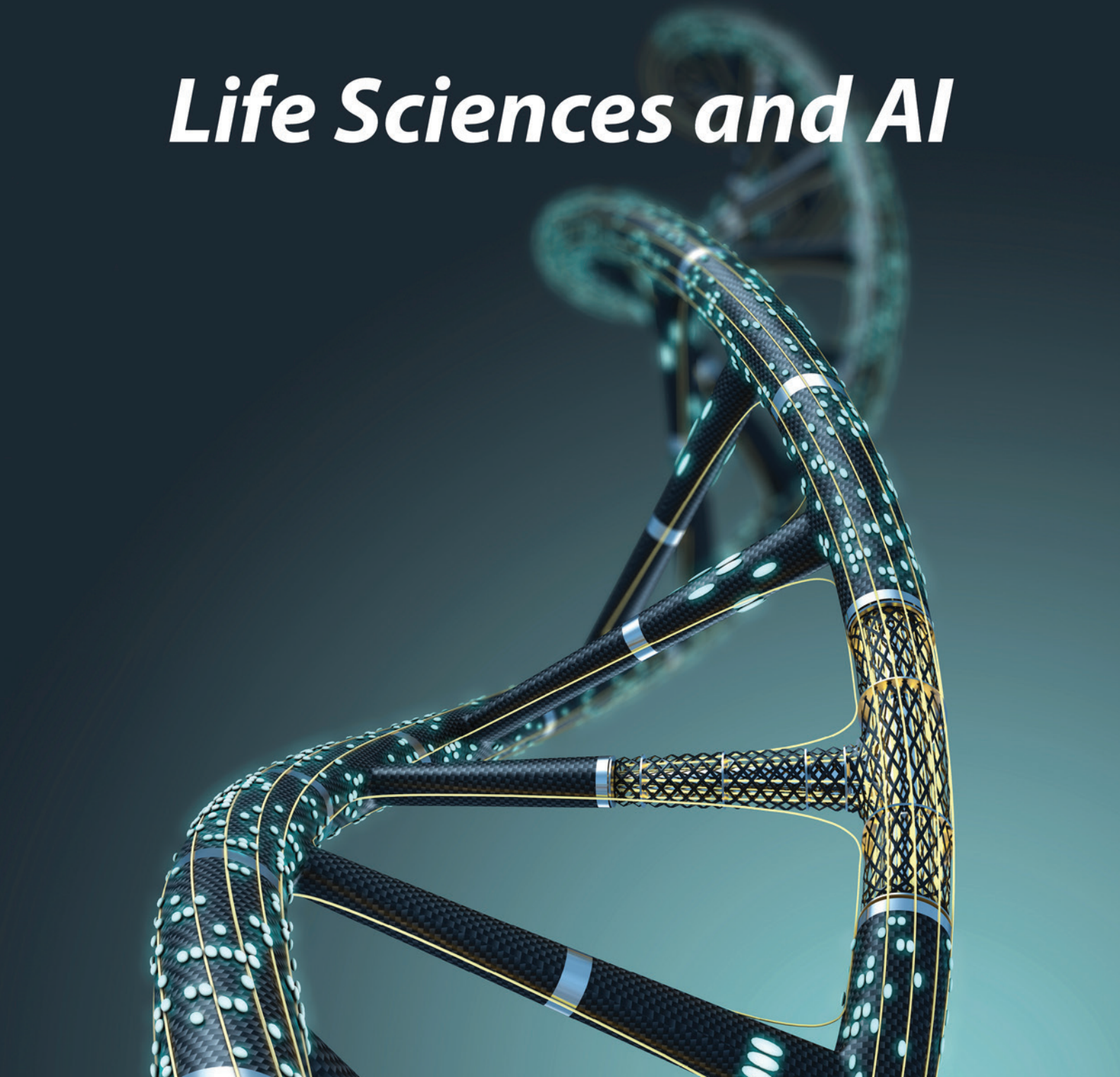
NEXT FRONTIER

JOURNAL OF NEXT FRONTIER FOR LIFE SCIENCES AND AI

2021 Volume 4 Issue 4

Life Sciences and AI



NEXT FRONTIER

JOURNAL OF NEXT FRONTIER FOR LIFE SCIENCES AND AI

2021 Volume 4 Issue 4

PUBLISHER

ABA Yayıncılık Dağıtım ve Pazarlama A.Ş.

abayayin.com

EXECUTIVE OFFICE

İnşirah Caddesi No: 29/7 Bebek/İstanbul

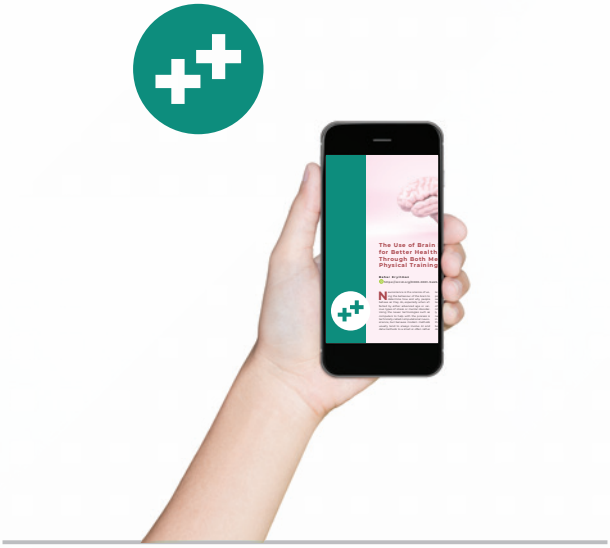
0 (212) 287 86 06

/NextFrontier

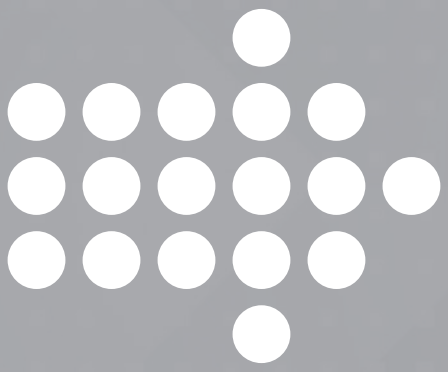
NEXT FRONTIER

JOURNAL OF NEXT FRONTIER FOR LIFE SCIENCES AND AI

2021 Volume 4 Issue 4



CONTENTS



The Improvement of Bio-Informatics Education through Use of Biological Concepts and Related Fields of Interest

Bahar Eryilmaz

https://orcid.org/0000-0001-9469-4151

Consisting of how and why we use biological functions and processes to inform ourselves about various applications of biological usefulness from treating cancer to less threatening aspects such as providing better plans for weight loss, we here state that bioinformatics as a field requires more foundations that draw from other fields as highly interrelated and necessary for learning about biology itself. Using the core competencies that make up biology schematics as bolstered by examples through its own shared field of viruses and its related but separate contributions to chemistry via the safety of food additives that provide both the variety and learning wealth needed of such a varying field, we begin with the words of Mulder et

al. Via the grasp of core concepts that produce the more basic learning material as needed of biology students.

Biology: Not so Simple, but Simply Necessary

Three steps were used by our authors in developing the core competencies: (1) defining the competencies needed for using bioinformatics, (2) defining a variety of user profiles describing distinct subgroups in need of training, and (3) defining how the competencies will apply to PLOS Computational Biology each user profile (scoring).[i] Mapping of bioinformatic competency was measured as applicable to: (1) the bioinformatics user; (2) the bioinformatics scientist; and (3) the bioinformatics engineer, with initial workshops finding these categories unrealistically narrow, as much discussion generated around separating bioinformatics scientists from bioinformatics engineers.

The roles were therefore expanded to include include: physicians, lab technicians, ethicists and biocurators, scientists (discovery biologist, academic bioinformatics researcher and core facility scientist), and engineers (bioinformatician in academia/research institute or software engineer). Similarly, scoring methods were also inefficient, settling on the Bloom's Revised Taxonomy terms that are knowledge, comprehension, application, analysis, synthesis, and evaluation.[ii]

The current modelled frame as mapped to competencies is given below (Table 1).

Competency \ Persona	Physician	Lab technician	Ethicist	Biocurator
A. General biology	knowledge to application	comprehension	knowledge	comprehension
B. Depth in at least one area of biology (e.g., evolutionary biology, genetics, molecular biology, biochemistry, anatomy, physiology)	application	application to evaluation	evaluation	application to evaluation
C. Biological data generation technologies.	knowledge	knowledge to evaluation	knowledge	knowledge
D. Details of the scientific discovery process and of the role of bioinformatics in it.	application to analysis	comprehension to analysis	knowledge to comprehension	comprehension to evaluation
E. Statistical research methods in the context of molecular biology, genomics, medical, and population genetics research.	knowledge to application	knowledge to application	knowledge to comprehension	comprehension
F. Bioinformatics tools and their usage.	comprehension	knowledge to analysis	knowledge	application
G. The ability of a computer-based system, process, algorithm, component, or program to meet desired needs in scientific environments/problem.	N/A	knowledge	N/A	comprehension to application
H. Computing requirements appropriate to solve a given scientific problem (e.g., system, process, algorithm, component or program; define algorithmic time and space complexities and hardware resources required to solve a problem).	N/A	knowledge	N/A	comprehension to application
I. GUI/Web-based computing skills appropriate to the discipline (e.g., effectively use bioinformatics and analysis tools through web).	knowledge	application	comprehension	application to evaluation
J. Command line and scripting-based computing skills appropriate to the discipline.	N/A	knowledge	N/A	comprehension
K. Construction of software systems of varying complexity based on design and development principles.	N/A	N/A	N/A	knowledge
L. Local and global impact of bioinformatics and genomics on individuals, organizations, and society.	knowledge	comprehension	application	comprehension
M. Professional, ethical, legal, security and social issues and responsibilities of bioinformatics and genomic data in the workplace.	application	evaluation	evaluation	analysis
N. Effective communication of bioinformatics and genomics problem/issue/topics with a range of audiences, including, but not limited to, other bioinformatics professionals	comprehension	application	application	application to evaluation
O. Effective teamwork to accomplish a common scientific goal.	knowledge	analysis	knowledge	analysis
P. Engage in continuing professional development in bioinformatics.	evaluation to analysis	application	application to evaluation	application

Table 1.
Mapping of competencies to bioinformatics user personas via Bloom's Taxonomy.[iii]

Levels of Use

Examples from (1) complete degree programs for which the competencies have proven valuable to overall curriculum design or refinement; (2) supplements to existing degree programs (i.e., specializations, tracks, certificates); and (3) training resources outside the context of specific degree programs all work as a cohesive means to further working systems of educational designs, most especially in fields of science.

