imaging procedures analysis", over 75 million CT scans are performed each year in the United States. This number is forecasted to grow to reach 84 million procedures by 2022. (Research, 2018) (22), and in the same way more than 95 million MRI (Magnetic resonance imaging) scans per year worldwide.

(Lakrimi, 2018) (23)

Platform's main functionalities

To meet the proposed objective of the platform, we can mention two perfectly defined functionalities that it must implement:

1) Data Gathering

   This functionality of the platform is responsible for obtaining the data, the essential fuel required by the AI engine implemented in the platform to be able to fulfill its mission of accurate early diagnosis.

   The mission of the data gathering module is the "ingestion" of the information provided by medical centers, diagnostic centers, and any other institution related to medicine that has been contracted to provide images and laboratory results such as CT scans, MRIs, X-rays, and lab results. The module is the one who presents the front end through which these medical

institutions send the information to the platform with its corresponding registration, validation and storage in the platform's database.

When the information is correctly stored, the module sends the confirmation to the information provider, executes the accounting registration process of the information received for subsequent payment management to the provider and sends the communication to the AI module about the information received.

2) AI engine core functionality.

This is the Deep Learning algorithm implemented in the platform, the AI engine that, based on the information it receives from the previous module and uses it to "learn" about the data and raise the level of precision of the queries made by users. medical specialists in the next module.

The complete system will achieve acceptance and use by the community to which its service is aimed, to the extent that this module manages to develop with the necessary precision so that the learning process on the ingested data provides answers to early diagnosis queries, allowing rapid treatment that increases the chances of overcoming the disease in patients.

3) Patient information AI check

This module is the front end of the doctor who connects to the platform to validate the results of his patient's clinical studies. The doctor seeks to connect, upload the images and diagnostic data of his patient and get the answer from the AI about the possibility of having detected cancer. In addition to delivering the report to the doctor, the platform validates with the doctor if it is authorized to load these images and results in the application's database to increase the sample of data on which the AI engine learns.

4) Backoffice

The Backoffice module is in charge of all the administrative tasks, for example customer management, finance and accountability, business

administration, any other process needed for the business but not directly related to the platform's core modules.

5) Statistics

Module responsible for all the metric dashboards, reporting for all user levels, statistics regarding the AI performance when it is used by customers, statistics regarding the data used to train the AI engine. Typically, this module uses different business intelligence and advanced analytics engines, ingestion and ETL technologies and custom code for data manipulation and data visualization needs.

Technical Approach

The state of maturity of different fields of technology together, makes it possible for us today to be sure that the platform we seek to create is implementable, and not only that, but we have more than ten possibilities of providers in the cloud that we can evaluate. to select the technological partner on which to develop our solution.

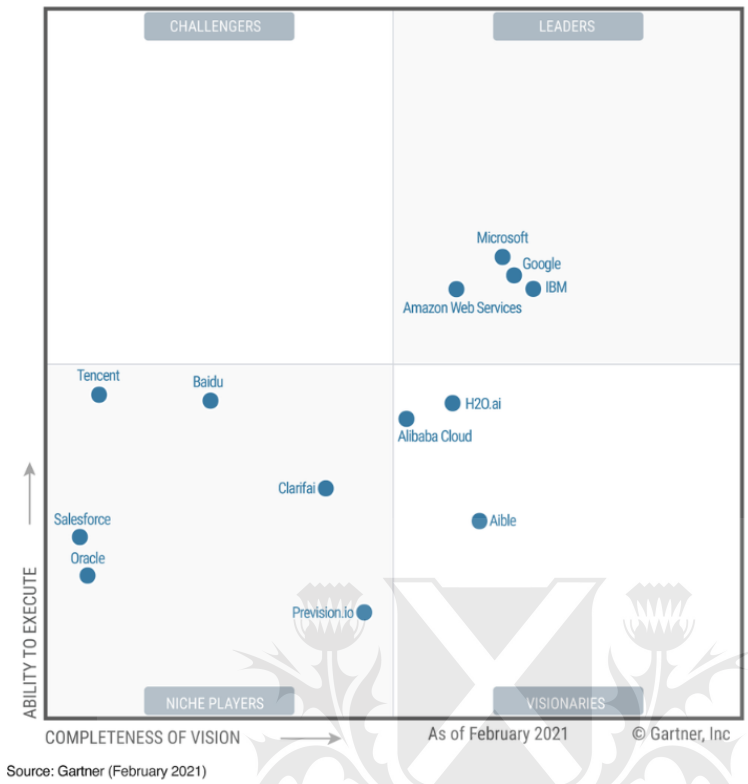
The technological fields to which we refer are:

- Internet connectivity.
- Cloud computing.
- AI software as a service.