Degree program-Africa

H3ABioNet (www.h3abionet.org) is a Pan African bioinformatics network for H3Africa that has "developed a bioinformatics training program for African scientists from the Human Heredity and Health in Africa (www.h3africa.org) consortium."[iv] Using bioinformatics training for a broad range of audiences through topics of genomics data analysis, selected existing master's courses were used to define and augment model with additional electives relevant to specific institutions based on their research priorities at mainly the MA level.

Planning these modules to skills needed of a bioinformatics specialist to generalize a subset of knowledge that would be needed by all students of the subject is critical, as was done here, with electives then being a good option for covering specific research goals via the institution. Currently in use at at least two universities in Africa starting their first master's programs, there emerged a need for more basic "introduction to bioinformatics" training as well. In response, H3ABioNet developed an Introduction to Bioinformatics course delivered remotely across multiple countries, the focus on simpler parts of the topic such as use of tools such as algorithms for biological solutions in a practical environment of solving problems via personal experience.

Undergraduate and graduate research degree programs-US

Carnegie Mellon University offers degree programs in computational biology at several levels, including a BS in computational biology (since 1989), an MS in computational biology (since 1999), a PhD in computational biology (offered jointly with the University of Pittsburgh since 2005), and required training in computational biology ""or students primarily training for work in experimental biology."[v] Making Introduction to Computational Biology (ICB) a core requirement of every undergraduate biological sciences major as a use of general requirement as universally applicable for all students in the field., while the Carnegie Mellon/University of Pittsburgh joint PhD in computational biology offers an example at another extreme of the spectrum: a full multi-year training program for students expected to become experts in computational biology of any expertise, from research to teaching.

Undergraduate degrees at a small liberal arts college

Saint Vincent college began an bioinformatics program in 2005. With less than 20 students in the major, the programme was split into two because students from the last 2 groups of the three that tended to enroll as: (1) students who enjoyed both biology and computation and were good at both; (2) students who enjoyed biology but struggled with the programming courses; and (3) students who enjoyed programming but struggled in the upper biology courses, particularly labs.[vi]

The split meant that key courses became somewhat divergent rather than the former mix of all maths and biology, corresponding to bioinformatics users and bioinformatics scientists, as of 2013.

There also exist several certifications offered by institutions, as well as projects such as the use of bioinformatics in clinical settings (such as the 100,000 genome project in the UK that seeks to sequence the said number for better-informed treatments) that allow a transition to the use of methods such as virology as a form bioinformatics that can, as genomics do, play an invaluable role through the power of adaptivity that both carry.

Viruses: Flexibility for Semi-Rigid "Separate Scientific Fields" (and Their Educational Tools)

Despite virology as an excellent partner to bioinformatics, little was done in collaboration apart from some pioneering work on HIV-1 and influenza[vii], but such partnership is required in our era of big data to deal with the vast quantity of information that makes up forms of information such as DNA sequences as used in virology in a fluid manner that prioritizes efficiency to the opposite that is bioinformatic use of detail at a slower pace. Below, the historical line of noteworthy viral problems that show how limited the current view of use actually is.

Virology is usually defined as the study of non-organisms that cause disease in living creatures, but the biosphere actually contains an estimated 10 to the 31st power, 10 times the amount of bacteria but few even identified due to their "non-ills" status (at least, as far as we currently presume). Therefore, viruses as "only" parasites is an invalidity that now means we can make use of their properties to add to research how viruses are able to safely transfer and store genetic information of their host population and influence entire biogeochemical cycles. without being pathogens.[ix]

New genome sequencing technologies that use tools of big data can now become methods to answer questions that might otherwise need lengthy and/or expensive processes rather than simple replicative sequences that many types of safe viruses might

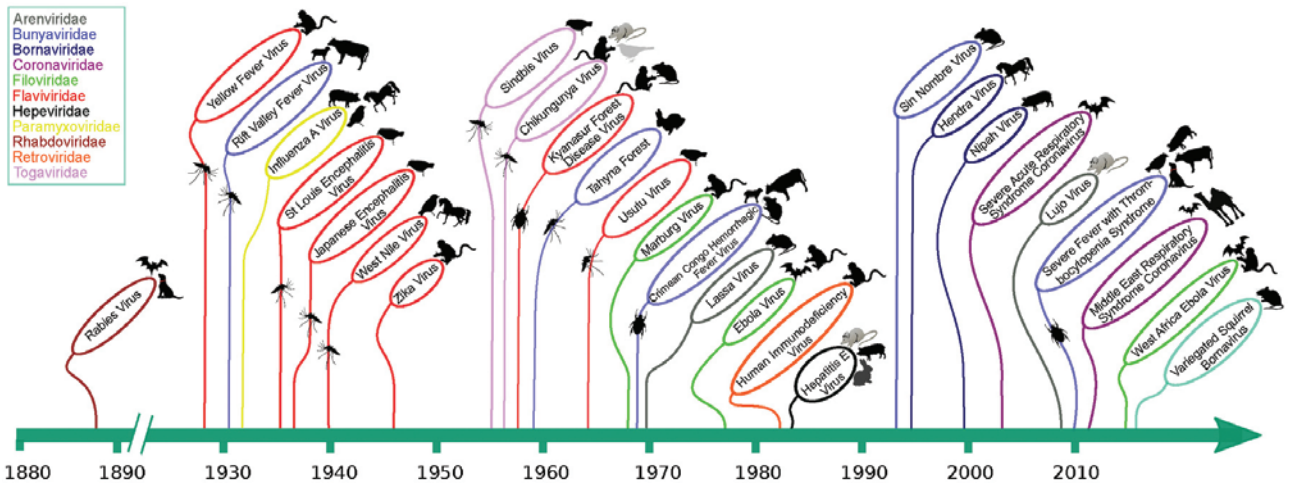


Figure 1.
Unknown/new viruses emerge all the time. Figure is an extended and redrawn version of https://www.microbiologysociety.org/publication/past-issues/zoonotic-diseases.html[viii]

provide scientists in many forms, with many differing styles of functional research uses based on the species/family group.

The Tools

Virus-specific databases

A few virus-specific databases exist so far for virologists but a general database for all viruses needs to be urgently developed, with EpiFlu is currently the most complete collection of genetic sequence data of influenza viruses and related clinical and epidemiological data, a result of the common incidence of colds versus all other types of ill-bearing viruses.

Viral genome de novo assembly tools

There have been many tools developed for whole genome assembly, but are meant to assimilate the repetitive elements in the viral UTR, as the low and uneven read coverage mean little information can be pulled from such a small genetic sample. However, algorithms dedicated for single-cell sequencing, such as SPAdes [22] or IDBA-UD [21] work well for tested samples, allowing at least a rudimentary form of educational efficiency until methods progress.

Viral Phylogeny

Phylogenetic trees are a standard graphical model for viral phylogenies in the literature, but their use cannot account for "variation in evolutionary rate, lack of physical "fossil records" of viruses, and confounding evolutionary relationships between viruses and their hosts"[x] as comprising the descriptive elements that allow methodical analysis of both evolutionary processes such the relationships between viruses and their hosts (including processes that create quick shifts such as horizontal gene transfer).

Virus annotation and genotyping

Genome annotation is the process of "identifying gene locations, functions, and the coding and non-coding regions of a genome."[xi] GLUE an open-source software toolkit works well for storage and interpretation of sequence data, even with multiple sequence alignments (MSAs) existing-usually an issue due to confines of available space.

With correct identification in all instances as is perhaps possible with GLUE students are therefore capable of determining pathogens with accuracy, and then multiplied for research use via PRiSM as a set of algorithms that creates primers for amplification and sequencing of short 7 viral genomes (with population diversity intact).

Last, we analyze the use of food additive categorization by Zhang et. al as a framework for how students in bioinformatics my make use of not only progressive AI for handling large datasets as an obvious solution, but learn to use resources from related fields to more qualitatively assess how biology works in tandem with other sciences as internally with its own concepts as extending in more creative ways tha usually imagined.

And now, Practical Application (of Food) with AdditiveChem

Data and Its Integration

The model was creating by first, because food additives varies among countries, but here we mean anything that is added to otherwise bland produce: Food nutrition fortifiers, processing aids, pigments, and spices,[xii] listing them in one place. Provided by

the United States (US), World Health Organization (WHO), and the European Union (EU) for representative dataset consisting of three: Additives approved by the US (FAAUS), approved by WHO (FAAW), and approved by the EU (FAAEU).

Substances are described with their additive name, FEMA number, Chemical Abstract Service (CAS) number, and related Code of Federal Regulation. In the EU databases, food additive data are identified by the additive name and E number (codes for substances that have been assessed for use as food additives within the EU), while the European Food Flavoring database are listed by the additive name etc. The database of 14,123 additives from the above there remained 9,064 unique food additives with a high level of food additive coverage for data use.[xiii]

Food additives regulation, usage specifications, and acceptable daily intake

Usage specifications are the most important reference data for the safe use of food additives, and numbered 8,896 usage specifications from the EU food additives and flavoring substances databases, and GSFA online database. Here, AdditiveChem also included 12,336 linkages between food additives and the Code of Federal Regulation (for example, fumaric acid as mentioned in laws 73.129, 73.3030, 73.34531, 73.61532, 173.21033, 175.10534, 175.32035, 176.18036, 176.30037, and 177.260038).[xiv]

Acceptable daily intake (ADI) is also a significant measurement that lists amount of a specific substance as ingested daily over a lifetime without an appreciable risk, the key measure for food additives on a global scale-the higher the value, the safer the additive.[xv]

Search methods and algorithm

Using AdditiveChem, there are several retrieval methods by using chemical information processing and multiple algorithms such as structure retrieval, fragment retrieval, similarity retrieval, maximum common substructure (MCS) retrieval, and text retrieval options.[xvi] Structure retrieval is based on the SMILES matching method. To cite one example used in the study, the SMILES of all food additives were pre-calculated and stored in the database that is then used via similarity retrieval to match it to the SMILES of the molecule that is requested through a search of any of the 5 kinds above. For another, the 'fragments' used for searches are a set of connected atoms that may have associated functional groups, with particular ones usually corresponding to physical and chemical structures (and, as such, related biological function), such as azobenzene or methylphenol.[xvii]

Users and applications

Increasing use of searches to quickly separate useful information from the extraneous is becoming the norm in scientific studies through computational biology use of large compound structure databases, such as the PubChem and ZINC, but the authors made use of AdditiveChem (http://www.rxnfinder.org/additivechem/) a more specific database "because it was manually corrected by food chemistry scientists."[xviii] Because it is seen and updated, this also means that it usually remains the most relevant source, with quality data always available through user participation-Updated reports are published on the AdditiveChem website homepage, while the various search functions are inclusive of many names that cover many chemical compounds in different scenarios (comparative table below).