Regarding this market, the Gartner Group refers to it as "Cloud AI Developer Services" and defines it as "cloud-hosted or containerized services/models that allow development teams and business users to leverage artificial intelligence (AI) models via APIs, software development kits (SDKs), or applications without requiring deep data science expertise." (Groll, 2022) (24)

In its latest edition of its classic "Magic Quadrant" market analysis, we see which technology providers participate in this market and what level of maturity / execution capacity Gartner Group assigns to each of them:

Figure 1: Magic Quadrant for Cloud AI Developer Services



Source: Gartner (February 2021)

The magic quadrant shows us that there are at least 13 cloud technology providers in the market today that we can evaluate for the development of our platform. In the upper right quadrant appear the best qualified, those providers whose solutions have the highest ability to execute combined with a high degree of maturity, in the same are the large global computer companies of today IBM, Microsoft, Amazon Web Services and Google, about them the report says “Leaders have robust offerings in all three key service areas: language, vision and autoML. Their services are API-accessible and do not require developers to have data science expertise. Leaders also have ancillary services that support or enhance the capabilities of their core services. Leaders serve multiple geographies and support multiple languages.”

Hiring any of these providers guarantees us that the solution has the latest technology, products that are in constant development and evolution, global technical support, the possibility of making commercial agreements tailored to our organization.

Some words about two of the software providers leaders IBM (the higher completeness of vision) and Microsoft (the higher Ability to execute), About IBM with its solution Watson, pioneer AI solution with more than 10 years in the market: “IBM Watson solutions come pre-integrated and pre-trained on a flexible information architecture optimized to accelerate production and deployment of AI. They’re designed to allow developers to build models and create applications to help business make more accurate predictions, automate processes, interact with users and customers, and augment expertise. Developer tools make it easy to incorporate conversation, language and search into applications.

For example, ready-to-use Watson APIs for language, vision, speech and data such as Watson Discovery, a cognitive search and content analytics engine that extracts value from unstructured data and IBM Watson™ Knowledge Studio.

Watson Discovery’s engine works easily with Watson Knowledge Studio – allowing users to integrate custom models suited to any industry. This provides flexibility to apply Discovery’s document-enhancing capabilities with domain specific information – drawing upon public data and proprietary data.”

In the case of Microsoft: “Gartner believes that enterprise development teams will increasingly incorporate models built using AI and ML into applications. These services currently fall into three main functional areas: language, vision and automated machine learning (autoML). The language services include natural language understanding (NLU), conversational agent frameworks, text analytics, sentiment analysis and other capabilities. The vision services include image recognition, video content analysis and optical character recognition (OCR). The autoML services include automated tools that will let developers do data preparation, feature engineering, create models, deploy, monitor and manage models without having to learn data science.” Azure AI enables you to develop AI applications on your terms, apply AI responsibly, and deploy mission-critical AI solutions.

(Gartner, 2021) (25)

If it were my responsibility to select the technological platform to develop this solution, I would recommend doing so on one of these four options within the leader's quadrant, favoring the technological strength and maturity of the solutions over other providers that are still on the path of development of your solutions.

Real cases today

In the previous sections it has been mentioned that the level of maturity of various fields of technology makes it possible today to implement a platform with these characteristics. Next, I will present examples that are happening today from the digitalization of health in the case from Estonia to Silicon Valley startups that offer precisely the service that we are describing in this work:

The Estonia's digital health system

Estonia is known for being the most digital country in the world, the government has been carrying out a digitization program for years of all aspects of society and government services. Within this initiative to deploy digital government solutions called e-services, the health area has been considered one of the highest priorities in implementing this type of service, as described by the official Estonian site www.e-estonia.com in its health section https://e-estonia.com/solutions/healthcare/e-health-records/ patients, doctors, hospitals and the government itself benefit from having the information that the electronic health service delivers, from the point from the point of view of quick and accurate access to information as well as cost reduction.

Each person in Estonia that has visited a doctor has an online e-Health record that can be tracked. Identified by the electronic ID-card, the health information is kept completely secure and at the same time accessible to authorized individuals. The use of KSI Blockchain technology in the system ensures data integrity and mitigates internal threats to the data.

The Electronic Health Record (**e-Health Record**) is a nationwide system that integrates data from Estonia's different healthcare providers to create a common record that every patient can access online.

Functioning very much like a centralized, national database, the e-Health Record actually retrieves data as necessary from various providers, who may be using different systems, and presents it in a standard format via the e-Patient portal. A result is a powerful tool for doctors that allows them to access a patient's records easily from a single electronic file. Doctors can read test results as they are entered, including image files such as X-rays even from remote hospitals.

KSI Blockchain technology is used to ensure the integrity of retrieved electronic medical records as well as system access logs.

With the implementation of the e-health system, the initial benefits mentioned are real and contribute to a better quality of life for the population, but additionally, other solutions have been built on the basis of this system, such as the **e-Ambulance**, is a quick-response solution that can detect and position an emergency phone call for the responding ambulance within 30 seconds and send the emergency ambulance to the point of need quickly. In an emergency situation, a doctor can use a patient's ID code to read time-critical information, such as blood type, allergies, recent treatments, on-going medication, or pregnancy.

Also, they are implemented the **e-Prescription** application, a centralized paperless system for issuing and handling medical prescriptions. When a doctor prescribes medicine using the system, he or she does so electronically, with the aid of an online form. At the pharmacy, all a patient needs to do is present an ID-card. The pharmacist then retrieves the patient's information from the system and issues the medicine.

So, here we have a real case where the combination of a successful digital transformation project combined with blockchain technology has delivered a complete solution of digital health that is happening today.

Examples of current companies working under this idea

The "data commons" concept and the National Cancer Institute case:

In his article "How Data Commons Can Support Open Science "Robert Grossman explains the concept of "data commons", data commons are used by projects and communities to create open resources to accelerate the rate of discovery and increase the impact of the data they host. It is important to note that data commons are not designed for individual researchers working on an isolated projects to ignore FAIR principles and to dump their data to satisfy data management and data sharing requirements.

More formally, data commons are software platforms that co-locate: 1) data, 2) cloud-based computing infrastructure, and 3) commonly used software applications, tools and services to create a resource for managing, analyzing and sharing data with a community.

A good example of how data commons can support open science is the **Genomic Data Commons (GDC)** that was launched in 2016 by the **National Cancer Institute (NCI)**. The GDC has over 2.7 PB of harmonized genomic and associated clinical data and is used by over 100,000 researchers each year. In an average month, 1–2 PB or more of data are downloaded or accessed from it. The GDC also interoperates with three cloud computing platforms: Broad's FireCloud, the Seven Bridges Genomics Cancer Genomics Cloud, and ISB's Cancer Genomics Cloud.

The GDC supports an open data ecosystem that includes Jupyter notebooks, RStudio notebooks, and more specialized applications that access GDC data via the GDC API. The GDC saves the research community time and effort since research groups have access to harmonized data that have been curated with respect to a common data model and run with a set of common bioinformatics pipelines. By using a centralized cloud-based infrastructure, the GDC also reduces the total cost for the cancer researchers to work with large genomics data since each research group does not need to set up and operate their own large-scale computing infrastructure.

(Grossman, 2019) (26)

So, in this example we can see that today we have the possibility to access big quantities of specialized data “as a service” taken advantage of some of the last tendencies in technology like, big data, cloud computing, fast internet access and APIs.

Example of a startup that implemented this Platform idea

A very interesting example of how some startups are beginning to produce platforms like we are describing in this chapter is “Arterys” (**www.arterys.com**) a startup based in San Francisco who offers different AI services to analyze diagnostic images. According to its web page: The Arterys platform extracts actionable insights from medical images to add clinical value, improve diagnostic decision making, efficiency and productivity. Arterys is the medical imaging AI platform allowing you to weave leading AI clinical applications directly into your existing PACS or EHR driven workflow to make it a natural extension of what you already do. Accessible anywhere from any validated device via the cloud for faster performance, ease of deployment with no PHI exchange and completely secure. Arterys is the market leader in bringing human and AI together to improve patient outcomes through precision medicine and insights not previously achievable. Physician experience is improved by automating findings and results – removing the tedium of radiology.

One of the modules that is part of the platform is “Breast AI”, it has the objective of provide early detection of breast cancer, as the web site says “Breast AI provides innovative solutions for breast cancer detection, measure breast density and assess personalized risk that offer clinically proven benefits to clinicians and patients, and are designed to optimize efficiency, enhance the patient experience, and improve outcomes. Breast AI uses deep learning technology that is intended to be used concurrently by radiologists while reading digital breast tomosynthesis (DBT) exams. The algorithm detects soft tissue densities (masses, architectural distortions and asymmetries) and calcifications in 3D DBT slices. The suspicious areas that are detected and highlighted and the unique certainty of finding and case scores assist

radiologists in identifying and assessing soft tissue densities and calcifications that may be confirmed or dismissed by the radiologist”