	JECFA	GSFA	EAFUS	EU	FADB	AdditiveChem
Laws and usage specification	x	√	√	√	x	√
ADI	√	x	x	x	x	√
Molecule structure	x	x	x	x	√	√
Physicochemical properties	x	x	x	x	√	√
Toxicology	x	x	x	x	x	√
Targets	x	x	x	x	x	√
ADMET properties	x	x	x	x	x	√
Taxonomy	x	x	x	x	x	√
Total entries	2911	296	3968	2928	2540	9064

Table 2.
Summary of information provided by different food additive databases.[xix]

Ranging from detailed data for additives to possible side-effects, most food additive molecules are pigments. After retrieval of the molecule (possibly even by drawing it), users can browse through the detailed data of each food additive via "molecular structure, chemical and physical properties, absorption and distribution potential, metabolism, excretion and toxicity properties, biosynthesis and biodegradation methods, usage specifications, toxicological and risk assessment data, and targets in the human body."[xx]

This information base then ensures that users can readily investigate if illegal food additives are potentially toxic or influence certain tissues. For example, "in-vivo absorption of small molecule compounds across the gut wall can be estimated from the permeability of Caco –2"[xxi], and a download function is made available so the user may have a compendium on hand as required to cover more advanced concepts such as virtual screening or molecular docking.[xxii]

Conclusion

From the use of biology in 'unrelated' food additive chemistry to the use of viruses in a world absolutely filled with them, aside from the common cold as we tend to catch it, we see observe that there is more than one way to skin a cat; indeed, this is absolutely required nowadays, because the cat has become the size of a large mountain that threatens to crush any non-use of not only big data through AI means, but also absolutely requires use of information from many other fields of science if we are to solve larger questions such as cures for cancer by educating the next generation of students of bioinformatics with the appropriate training and responses.

Endnotes

i Mulder, N., Schwartz, R., Brazas, M. D., Brooksbank, C., Gaeta, B., Morgan, S. L., ... & Welch, L. (2018). The development and application of bioinformatics core competencies to improve bioinformatics training and education. PLoS computational biology, 14(2), e1005772

ii Ibid.

iii Ibid.

iv Ibid.

v Ibid.

vi Ibid.

vii Ibrahim, B., McMahon, D. P., Hufsky, F., Beer, M., Deng, L., Le Mercier, P., ... & Marz, M. (2018). A new era of virus bioinformatics. Virus research, 251, 86-90.

viii Ibid.

ix Ibid.

x Ibid.

xi Ibid.

xii Zhang, D., Cheng, X., Sun, D., Ding, S., Cai, P., Yuan, L., ... & Hu, Q. N. (2020). AdditiveChem: A comprehensive bioinformatics knowledge-base for food additive chemicals. Food chemistry, 308, 125519.

xiii Ibid.

xiv Ibid.

xv Ibid.

xvi Ibid.

xvii Ibid.

xviii Ibid.

xix Ibid.

xx Ibid.

xxi Ibid.

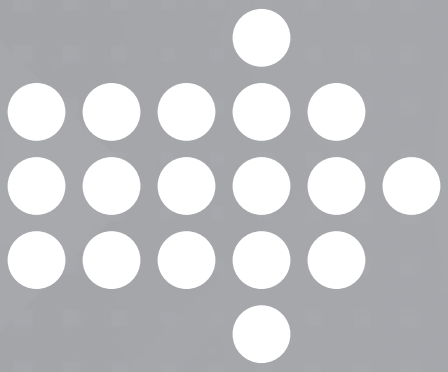
xxii Ibid.

References

[1] Ibrahim, B., McMahon, D. P., Hufsky, F., Beer, M., Deng, L., Le Mercier, P., ... & Marz, M. (2018). A new era of virus bioinformatics. Virus research, 251, 86-90.

[2] Mulder, N., Schwartz, R., Brazas, M. D., Brooksbank, C., Gaeta, B., Morgan, S. L., ... & Welch, L. (2018). The development and application of bioinformatics core competencies to improve bioinformatics training and education. PLoS computational biology, 14(2), e1005772

[3] Zhang, D., Cheng, X., Sun, D., Ding, S., Cai, P., Yuan, L., ... & Hu, Q. N. (2020). AdditiveChem: A comprehensive bioinformatics knowledge-base for food additive chemicals. Food chemistry, 308, 125519.



Artificial Intelligence Applications for Breast Cancer

Yasemin Yüksel

https://orcid.org/0000-0001-6106-3142

How can an age-old disease such as breast cancer be supported with artificial intelligence? Today breast cancer is a predominant type of cancer that affects many women (and sometimes men) all over the world and the morbidity rates are still high. Artificial intelligence holds the power to support medical technologies that can detect breast cancer lesions early on which is a significant way to reduce the yearly number of deaths. Artificial intelligence-enhanced applications for devices such as MRIs and other imaging machines like mammography and ultrasounds can offer more accurate and precise readings for physicians and radiologists in order to treat breast cancer early on and make the necessary changes in ther-

apy to prevent women all over the world from getting cancer or stopping it cold in its tracks (see Table 1).

Introduction

How often have we heard a friend or loved one get breast cancer? Breast cancer has been prevalent in women for many decades if not hundreds of years. But how many of us actually know what breast cancer is? Breast cancer affects the tissue of the breast and revolves around the uncontrollable growth of cells in this area that result in lesions or masses of cells that form a tumor. Breast cancer can remain in the breast area or move to other parts of the body. The statistics of breast cancer affecting women boil down to a 1 in 8 chance of getting it during their lifetime and for men the chances are 1 in 1000. Breast cancer tends to be genetically based and it is widely known that mothers can pass on the genes to their daughters and sisters can also share the genes prone to breast cancer. Many risk factors include familial ties and also revolve around whether or not a woman has had a benign case of cancer or if she has ever been treated with radiation or if she drinks alcohol. Many treatments for breast cancer are readily available that include, in severe cases, a mastectomy which involves removal of the entire breast tissue, removal of tumor(s), chemo or radiotherapy as well as hormone therapy.[i]

Breast cancer is widely common and early diagnosis is important while it is known to be difficult to diagnose. The more traditional methods of screening breast cancer include ultrasound, mammography, and thermography. Image processing today is enabled with artificial intelligence (AI) and support vector machine accuracies can be checked and compared. Computer-assisted methods such as these can aid in "reducing false positives."[ii] Advancements in AI have contributed to the ways we diagnose and treat breast cancer as well as identify new algorithms to improve former software tasks to ensure precision and accuracy in test results. But before AI was integrated into imaging for breast cancer it had humble beginnings at Stanford University when Professor John McCarthy coined the term AI in 1955. McCarthy envisioned an AI that with the proper study of learning, and every aspect of it, to include "features of intelligence" that we owls eventually be able to describe it to a machine and it "can be made to simulate it."[iii] Since McCarthy AI has grown its own subsets known as machine learning and deep learning among others (such as convolutional neural networks, artificial neural networks, and expert systems).

Machine learning can be described as the subset of AI that has the capacity, or intelligence, to "imitate intelligent human behavior" in problem solving. The functions of a machine learning system can be descriptive, predictive or prescriptive such that machine learning is able to use data to explain a phenomenon, or to predict what will happen in the situation, or it can make recommendations as to what activity or action should be taken in light of the data. The three types of machine learning are called supervised learning, unsupervised learning and reinforcement learning (RL). The first can be trained to identify labelled data sets, while the second finds its own patterns in unlabeled data. The final kind of machine learning is RL works on a trial-and-error system where it can notify about correct or incorrect decisions made in any one case. Deep

learning varies from machine learning in that it utilizes neural networks with multitude layers to process data. This is particularly helpful when image recognition involves "individual features" such as those in faces, and other body parts. Deep learning is particularly useful in the medical field and in diagnostics.[iv]

2-Imaging and Artificial Intelligence for Breast Cancer

AI is already being used for imaging purposes and has made great strides in the medical field. Conventional imaging machines such as MRIs (magnetic resonance imaging), ultrasound machines, and CTs (computed tomography) can use AI to give us enhanced-clearer, more accurate and precise images. For breast cancer, in particular, Sadoughi (2018) describes the various breast imaging methods: mammography, ultrasound and thermography. He presents the problems related to imaging methods such as the "presence of noise in images" and the problem with radiologists who cannot see the images clearly enough due to this "noise." With the advent of neural networks, a subset of AI, in the 1980s, imaging methods have improved so that the difficulty in the diagnosis of breast cancer has transformed. AI gives image processing the required "pattern recognition" as well as accuracy in the detection of masses (both benign and malignant). He argues that the aim of image processing is to address the "inherent problems associated with an image, including poor contrast, noise, and lack of recognition with the eye" and that we should "use techniques for making proper images of the human body, which are reliable for use in the diagnosis and treatment processes."[v]

Ultrasounds for Breast Cancer

Ultrasound machines are newly being enabled with AI. The first to be FDA approved was used to "capture images of acceptable diagnostic quality during adult echocardiography" in 2020.[vi] Current research is underway, since 2019, on how AI can enhance ultrasound for analysis of the breast (as well as for the prostate, liver and heart).[vii] Huang et al. (2005), in their study of "600" ultrasound images from diverse ultrasonic machines, were particularly interested in the ways Computer-aided diagnostic (or CAD) worked to better locate and identify both malignant and benign tumors in breasts. Yet, their studies prove that CAD can be effective for this purpose, but they do not speak of enhanced CAD systems (those enabled with AI) for their purposes.[viii] CAD has undergone software development and can now work with AI technologies (in sectors such as architecture), but its uses in medicine and for medical purposes is undetermined.

However, the problem with ultrasounds lies in the fact that they are not used for breast cancer detection, according to Johns Hopkins Medicine. The reason being that ultrasound (with or without AI?), cannot detect the signs of cancer early on, such as "tiny calcium deposits" or what are called microcalcifications.[ix] Because ultrasound lacks the radiation that is present in mammography screening it can be used for the detection of larger masses and is also nor reliable for women who have larger breast density.[x] It is also important to note that because ultrasound is ultimately controlled by the human technician, it would be difficult to implement AI into it though studies continue in this regard.[xi]

n	Task	Algorithms	No. of Cases	Results	Ref.
1	detect, characterize and categorize lesions	a supervised-attention model with deep learning	335	AUC=81.6%	(60)
2	classify lesions	radiomic analysis and CNN	1294	AUC=98%	(62)
3	characterize and classify lesions	the combination of unsupervised dimensionality reduction and embedded space clustering followed by a supervised classifier	792	AUC=81%	(63)
4	classify breast tumors	QuantX	111	AUC=76%	(67)
5	assess and diagnose contralateral BI-RADS 4 lessions	MRI radiomics-based machine learning	178	AUC=77% ACC=74.1%	(69)
6	assess tumor extent and multifocality	CADstream software (version 5.2.8.591)	86	AUC=88.8% Spe=92.1% PPV=90.0%	(70)
7	early predict pathological complete response to neoadjuvant chemotherapy and survival outcomes	linear support vector machine, linear discriminant analysis, logistic regression, random forests, stochastic gradient descent, decision tree, adaptive boosting and extreme gradient boosting	38	AUC=86%	(71)

Table 1-Applications of AI in Breast MRI

AI, artifical intelligence; MRI, magnetic resonance imaging; AUC, the area under the receiver operating characteristic curve; CNN, convolutional neural network; BI-RADS, Breast Imaging Reporting and Data System; ACC, accuracy; CAD, computer-aided detection; Spe, specificity; PPV, positive predictive value.