A second example is a module called “Neuro AI” who allows “The Neuro AI platform is a vendor neutral comprehensive suite of clinically useful AI neuroimaging applications that provide physicians with fully automated, and easy-to-interpret customizable reporting -- facilitating fast and accurate diagnostic and treatment decisions for stroke, neurodegenerative disease, multiple sclerosis and brain tumor patients. The neuro oncology suite provides you with the tools to non-invasively distinguish tumor from pseudo-progression with >95% accuracy. IB Neuro offers the most robust algorithm to post-process your DSC perfusion MRI, IB DeltaT1 allows rapid and objective identification of TRUE enhancing regions in pre- and post-contrast T1 MRI. IB Diffusion calculates (ADC) maps, and other diffusion parameters and IB DCE for Automated generation of perfusion parameter maps (Ktrans, Vp) all with zero clicks.”

Then, we can see that the idea of a platform dedicated to provide early diagnosis to different type of cancer with the objective of grow the success patient’s treatment is today a reality, a company (surely not the unique one) is providing the service thru the web, and although is incipient we are looking the birth of a new era in cancer detection.

Non happy case

Although throughout the previous chapters multiple success stories have been presented in the use of deep learning as a technological support to achieve early diagnoses and reduce the research cycle of new drugs against cancer, not all of them are success stories in the application of these new technologies in this field. It should not remain in the reader's mind that the implementation and training process of an AI engine in a field as sensitive as the one we are studying is a simple task or that it can be done with "out of the box" software from minimal settings. Next, I will summarize the case of the “Oncology Expert

Advisor (OEA)" project developed by the MD Anderson Cancer Center (MDACC) in the USA.

The MADACC:

Founded at the University of Texas in 1941, MDACC was the first comprehensive cancer hospital to affiliate with a major university and was considered one of the most prominent cancer centers in the world. MDACC's mission was to uncover knowledge and share discoveries about cancer research throughout the world, eventually eradicating cancer altogether. In 2014, MDACC employed more than 20,000 people and treated 127,000 patients. In 2016, MDACC's total revenue was more than $4.4 billion. Some of the MDACC big contributions to the science were, the development of the first affordable radiotherapy machine in the 1960s, the beginning of use combination chemotherapy to treat adult cancers, by the 2000s, MDACC had become known as a vanguard trial site for smart drugs that were able to target cancer cells at the molecular level.

The technology partner, IBM Watson

Since 2000s, IBM researcher had been searching for a major project that could also reveal new business opportunities. Following its groundbreaking innovation with Deep Blue, the computer that beat the chess champion Garry Kasparov in 1997, the research team was looking for challenges that had the potential to garner wide media attention. The team landed on the TV Q&A show *Jeopardy*, which would require rapid technology development in the field of Q&A. Over the next three years, IBM's team constructed the foundational architectures of the AI product Watson, called "DeepQA". At that time, computers were exceptionally good at making calculations but struggled with tasks such as interpreting human language, so IBM's research team implemented new ways of processing texts and contextual information implementing new algorithms for parsing, question classification, question decomposition, automatic source acquisition and evaluation, entity and relation detection, logical form generation and knowledge representation and reasoning obtaining the base of the new AI

product called “Watson”. IBM research’s core algorithmic team of 20 researchers spent nearly three years testing and tweaking the technology and eventually overhauled everything about their original Q&A architecture. By 2011, Watson had played over 100 rounds of Jeopardy against past winners but shocked the world after besting two grand champions in a $1 million match.

In August 2011, IBM believed it could leverage Watson’s advanced Q&A capabilities to assist healthcare professionals in decision-making, in three months 107 new staffers who specialized in NLP and ML joined the effort. By 2012, IBM had already begun piloting Watson with two healthcare organization, one of them an competitor of MDACC for which Watson helped choose the best therapy options for cancer patients.

The OEA Project

In early 2012 some University of Texas senior officials began to discuss whether MDACC should engage IBM’s Watson unit in a partnership to study how they could apply cloud computing technology to increase cancer care quality and access. **MDACC’s goal was to help researchers make new discoveries in cancer diagnosis, care, and treatment process as well as to recognize and avoid adverse events throughout the care continuum.** On June 2012 MDACC and IBM agreed to explore ways to pilot Watson technology at MDACC. By October 2012, MDACC and IBM began developing a plan to build a tool based on Watson. The MDACC team hoped the OEA would assist doctors worldwide with cancer diagnosis and treatment recommendations. If everything went according to plan, even cancer patients in rural communities would have access to world-class cancer expertise and the latest research. For example, a physician without direct access to a leukemia expert would be able to insert a patient’s data into the OEA and receive personalized advice as if they were consulting with a human expert who had also recently reviewed millions of the latest papers on leukemia and recalled every important detail. In October 2013, MDACC revealed that it had collaborated with IBM to build the OEA.

Project Achievements

The team programmed the OEA with a knowledge base of over 23 million summaries from medical journals and MDACC's internal HER. They also loaded a training set into the OEA, preparing it to identify and diagnose a select number of conditions and treatment options. After a physician submitted a case to the OEA, ML algorithms examined its knowledge base to build a profile of a patient's medical history and make recommendations for treatment or clinical trials. In October 2014, the system had ingested data from 10.000 leukemia patients, including additional medical research on the subjected, however the team leader considered that more data from sources around the world would be needed to advance the OEA's knowledge. In December 2014, he requested an additional budget to support the project and start an expansion to begin developing a tool for lung cancer. By the end of 2015, and after apply several modifications, the OEA's matching performance for approved therapy options improved to 99,9% recall and 88% precision. Additionally, the OEA's clinical trial screening boasted a recall and precision of 97,9% and 96,9%.

Problems and risks.

In October 2012, MDACC launched the "Adaptive Patient-Oriented Longitudinal Learning and Optimization" (APOLLO) program, which aimed to centralize the center's medical data, make workflows more efficient, and build a standard process for collecting patient history. To advance this strategy, MDACC also began scouting for a new EHR system. Officials thought a new system would better integrate their patient data in to their clinical practice, leading to better patient outcomes. By May 2016, the new EHR system rollout was completed, and the OEA could no longer interpret data from the MDACC's patient records, so the team has to invest time in remapping the OEA to the new EHR.

Many doctors were unsure about using AI to aid physicians in a clinical setting. The implementation of technology like the OEA represented a paradigm shift in

medical care in MDACC, success for team require changing existing workflows, learning strategies for clinicians, and adopting new behaviors in clinical tasks.

MDACC had not also tested the OEA outside of their own facilities, making it impossible to generalize the results of the controlled introduction to a broader audience.

Integrating and developing the OEA into the MDACC's clinical process was already costly, by 2016 the MDACC has expended more than u$s 62.000.000 in the OEA project and the result is an application that they just can use for leukemia and some lung cancer cases, with the concern about the chance of use the solution in other medical centers. The next picture shows the OEA project financials chart:

Exhibit 11 OEA Project Financials (in USD), 2016

Fund Used	Current Free Balance	Contract	Total Paid from Fund
Genome Medicine Initiative	43.19	PwC Network Democratization PwC Value Capture	71,923.37 630,140.00
Artificial Oncology Intelligence	N/A	IBM Watson PwC Business Plan	15,400,000.00 2,223,702.88
Big Data Platform/TRA	(1,513,121.30)	PwC BDI Bridge PwC Network Democratization	2,201,803.00 2,459,402.91
Bosarge Apollo Watson Fund	330.00	IBM Watson PwC Value Capture	2,000,000.00 1,699,920.00
Low OEA Promo Video	60.00	PwC Value Capture	149,940.00
Oncology Expert Advisor Development Fund	(9,772,671.53)	IBM Watson PwC Network Democratization	11,985,559.94 12,896,174.26
OEA Democratization	23.44	PwC Network Democratization	424,976.53
Jordan Network Democratization Pilot	41.67	PwC Network Democratization	84,958.33
Lung OEA	(300,000.00)	IBM Watson	9,800,000.00
Various Donors/Research	41.67	PwC Network Democratization	84,958.33
Project Total			**62,113,459.55**

Source: MD Anderson Development Office, "Special Review of Procurement Procedures Related to the M.D. Anderson Cancer Center Oncology Expert Advisor Project," The University of Texas System Administration, Appendix C, p. C-1, November 2016, https://bit.ly/3ijXgtq, accessed May 2020.

Some lessons about this non happy case:

MDACC's decision to start a project along the lines of the OEA was reasonable on the grounds that the medical center has 70 years of experience in cancer

treatment and research, possesses vast amounts of data, patient medical records, and literature on cancer. topic, AI was at the time an accessible technology for large organizations, so as an idea it had a reasonable basis. By the beginning of 2012, choosing IBM as a technology partner was a natural decision since, of the technology giants, it was the most advanced company in this field due to the success demonstrated with DeepBlue and DeepQA.

We can cite the following points as risks that led to the project in mid-2016 being far from the company's expectations:

Watson was very successful in Q&A but for health scenarios it was still in the experimental phase, in fact Watson for health as a product was partly developed in conjunction with the MDACC in this project, so we can conclude that the state of maturity was initial at that moment.