Magnetic Resonance Imaging (MRI) and Artificial Intelligence

Magnetic resonance imaging is a more expensive way to detect breast cancer, but it does offer better detection at earlier stages than mammograms.[xii] Newer types of MRI, such as the "fast" kind are, however, better at detecting cancer in breasts, to include dense ones, at an earlier stage.[xiii] Breast cancer is the number one killer of women worldwide and early detection of lesions and cancerous cells is significant to cure it. New types of imaging technologies are now increasingly dependent on the use of AI for better and less invasive testing of women no matter the size of their breasts which leads to better diagnosis and treatment.

An MRI machine consists of a "large, cylindrical (tube-shaped) machine that creates a strong magnetic field around the patient." MRIs do not use radiation and are dependent on radio

waves to work. "The magnetic field, along with radio waves, alters the hydrogen atoms' natural alignment in the body. Computers are then used to form a two-dimensional (2D) image of a body structure or organ based on the activity of the hydrogen atoms." MRIs can be used to detect breast cancer, however because the more conventional MRIs are not enhanced with AI, they are not able to always locate small breast lesions like mammography can. Breast cancer can be tracked with the MRI after a woman has been injected with a die to locate lesions through coloring of blood in the veins. MRIs are better used to detect cancer in younger women (less than 40) and for those who have breast implants. MRIs are not always able to determine types of breast cancer and this may require further testing through biopsies. Moreover, error and "false positives" can be the outcome for breast cancer patients who use MRIs. MRIs, however, are now more advanced and can give better, more accurate results with AI (Johns Hopkins Medicine2).

Thermography for Breast Cancer: Artificial Neural Networks

Thermography is another form of imaging that can be used to detect breast cancer. It involves the use of temperature sensing as it "records temperature changes on the surface of the skin which is known as Digital Infrared Thermal Imaging (DITI)." Versus mammography and MRI thermography is a less invasive, "painless and cost-effective" form of imaging that offers a "high chance of recovery" for patients when detection is early. It is argued that no form of imaging is as predictable as biopsies, but today enabling technologies with AI and artifi-

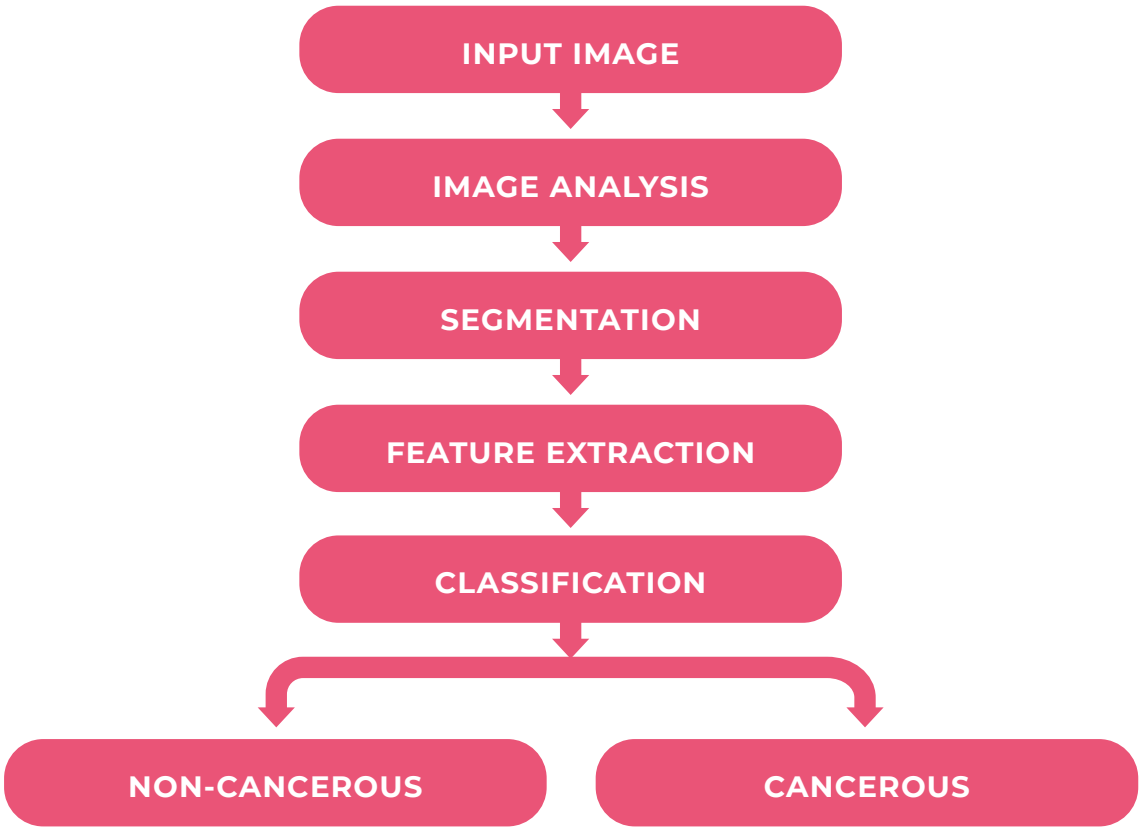


Image 1-The Effect of Thermography on Breast Cancer[xv]

cial neural networks (ANNs) can offer the tools for increased predictability that can also enhance thermography imaging. ANNs are a machine learning technique that can "detect trends and extract complex patterns." Many imaging machines are dependent on "classification algorithms'' that aid in determining whether or not a patient needs further testing done. Aside from ANNs, machine learning algorithms for imaging classification are "Naïve Bayes Classifier Algorithm, K Means Clustering Algorithm, Support Vector Machine Algorithm, A priori Algorithm, Linear Regression, Logistic Regression...Forests, Decision Trees and Nearest Neighbors." Yet many of these algorithms cannot be utilized for imaging processes. Not all classifiers are suitable for all image processing techniques and all have their own advantages and disadvantages.[xiv]

3-Detection and Risk Prediction

Breast cancer needs to be detected earlier in order to cure the disease. For many women around the world, the various types of imaging offered are either dependent on the detection of larger masses of cancerous cells or tumors or they are problematic in terms of breast size and detection of abnormalities within the breast tissue itself. In this arena, artificial intelligence holds much potential. Dr. Constance Lehman, Harvard University Medical School, finds that current cancer risk models are inadequate and he, as head of Breast Imaging at Massachusetts General Hospital, wants to apply more AI to "improve prediction of future risk of breast cancer based on mammography alone, including validation trials in multi-ethnic populations" and he wants this to be the future of breast cancer imaging applications so that patients can have a more "personalized" and "precise" way to detect, diagnose and treat breast cancer.[xvi]

4- Ethics of Artificial Intelligence and Cancer

With the development of any advancement, ethics is always an issue that is debated. In areas of medicine, to be particular, breast cancer treatment, screening and diagnosis bring along "risk calculation, prognostication and clinical decision-support, management planning, and precision medicine." Artificial intelligence (AI) systems have caused immense excitement with their rapid development as the care for breast cancer greatly increases. With this, implementations also naturally seem to happen in rushed manners. The ethics aspect does not only require accuracy from AI systems because "accuracy alone cannot justify clinical use." These systems also have to be evaluated in regards to their legality and their position according to social as well as ethical criteria. With the review of multiple areas of implications, the values encoded in algorithms of the systems are evaluated regarding their outcomes, any kinds of biases, consensuality, morality, and responsibility.

Stakeholders in Healthcare AI

Health systems adopting AI should do so in rigorous trials with public deliberation. We consider potential effects for patients, including on trust in healthcare, and provide some social science explanations for the apparent rush to implement AI solutions. We conclude by anticipating future directions for AI in breast cancer care. Stakeholders in healthcare AI should acknowledge that their enterprise is an ethical, legal and social challenge, not just a technical

challenge. Taking these challenges seriously will require broad engagement, imposition of conditions on implementation, and pre-emptive systems of oversight to ensure that development does not run ahead of evaluation and deliberation. Once artificial intelligence becomes institutionalized, it may be difficult to reverse: a proactive role for government, regulators and professional groups will help ensure introduction in robust research contexts, and the development of a sound evidence base regarding real-world effectiveness. Detailed public discussion is required to consider what kind of AI is acceptable rather than simply accepting what is offered, thus optimizing outcomes for health systems, professionals, society, and those receiving care.[xvii]

Conclusion

Breast cancer is not a novel disease it is one that has plagued women all over the world decades if not longer. Medical technologies, physicians and researchers as well as the larger public have been moving to end the deaths of women due to this detectable and treatable illness. The problem, however, with all the funds the world has put in for research of breast cancer has, though saved many women's' lives worldwide, has not improved as much as we would like nor at the pace of the disease itself. As diseases evolve so must our technologies and the ways we detect, diagnose and treat diseases like breast cancer. Artificial intelligence, favorable for its data imaging and predictability of data which is fast, accurate and efficient, offers clinicians a way to detect and learn more about cancerous tumors and lesions not only in the breast but commonly for the other types of cancer that can be detected better with AI such as prostate, brain and lung cancer. Using algorithm enhanced AI machine applications, MRIs, mammograms and ultrasounds are currently being trained with AI to improve data outcomes for the health of women everywhere. It is hoped that the ethical concerns that revolve around the use of AI for the detection, diagnosis and treatment of breast cancer do not outweigh the future benefits and issues of privacy and security will be addressed and resolved by governments and medical institutions so that AI can improve health and well-being without being limited and so that researchers can offer less invasive, painful and more precise and accurate information for patient health.

Endnotes

i Harvard University.
ii Sadoughi.
iii Stanford University.
iv MIT.
v See endnote ii.
vi Drukker et al.
vii Ibid.
viii Huang et al.
ix Johns Hopkins University.
x Johns Hopkins Medicine.
xi Towards Data Science.
xii Penn Medicine.
xiii Ibid.
xiv Pavithra et al.
xv Ibid.
xvi Harvard Medical School.
xvii Carter et al.