The medical center makes a strategic mistake by launching the OAS project and the APOLLO project at the same time, since the AI engine needs data as its fuel, changing the transactional system that provides the data from the medical records of the patients to the AI without having made an adequate integration (in light of the results at the time of implementing the APOLLO project) was a mishandled risk.

Financially, with more than u$s62.000.000 invested on the project just to have a solution that works with two types of cancer and, its applicability is only within the MDACC, creates a very bleak scenario for the MDACC board.

(IBM Watson at MD Anderson Cancer Center, 2021) (27)

Therefore, it is in my interest to clarify with this case that not all AI implementations in fields as sensitive as cancer diagnosis lead to success stories, risks and context conditions must be considered and handled appropriately as in any other technology project in order to reach a satisfactory end.

Conclusions

In this paper we have reviewed the state of progress and maturity of artificial intelligence in today's world, as it is present in topics as diverse as virtual assistants, autonomous cars or search and recommendation engines. In particular, the technology called "deep learning", a subset of machine learning oriented to the processing of images and unstructured information, has made it possible, since its widespread use in software platforms and cloud computing providers, to tackle a previously impossible approach in the fight against cancer, the massive analysis of diagnostic images, laboratory data and medical literature to train AI applications to help health professionals in the early diagnosis of the disease, thus increasing the possibility of successful treatment.

But it is not only about the technological development of artificial intelligence, other factors in parallel have contributed so that together these solutions can be a reality, we can mention the massive deployment of the Internet to practically the whole world, the massification of diagnostic devices such as being magnetic resonators, tomographers and X-ray devices, cloud computing and the platform economy all at the same time means that for the first time in history there is the possibility of developing applications of these characteristics and having the information to run them in a successful way.

The six types of cancer that we have presented in chapter IV (cervical, brain, prostate, breast, lung and skin) added to the case of histopathological diagnosis and a brief example applied to research on COVID-19, show that it is not about an isolated case but rather it is a technological approach of increasing applicability and that is beginning to be used in treatment centers in different parts of the world.

In the development of chapters IV and V we have stated that, in fact, there are three key areas where AI can play a role in addressing cancer. Firstly, AI enables an early diagnosis through diagnostic images analysis.

Secondly, AI and deep learning can be used for personalized treatment and medicines by making use of the patient's electronic health records, data from sensors and wearables. By using medical history and the characteristics of the tumor, AI has the potential to come up with multiple treatment options for patients. The Natural Language Processing techniques based on AI have shown potential in predicting the development of diseases across healthcare systems.

Thirdly, AI can be used in drug development and clinical trial designs. There are multiple stages of drug discovery and AI can be used in new drug discovery by designing protein structures, target validation and managing drug trials. It is expected that with the introduction of AI, not only will the costs of drugs reduce, but it will also enhance the drug discovery process, which currently takes as long as 10-15 years.

Related to business aspects, we can say that platform like that we described in chapter VI could deliver a set on services to different communities, like:

Data ingestion service: Service dedicated to capture the data (images, test results, medical literature) that could be a source of money for medical centers, diagnosis centers, research centers and individual doctors.

Diagnosis validation service: service dedicated to validate the patient tests results in an early way giving to the doctor a suggested opinion about the patient's possibility to develop cancer and suggestions of treatments. This service is the main platform's income source.

Drugs research service: Service dedicated to assist researchers and investigation laboratories in the drugs development process. This service is the secondary platform's income source.

Statistics service: Service dedicated to provide statistics about cancer based on the data accumulated by the platform and the capacity of the AI engine to do

different types of calculus, projections and statistics. This service is a secondary platform's source of income.

All with the objective od accomplish with the platform's value proposition "to unite producers of medical diagnostic images, diagnostic labs results and medical literature with health professionals who wish to validate the tests performed on a patient in order to detect possible forms of cancer early and act accordingly to prevent cancer disease development".

In economic aspects, we saw that the development of the platform requires a big investment, that could be done by international health organizations like WHO, PAHO in a first stage and then continue with investment rounds up to the third year where the projections shows that revenues start to grow.

So, in base to this work's investigation question "Can the application of AI, in particular deep learning models, help reduce deaths from certain types of cancer?" we can conclude that taking into account that is incipient, we are able to say that with the multiple technological advances that are happening in parallel for a few years, the path of the delivery of AI on this area will contribute to a continuous decline in the percentage of deaths caused by cancer in the next years.

Expectations to 3 - 5 years.

Since the technological advances that make AI possible, helping to reduce cancer deaths, are recent, but already mature, we can say that the near future will show a massification of this trend. The factors that support this trend are the following:

• The evolution of communications with the implementation of 5G networks whose main benefits with respect to current data networks are tailored to the platforms we propose for the AI solution, namely:

- An excellent network speed estimated between 15 and 20 gigabytes per second.

- A lower latency that can reach a millisecond. The goal of immediacy is much easier to achieve.
- Greater network security. It is easier to control all access to the network and receive real-time warning of any threat.
- The connection of multiple devices without alterations in connection speed.
- A higher bandwidth that allows not only to receive information, but also to send it correctly.
- Facilitates the use of robotics, virtual reality, artificial intelligence and the Internet of Things.
- It favors the automation of processes and allows each employee's working day to be streamlined without wasting time.
- The conversion is much simpler, since the client will hardly need help to know how to acquire a product or service given the speed of response of the corresponding website.
- Energy consumption will drop by around 90%, an important factor to save on monthly expenses.

- The continuous massification of imaging and ray diagnostic equipment, whose new generations are digital equipment connected to networks and the Internet and are the main providers of data for the AI engine.
- The digitization of medicine that we are already seeing today, as in the Estonian e-health case presented, will continue to expand to more countries and regions, enabling electronic health record systems to be integrated with the AI platform required your data, either on an ad hoc basis as a service or permanently as part of a common solution.
- The irreversible trend of moving corporate systems and developing new projects in the cloud that companies of all sizes and sectors are carrying out and that will make the cloud the natural environment for deploying complex software and service solutions, given its State-of-the-art Software as a Service, automatic horizontal and vertical scalability, and pay-per-use scheme, among other benefits of cloud computing.

- The permanent development and evolution of AI, which has already reached a point of massification such that the technology giants offer lines of products and services of ML, DL, NLP, specific programming languages, development frameworks and vertical applications based on in AI such as Watson for Health, Watson for financials, Watson for customer service etc. Also, companies focused on AI have emerged that provide AI services and products as a core business, with which the AI offer in the market is more than ample to absorb the growing market demand. At the same time, university careers, technical specializations and courses have emerged to train professionals who can dedicate themselves to a market that is increasingly in demand.

Therefore, we can conclude that based on what has been developed today and is being used in AI systems dedicated to early diagnosis of cancer and in the development of new drugs, added to the advances that we detail, there will be in the next 5 years is when humanity has begun the path that will allow it to lower cancer mortality in the short term and have real chances of reaching a cure in the next 20 to 30 years.

References

0. World Health Organization (2022). Retrieved from https://www.who.int/news-room/factsheets/detail/cancer#:~:text=Other%20terms%20used%20are%20malignant,is%20referred%20to%20as%20metastasis.

1. Our World In data – Max Roser, Hannah Ritchie (2019). Retrieved from https://ourworldindata.org/cancer#deaths-from-cancer

2. Statista – Matej Mikulic (2021). Retrieved from https://www.statista.com/statistics/696208/oncology-costs-worldwide/

3. Statista – Matej Mikulic (2021). Retrieved from https://www.statista.com/statistics/373946/global-spending-and-growth-in-oncology-market/

4. Statista (2016). Retrieved from https://www.statista.com/statistics/696174/global-oncology-spending-by-region/

5. Statista – Matej Mikulic (2017). Retrieved from https://www.statista.com/statistics/788894/randd-costs-for-select-cancer-drugs/

6. Statista – Matej Mikulic (2017). Retrieved from https://www.statista.com/statistics/788787/randd-start-to-approval-time-range-of-select-cancer-drugs/

7. Gartner Group (2020). Retrieved from https://www.gartner.com.au/en/articles/2-megatrends-dominate-the-gartner-hype-cycle-for-artificial-intelligence-2020.

8. Artem Oppermann (2019). What is Deep Learning and how does it work? Retrieved from https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac

9. Artificial Intelligence – Opportunities in cancer research (August 31, 2020). Retrieved from https://www.cancer.gov/research/areas/diagnosis/artificial-intelligence

10. Hu L (2019) – Leveraging AI to improve detection of cervical precancer infographic. Retrieved from https://www.cancer.gov/research/improve-detection-of-cervical-precancer-leveraging-ai-infographic

11. Neil Savage (2020) - How AI is improving cancer diagnostics (March 25, 2020). Retrieved from https://www.nature.com/articles/d41586-020-00847-2.