References

[1] AI and the Future of Breast Cancer Detection and Risk ... (n.d.). Retrieved September 27, 2021, from https://executiveeducation.hms.harvard.edu/thought-leadership/webinar-series/recorded-presentations/ai-future-breast-cancer-detection-risk-prediction

[2] Breast Magnetic Resonance Imaging (MRI). (n.d.). Retrieved September 15, 2021, from https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/breast-mri

[3] Breast Ultrasound. (n.d.). Retrieved September 10, 2021, from https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/breast-ultrasound

[4] Brown, S. (2021, April 21). Machine learning, explained. Retrieved September 25, 2021, from https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained#:~:text=Machine learning is a subfield of artificial intelligence, which is, to how humans solve problems

[5] Carter, S. M., Rogers, W., Win, K. T., Frazer, H., Richards, B., & Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. The Breast, 49, 25-32.

[6] Drukker, L., Noble, J. A., & Papageorghiou, A. T. (2021). Introduction to Artificial Intelligence in Ultrasound Imaging in Obstetrics and Gynecology. Obstetrical & Gynecological Survey, 76(3), 127-129.

[7] 'Fast' MRI Detects Breast Cancers that 3-D Mammograms May Miss. (n.d.). Retrieved September 27, 2021, from https://www.pennmedicine.org/news/news-releases/2020/october/fast-mri-detects-breast-cancers-that-3d-mammograms-may-miss

[8] Harvard Health. (n.d.). Retrieved September 10, 2021, from https://www.health.harvard.edu/topics/breast-cancer.

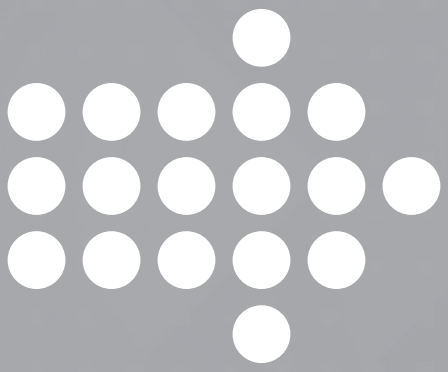
[9] Huang, Y. L., Kuo, S. J., Chang, C. S., Liu, Y. K., Moon, W. K., & Chen, D. R. (2005). Image retrieval with principal component analysis for breast cancer diagnosis on various ultrasonic systems. Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology, 26(5), 558-566.

[10] Holländer, B. (2020, July 21). How Medical Ultrasound May Become the Preferred Imaging Method with AI. Retrieved September 15, 2021, from https://towardsdatascience.com/how-medical-ultrasound-may-become-the-preferred-imaging-method-with-ai-323a689a1997

[11] Lei, Y., Yin, M., Yu, M., Yu, J., Zeng, S., Lv, W., . . . Dietrich, C. F. (0001, January 01). Artificial Intelligence in Medical Imaging of the Breast. Retrieved September 27, 2021, from https://www.frontiersin.org/articles/10.3389/fonc.2021.600557/full

[12] Myers, A., & Myers, A. (2011, October 25). Stanford's John McCarthy, seminal figure of artificial intelligence, dies at 84. Retrieved September 15, 2021, from https://news.stanford.edu/news/2011/october/john-mccarthy-obit-102511.html

[13] Pavithra, P. R., Ravichandran, K. S., Sekar, K. R., & Manikandan, R. (2018). The effect of thermography on breast cancer detection. Systematic Reviews in Pharmacy, 9(1), 10-16.



Machine Learning and Bioinformatics

Batu Engin

Introduction

Biology, the study of life, is one of the most complex and intriguing sciences known to humankind. For our knowledge of Biology to expand, scientists have to analyze and study large amounts of data for the smallest objects such as DNA strands, proteins, and genes. This massive amount of data is not easy to analyse and interpret without the use of computers. As a result, an interdisciplinary field called Bioinformatics, or the combination of Data Science and Biology, is beneficial. Obstacles in analysis and interpretation can be efficiently tackled by applications of machine learning. Ever since the term “machine learning” (ML) was coined, it has been a hot topic in many industries such as

in the areas of industrial production, finance or healthcare. The main reasons for the rise in ML's applications in Bioinformatics is because scientists are aiming to increase their efficiency and quality when analyzing and interpreting biological data in research for gene finding, sequence alignment, protein structure prediction, drug discovery and design, and simulations. Thus raising the question as to what extent can Artificial Intelligence (AI) and ML, specifically, benefit Bioinformatics? This paper will specifically evaluate the benefits of ML to the aforementioned applications within Bioinformatics.

Machine Learning

Machine learning is a subfield of Artificial Intelligence and Computer Science (See Figure-1) in which data and different algorithms that mostly rely on statistics are used in order to find patterns in data, achieve human-like learning capabilities and ultimately attempt to increase their accuracy through feedback loop systems (IBM, n.d.).

There are two predominantly used types of machine learning: supervised learning and unsupervised learning. Supervised learning can be thought of as an interventionist approach by humans in which models receive already labelled data to predict the outcome of a certain situation. The most well-known and basic example being: a model receives a dataset full of house features such as size, amount of bedrooms and location; and this data is labelled with the price of the houses. The ultimate goal for this model would

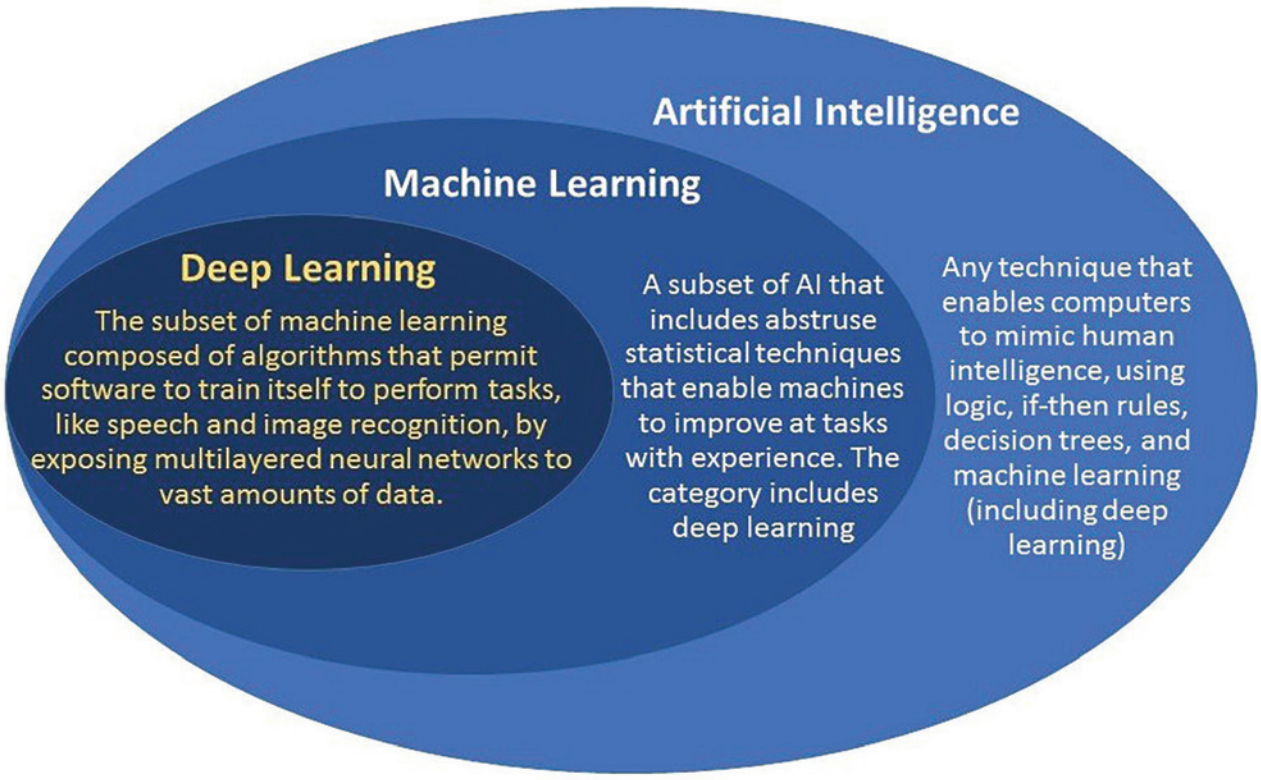


Figure 1.
Note. From What is the difference between AI, machine learning and deep learning? by M. Dhande, 2020 (https://www.geospatialworld.net/blogs/difference-between-ai%EF%BB%BF-machine-learning-and-deep-learning/). Copyright by GEOSPATIAL MEDIA AND COMMUNICATIONS.

be to predict the price of a house given their features. On the other hand, unsupervised learning is the opposite of supervised learning in which models are tasked with finding the patterns between different data and labels themselves. One example of this heavily relates to bioinformatics: given a collection of millions of genes, the model would be tasked with categorizing these genes into what their functions and locations would be (Ng, n.d.).

Foote's (2019) history of Artificial Intelligence reveals that the first machine learning application was an AI for playing checkers, implemented by Arthur Samuel, who was working at IBM, in the 1950s. The object of this machine learning algorithm was to find the best possible outcome and play accordingly, known as the process of alpha-beta pruning. Samuel himself coined the term "machine learning" in 1952. Furthermore, the first unsupervised algorithm called "The Nearest Neighbor Algorithm" was conceived in 1967 as an attempt to solve travel route problems for people such as salespeople. It was not until the early 1980s that ML was separated into its own distinct field due to the differences between its goals and those of AI. Finally, Foote also explains that ML shifted its focus to providing services via its statistical methods, such as regression, and focused on conducting research on Neural Networks which is still one of the most recognized machine learning methods today.

Throughout its history, machine learning has achieved and supported many discoveries from recognizing images of cats on YouTube to supporting self-driving cars; yet there is so much more to discover. Machine learning applications have countless benefits to multiple industries. Some of these include: identification of trends and patterns, mostly simple to implement and use, little to none human intervention needed as models can learn themselves to boost their accuracy and performance. For example, one of the applications of ML is in finance; ML can be used to predict which stock may be worth buying, recognize anomalies or detect people eligible for loans. Another example is in transportation, one of the most recognized applications being self-driving cars which can use object classification algorithms to identify objects such as roads, cars, pedestrians in order to guide them through the route and reach their designated destination (Mindy Support, n.d.).

Bioinformatics

According to Christopher P. Austin, M.D. from the National Human Genome Research Institute, "Bioinformatics is a field of computational science that has to do with the analysis of sequences of biological molecules." (National Human Genome Research Institute, n.d.)

Gauthier's (2018) "A brief history of bioinformatics" informs us that bioinformatics is almost as old as machine learning even though some may think of it as a recently established field. Margaret Dayhoff who was originally a physical chemist "who pioneered the application of computational methods to the field of biochemistry" is regarded as the "first bioinformatician". During the years 1958 to 1962, she and a physicist, Robert S. Ledley, implemented a computer program called COMPROTEIN for the IBM 7090 which can be seen in Figure-2.

This program, according to Gauthier, was used to "determine protein primary structure using Edman peptide sequencing data". Furthermore, ac-

Figure 2.
Note. From IBM 7090 Data Processing System,, n.d. (https://www.computerhistory.org/collections/catalog/102646612). Copyright 2021 Computer History Museum.

cording to Gauthier, two scientists, Emile Zuckerkandl and Linus Pauling, focused on investigating biomolecular sequences as 'carriers of information' whereas most research focused on the mechanistic modeling of enzymes. 64 codons of the genetic code were deciphered by 1968, paving a way for sequencing DNA. The first method of DNA sequencing was the "Maxam-Gilbert sequencing method" developed in 1976. In 1978, the first "probabilistic model of amino acid substitutions" was developed by Dayhoff and two other scientists. Finally in the 1980s-1990s, computer science and bioinformatics started to come closer together amid the popularization of commercial computers. With more access to computers by consumers, the rate of development for software for bioinformatics and biology increased substantially.

The benefit of data science could be seen evident when the Human Genome Project (HGP) was initiated by the National Institutes of Health (NIH). HGP was "the international research effort to determine the DNA sequence of the entire human genome" (National Human Genome Research Institute, n.d.). Whilst the sequencing of the human genome would be much faster and cheaper nowadays, at that time

it took a "rough minimum of about 40000 runs" in order to achieve only the first step of processing the data (Gauthier). Therefore Perl (a high-level programming language) based software was created to address the massive chunk of data. In the past, many scientists such as Dayhoff made major strides towards expanding the field of bioinformatics which was ultimately sped up by the introduction of computers entering the consumer market, new statistical models and initiatives such as the HGP.

In the present, however, as Janice Glasgow, Igor Jurisica, and Burkhard Rost (2004) argue, bioinformatics has become a "truly interdisciplinary field" meaning that it is created from many other branches of science. Due to this feature of the field, a lot of different names are used for it such as "theoretical biology, biocomputing or computational biology." Whilst this might be confusing, I believe that it is a testament to the fact that when two or more branches of knowledge are combined, the outcome will be more powerful and efficient.

How does ML benefit Bioinformatics?

We have gone through the history of how computer science and machine learning was applied by bioinformatics but through which techniques and applications does it specifically benefit bioinformatics? First of all, the word "technique" will be used to refer to different machine learning algorithms whereas the word "application" will be referring to the way in which machine learning techniques are applied to different bioinformatics methods.

There are many machine learning algorithms used within bioinformatics however the most common 3 are: "K-Nearest Neighbours", "Decision Trees" and "Regression". K-Nearest Neighbours (KNN) is a supervised machine learning algorithm that can be used to solve classification and regression problems. KNN works by the principle that "similar things exist in close proximity". The "K" in the KNN refers to the parameter which decides how many neighbours should be considered when deciding the output of a given input. For example, as seen in Figure 3, we have two classes, A and B. A "K" value of 3 would include 3 neighbours, 1 of class A and 2 of class B; on the other hand, A "K" value of 6 would include 6 neighbours, 4 of class A and 2 of class B. Based on this information, the k-nearest neighbours algorithm would evaluate the class in which the red star would fit (Harrison, 2019).

Furthermore another machine learning algorithm utilized would be decision trees. Decision trees are also supervised algorithms and are split into two types: classification trees and regression trees. In classification trees, given a certain input, the output would be binary. For example, classifying if a person is healthy or unhealthy. However regression trees output a "continuous variable", e.g an integer value or a percentage. For instance, regression trees could be used to determine the age of a student based on their habits such as if they exercise or eat properly. Whilst it could be easily implemented by a high level programming language such as Python with minimal lines of code, the inner workings of decision trees are complex. However, essentially as George Seif puts it, the objective of the decision tree algorithm is "to build a tree with a set of hierarchical decisions which eventually give us a final result, i.e our classification or regression prediction" (George Seif,

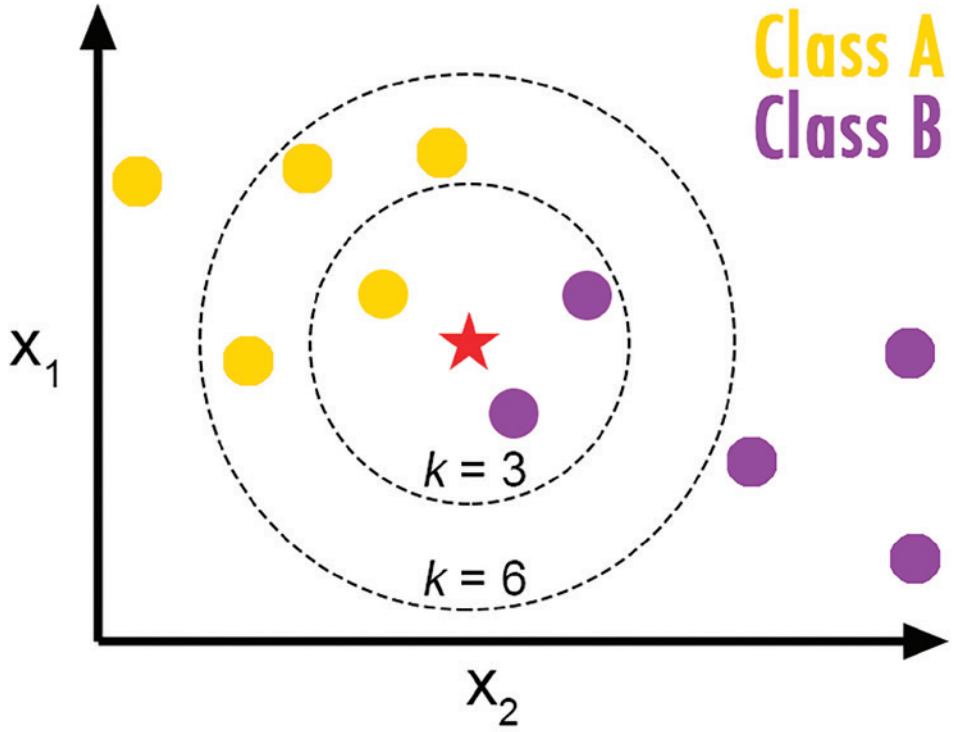


Figure 3.
Note. From Introducción al Machine Learning #9 - K Vecinos más cercanos (Clasificación y Regresión), by L. Salcedo, 2020 (https://pythondiario.com/2018/01/introduccion-al-machine-learning-9-k.html).

2018). Here are the simplified steps of training a decision tree model as written in the article:

> "1- Begin with your training dataset, which should have some feature variables and classification or regression output. 2- Determine the "best feature" in the dataset to split the data on; more on how we define "best feature" later. 3- Split the data into subsets that contain the possible values for this best feature. This splitting basically defines a node on the tree i.e each node is a splitting point based on a certain feature from our data. 4- Recursively generate new tree nodes by using the subset of data created from step 3 We keep splitting until we reach a point where we have optimised, by some measure, maximum accuracy while minimising the number of splits / nodes" (George Seif, 2018).

Lastly, regression is the most widely used machine learning algorithm not only in bioinformatics but also in finance, economics, sciences and other fields. For example the Capital Asset Pricing Model (CAPM) is a linear regression model that predicts the return of a portfolio with respect to its risk (Corporate Finance Institute, 2020). For the sake of simplicity, I will be talking primarily throughout this paper about the simple linear regression model that can be represented by:

$$Y = a\_0 + a\_1x$$

This type of regression analyses the linear relationship between the independent value (x) and the dependent value (Y). The "a_0" value refers to the initial value of Y or rather value of Y when x

is equal to 0. The "a_1" value refers to the gradient of the function. Understanding the formula is quite simple, however the complexities of linear regression lie in the cost function. As Gandhi (2018) states in the article, "The cost function helps us to figure out the best possible values for a_0 and a_1 which would provide the best fit line for the data points." In data such as the one demonstrated in the graph (Figure-4), what the cost function does is minimize the distance between the line and the points so we can predict more accurate values. Last concept to better understand linear regression is the gradient descent, which explains how exactly the a_o and a_1 values are determined. Simply gradient descent works by picking initial values of a_0 and a_1(these values can be both high or low). In an iterative process: values are changed, cost function is checked and at the end, values of a_0 and a_1 that have the lowest cost function are chosen for linear regression.

These machine learning techniques have been utilized in many bioinformatic methods such as drug discovery, gene finding, sequence alignment, drug design, protein structure prediction, protein structure alignment, evolution modelling and cell division modelling.

Gene Finding and Prediction

According to the NIH (n.d), gene finding is the process of "identifying genes within a long DNA sequence". Before the field of bioinformatics was created, gene finding was done by complicated processes such as studying its functions through living organisms or analysing it through test tubes. Indubitably, it could be sensed that this is not the most efficient way of analysing and identifying genes. Gene prediction is especially important as it allows for

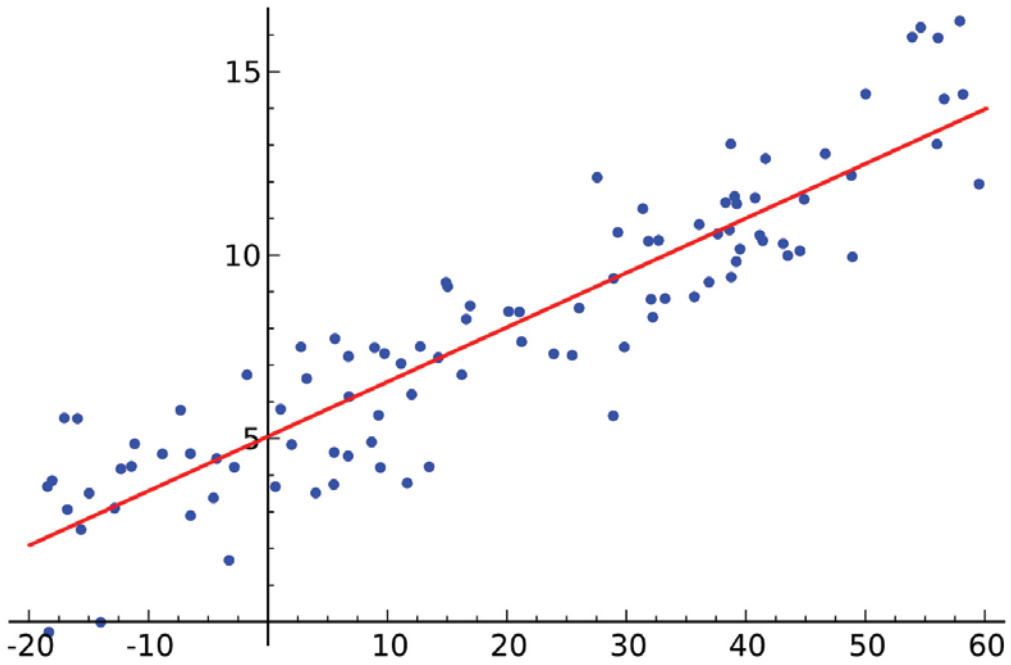


Figure 4.
Note. From Introduction to Machine Learning Algorithms: Linear Regression by R. Gandhi, 2018, (https://www.dataversity.net/a-brief-history-of-machine-learning/). Copyright by 2011 – 2021 Dataversity Digital LLC.