12. The Royal Marsden Cancer Charity (2020) – Using AI technology to benefit cancer patients during the COVID-19 pandemic.

https://www.royalmarsden.nhs.uk/using-ai-technology-benefit-cancer-patients-during-covid-19-pandemic

13. Litjens, G., Sánchez, C., Timofeeva, N. et al (2016) – Deep Learning as a tool for increased accuracy and efficiency of histopathological diagnosis (May 23, 2016). Retrieved from https://www.nature.com/articles/srep26286#Bib1

14. Kinnor Das, Clay J. Cockerell, Anant Patil, Pawel Pietkiewicz – Machine Learning and its application in skin cancer (December 2021). Retrieved from https://www.researchgate.net/publication/357194148_Machine_Learning_and_Its_Application_in_Skin_Cancer

15. Gregor Urban, Kevin M. Bache, Duc Phan, Agua Sobrino - Deep Learning for Drug Discovery and Cancer Research: Automated Analysis of Vascularization Images (May, 2018). Retrieved from https://www.researchgate.net/publication/325437124_Deep_Learning_for_Drug_Discovery_and_Cancer_Research_Automated_Analysis_of_Vascularization_Images

16. Stefan Harrer, Prstik Shah, Bhavna Antony, Jianying Hu - Artificial Intelligence for Clinical Trial Design (July, 2019). Retrieved from https://www.cell.com/trends/pharmacological-sciences/fulltext/S0165-6147(19)30130-0

17. Toivo Äijö, Olli Kuivalainen, Sami Saarenketo, Jani Lindqvist and Hanna Hanninen – Book "INTERNATIONALIZATION HANDBOOK FOR THE SOFTWARE BUSINESS" (2005)

18. Alexandre Gonfalonieri – Why is it hard to build AI & data network effects (January, 2020). Retrieved from https://towardsdatascience.com/why-is-it-hard-to-build-ai-data-network-effects-fa010497e2cb

19. Pan American Health Organization - Health technology assessment. Retrieved from https://www.paho.org/en/topics/health-technology-assessment

20. Conor Stewart – Medical imaging market size U.S 2015-2025, by product (February, 2019). Retrieved from https://www.statista.com/statistics/975731/medical-imaging-market-size-us-by-product/

21. Anthony D. Montagnese – The Evolving Role of Radiation-Dose Considerations in Diagnostic Imaging (Spring, 2012). Retrieved from http://www.jlgh.org/Past-Issues/Volume-7---Issue-1/The-Evolving-Role-of-Radiation-Dose-Considerations.aspx

22. iData Research – Over 75 million CT scans Are Performed Each Year and Growing Despite Radiation concerns (August, 2018). Retrieved from https://idataresearch.com/over-75-million-ct-scans-are-performed-each-year-and-growing-despite-radiation-concerns/

23. M. Lakrimi – Magnetic Resonance Imaging (MRI) and its global impact in healthcare (September, 2018). Retrieved from https://www.researchgate.net/publication/331563555_Magnetic_Resonance_Imaging_MRI_and_its_global_impact_in_healthcare

24. Jane Groll – How Cloud AI developer services empower developers (June, 2022). Retrieved from https://enterprisersproject.com/article/2022/6/caid-services

25. Gartner Group - Magic quadrant for cloud ai developer services (2021). Retrieved from www.gartner.com

26. Robert L. Grossman – How Data Commons Can Support Science (April, 2019). Retrieved from https://medium.com/@rgrossman1/how-data-commons-can-support-open-science-40b0d978d2fb

27. Harvard Business School – IBM Watson at MD Anderson cancer Center, case number 9-621-022.

28. Gunther Eysenbach – Characterizing artificial intelligence applications in cancer research: A latent Dirichlet allocations analysis (September 15, 2019). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6774235/

29. Bin Li – AI in oncology: when science fiction meets reality (October 4, 2018). Retrieved from https://www.aioncology.org/index.php/AIIO.

30. Davide Cirillo, Iker Nuñez-Carpintero, Alfonso Valencia – Artificial Intelligence in cancer research: learning at different levels of data granularity (February 02, 2021). Retrieved from: https://febs.onlinelibrary.wiley.com/doi/10.1002/1878-0261.12920

31. Different ideas and information retrieved from https://cancercenter.ai/.

32. Mario Coccia – Artificial intelligence technology in oncology: A new technological paradigm (September 26, 2018). Retrieved from https://arxiv.org/ftp/arxiv/papers/1905/1905.06871.pdf

33. Ruma Bhargava – An expert explains: 3 ways AI can help in the global battle against cancer (Febreary 26, 2021). Retrieved from https://www.weforum.org/agenda/2021/02/cancer-treatment-ai-machine-learning/