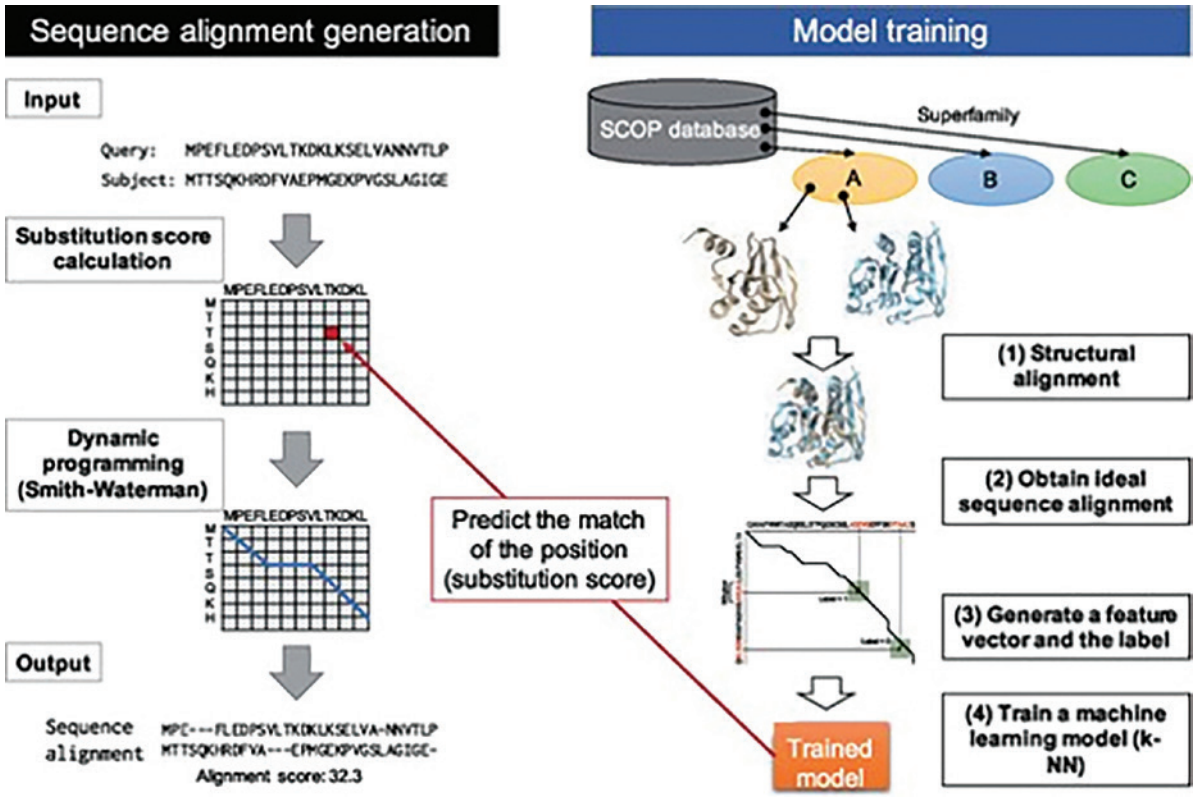
"understanding the minimal requirements of an organism, identifying disease genes, and finding new drug targets." (Xia, 2018)
To begin with, gene prediction has two types: ab initio and evidence-based. Ab initio is the process in which a computer program is fed data and instructions for finding genes. The computer tries to locate common sequences at the start and end of genes. Evidence-based gene finding is as follows in the article:

> "It involves gathering various pieces of genetic information from the transcript sequence and known protein sequences of the genome. With these pieces of evidence it is then possible to get an idea of the original DNA sequence by working backwards through transcription and translation." (How do you identify the genes in a genome? 2021)

The information from both techniques are then "combined and lined up with the sequence genome".
In Elizabeth H.Mahood's (2020) paper, the applications and benefits of machine learning for gene finding is demonstrated. She states that for gene function prediction, supervised algorithms are used because, in order to locate the function of a gene, data must be fed for the algorithm to find a pattern . For instance, a software used for the identification of "genic elements" utilizes Support Vector Machine, which is a supervised machine learning algorithm". Furthermore, in one study attempting to "predict metabolic pathways in an organism from its genome-wide gene complement", decision trees were used.
However an article by Zhang (2020) reveals that applications of machine learning may have a few limitations. Instead, he proposes a deep learning solution, explaining that machine learning's "prediction powers either alone or in integration mode (in machine learning methods) are still limited compared with those automatically learned by some deep learning frameworks".

Sequencing

Andrew D. Prjibelski (2019) reveals through his article that, "Sequence alignment is the process of comparing and detecting similarities between biological sequences" It is about finding evolutionary relationships through similarity functions when DNA sequences are aligned.
When comparing two strings, similarity scoring algorithms such as Hamming distance or Cosine difference may be used, however the Hamming distance is the most frequently used one. Macleod (1993) provides us the definition, "The Hamming distance between two codewords is simply the number of bit positions in which they differ."
Machine learning is beneficial in sequencing as past methods of sequence alignment such as similarity scoring algorithms do not provide "sufficiently accurate structure models" as Shuichiro Makigaki (2019) notes. As a high accuracy of sequence alignment is significant for modeling, another method is required. In his article, Makigaki proposes the training of a KNN model to boost the accuracy of sequence alignment. He uses the process of "re-alignment" which is basically, receiving an input of two sequences and outputting two sequences that are more suitable for "homology modeling". The KNN is trained on the SCOP database which stores all classification for protein structure. The objective of the machine learning model is to predict the match score for the two sequences. This match score is



then used for re-aligning the sequences creating a suitable input for "template-based modeling". The illustrated steps can be seen in Figure 4 below.

The application of machine learning was particularly successful as this new developed method outperformed the traditional distance functions being used before. On the other hand, there is a trade-off between accuracy and time. KNN is a relatively slow algorithm. Training a KNN model on a dataset such as the SCOP is time consuming. Therefore Makigaki was forced to reduce the amount of training data. Whilst machine learning can most certainly increase performance, it is also quite slow.

Drug Design and Discovery

During the Covid-19 crisis, we saw first-hand the need for rapid drug development. Bioinformatics has, for long, contributed to drug design and discovery for many diseases. As Xuhua Xia (2017) has stated, "Bioinformatic analysis can not only accelerate drug target identification and drug candidate screening and refinement, but also facilitate characterization of side effects and predict drug resistance." Bioinformatics aids in all the steps for drug discovery. For example, connecting disease symptoms to genetic and environmental factors, identifying the target of the drug, which could either be eliminating the harmful cells or restoring the cells' functions, or predicting the effects that a certain drug can have on the test candidates.

The contributions of machine learning to drug discovery spans, "predicting target structure", "identifying and optimizing hits", "exploring the biological activity of new ligands" and "designing models that predict the pharma-

cokinetic and toxicological properties of the drug candidates" (Ratanghayra, 2021). One recent case of the predicting function is the DeepCE project engineered by researchers at the Ohio State University. It is a neural network that is trained with L1000 which "is a National Institutes of Health-funded data repository" and DrugBank which contains a heavy load of chemical information for more than 10000 drugs. Evidently, processing this data with high accuracy requires machine learning. DeepCE explains how gene expression correlates with drug response which particularly helps with identifying drug repurposing candidates. This is especially useful when a drug is being developed for a new and unknown disease such as Covid-19. Another use of machine learning is finding patterns in signals from "image-based profiles". Ratanghayra explains that image-based profiling is the process of extracting and analysing biological images. A group of researchers and biologists lead by Anne Carpenter is also using deep learning to identify patterns in image-based profiles which can provide a better understanding of biological information and hence accelerate drug discovery.

The drawbacks of applying machine learning to drug development is that it cannot completely replace a human yet. As the editorial team (2021) argues, machine learning predictions can be harmed by the algorithm bias and the cost of the computer power required for deep learning and fast training of millions of data. Due to the fact that an algorithm cannot be fully trusted in important subjects such as developing drugs that could potentially save the lives of millions of people, scientists still need to verify the validity of predictions which can slow down the process.

The Future of Machine Learning for Bioinformatics

Throughout the article, there is a recurring theme of AI having its advantages but also disadvantages mostly due to its limitations of time. Whilst linear machine learning algorithms can be trained on a relatively small dataset with commercial computers, in the field of Bioinformatics, data is very large. As years pass, we are gathering more and more data and the processing of such data is complicated. This is why Quantum computing is one of the most anticipated and promising technologies.

As IBM (n.d) puts it, "Quantum computing harnesses the phenomena of quantum mechanics to deliver a huge leap forward in computation to solve certain problems." Quantum computers are built upon the principles of quantum mechanics. Instead of bits in normal computers, quantum computers use "qubits" which are the basic unit of information stored in the quantum state. The reason why quantum computers can analyze data is due to the vast 3 dimensional spaces it can create via the principle of superposition which is the ability of particles to be at multiple states simultaneously. Furthermore quantum algorithms can "exploit quantum entanglement" which allows randomly behaving qubits to be "perfectly correlated with each other". The ability of quantum computers to make mathematical calculations efficiently at the smallest possible levels allows for machine learning algorithms to perform faster. I strongly believe that quantum computing will positively impact the field of bioinformatics once it is fully integrated into bioinformatic methods, most importantly in drug discovery.

Conclusion

The increase in application in machine learning has long impacted the interdisciplinary field of bioinformatics. Throughout my paper, it is clear that these applications had a positive impact due to increased efficiency levels when it comes to processing the data and producing outputs. With the use of both machine learning and deep learning, methods of bioinformatics such as sequencing, gene finding and drug discovery have been accelerated. Whilst there are time constraints to the use of these statistical models, in the future, as quantum computing is more spread, this certainly will not be an issue. I believe that if we continue to develop and utilize machine learning solutions for bioinformatic applications, the world can further benefit from the biological and research advances.

Bibliography

Corporate Finance Institute. (2020, April 13). Capital Asset Pricing Model (CAPM). https://corporatefinanceinstitute.com/resources/knowledge/finance/what-is-capm-formula/

Editorial. (2021, April 12). AI in drug discovery - Benefits, drawbacks, and challenges. RoboticsBiz. https://roboticsbiz.com/ai-in-drug-discovery-benefits-drawback-and-challenges/

Fisher, C. (n.d.). What is Quantum Computing? IBM Quantum. https://www.ibm.com/quantum-computing/what-is-quantum-computing/

Foote, K. D. (2019, March 13). A Brief History of Machine Learning. DATAVERSITY. https://www.dataversity.net/a-brief-history-of-machine-learning/

Gandhi, R. (2018, June 21). Introduction to Machine Learning Algorithms: Linear Regression. Medium. https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a

Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2018). A brief history of bioinformatics. Briefings in Bioinformatics, 20(6). https://doi.org/10.1093/bib/bby063

Glasglow, J., Jurisica, I., & Rost, B. (2004). AI and Bioinformatics. AI and Bioinformatics. Published.

Harrison, O. (2019, July 14). Machine Learning Basics with the K-Nearest Neighbors Algorithm. Medium. https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761

How do you identify the genes in a genome? (2021, July 21). Yourgenome. https://www.yourgenome.org/facts/how-do-you-identify-the-genes-in-a-genome

IBM. (2021, August 12). Machine Learning. https://www.ibm.com/cloud/learn/machine-learning#toc-machine-le-K7VszOk6

Mahood, E. H., Kruse, L. H., & Moghe, G. D. (2020). Machine learning: A powerful tool for gene function prediction in plants. Applications in Plant Sciences, 8(7). https://doi.org/10.1002/aps3.11376

Makigaki, S., & Ishida, T. (2020). Sequence Alignment Using Machine Learning for Accurate Template-based Protein Structure Prediction. BIO-PROTOCOL, 10(9). https://doi.org/10.21769/bioprotoc.3600

Mazda, F. (1993). Telecommunications Engineer's Reference Book. Focal Pr.

Mindy Support. (n.d.). How Machine Learning in Automotive Makes Self-Driving Cars a Reality. https://mindy-support.com/news-post/how-machine-learning-in-automotive-makes-self-driving-cars-a-reality/

National Human Genome Research Institute. (n.d.-a). Bioinformatics. Genome.Gov. https://www.genome.gov/genetics-glossary/Bioinformatics

National Human Genome Research Institute. (n.d.-b). Bioinformatics: Finding Genes. Genome.Gov. https://www.genome.gov/25020001/online-education-kit-bioinformatics-finding-genes

National Human Genome Research Institute. (n.d.-c). The Human Genome Project. Genome.Gov. https://www.genome.gov/human-genome-project#:%7E:text=The%20Human%20Genome%20Project%20was,than%20the%20original%20estimated%20budget

Ng, A. (2011). Machine Learning. Coursera. https://www.coursera.org/learn/machine-learning/home/welcome

Ranganathan, S., Gribskov, M., Schönbach, C., & Nakai, K. (2019). Encyclopedia of Bioinformatics and Computational Biology. Elsevier Gezondheidszorg.

Ratanghayra, N. M. (2021, June 24). Automating Drug Discovery With Machine Learning. Drug Discovery from Technology Networks. https://www.technologynetworks.com/drug-discovery/articles/automating-drug-discovery-with-machine-learning-347763

Seif, G. (2021, July 18). A Guide to Decision Trees for Machine Learning and Data Science. Medium. https://towardsdatascience.com/a-guide-to-decision-trees-for-machine-learning-and-data-science-fe2607241956

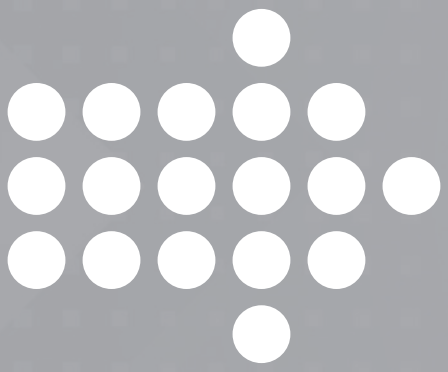
Xia, X. (2018). Bioinformatics and Drug Discovery. Curr Top Med Chem. Published.

Zhang, X., Xiao, W., & Xiao, W. (2020). DeepHE: Accurately predicting human essential genes based on deep learning. PLOS Computational Biology, 16(9). https://doi.org/10.1371/journal.pcbi.1008229

Dhande, M. (2020, July 3). What is the difference between AI, machine learning and deep learning? Geospatial World. https://www.geospatialworld.net/blogs/difference-between-ai%EF%BB%BF-machine-learning-and-deep-learning/

IBM 7090 Data Processing System (n.d.). Computer History. https://www.computerhistory.org/collections/catalog/102646612

Salcedo, L. (2020, December 22). Introducción al Machine Learning #9 - K Vecinos más cercanos (Clasificación y Regresión). Mi Diario Python. https://pythondiario.com/2018/01/introduccion-al-machine-learning-9-k.html



Computational Biology: Introduction, Methods, and a Means of Treating Cancer Patients

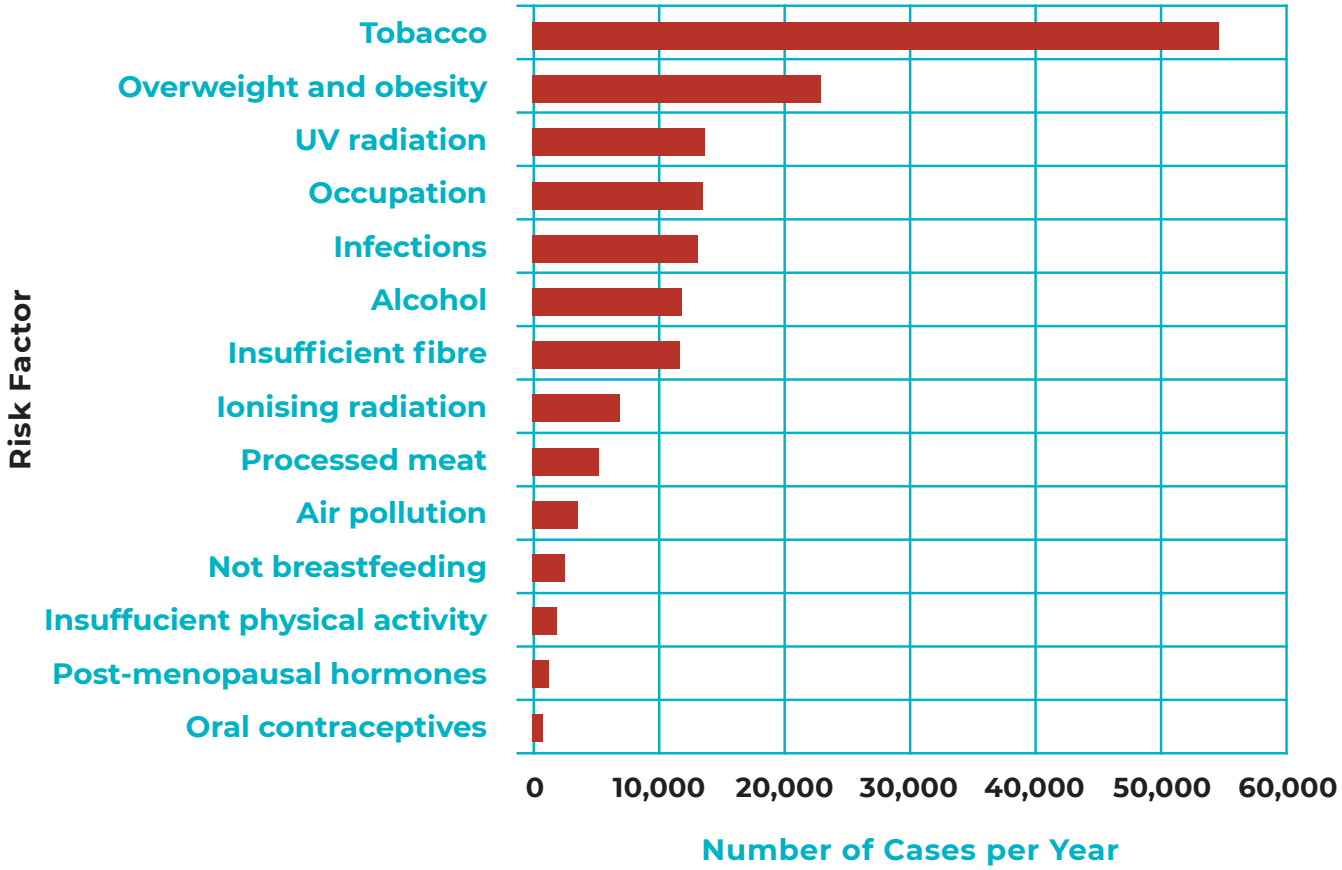
Halit Mert Bİldirici

https://orcid.org/0000-0003-0047-9229

What is Computational Biology?

It is, in the words of computational biologist Florian Markovetz, actually a manifestation of all biology, past and present. How so? He claims that, because computational biology makes use of technology to, in essence, classify and sort various things into groups. Markovetz says that

"One key example of how computers reshaped biological research is the use of databases and ontologies. Biological knowledge today is defined, organised, and accessed through computation. If Carl von Linné (also known as Carl Linnaeus), the Swedish botanist and father of taxonomy, lived today, he would be a computational biologist."



He also makes the excellent point that much of biology in the modern sense, such as treating cancer for example, depends upon thorough quantitative, not simply qualitative, use of large-scale (and usually global or at least regional) data. For example, what is the occurrence rate of lung cancer, for example? How often is it seen in which demograph according to area of residence, which of course would be accounted for in terms of exposure to pollutants such as smog or otherwise, the cleansing effect of sunlight (if any exists) as well as a varied plentitude of other factors such as general diet, exercise and standard of general health for the area. For example, the graph below shows us preventable incidence of cancers within the UK, to give one instance of how large data greatly benefits treatment plans and other solutions within the modern frame of biology studies that encompasses medical research.

In benefiting general knowledge of how exactly computational biology aids scientists in various forms of categorization and research, looking at it from an evolutionary perspective, similar to how Linnaeus divided animals and plants into groups and families for ease of identification, so shall examining the evolutionary perspective of cancer via computational biology help grasp how the large data is used to achieve smaller-scale but significant conclusions that apply more widely to the field(s) in question. This will be followed by a practical summary of the modelling methods and their applications as used in the real world by researchers, as well as recent advances made in the field via newer uses of biological computation in an experimental environment.

Computational Biology: Cancer from an Evolutionary Perspective

Firstly, what exactly is cancer, and why does it spread through the body as it does? Simply said, cancerous tumours grow in size and spread through the body due to essentially rapid and uncontrolled cell multiplication. Progression. This progression of cells emerges because "of the accumulation of selectively advantageous mutations, and expanding clones give rise to new cell subpopulations with increasingly higher somatic fitness." This somatic evolution of cancer cells was established in the 1970s Nowell and others, and today, computation biology allows researchers to use large-scale molecular profiling data to create principles and hallmarks of tumor evolution, and to also understand how it manifests across various types of cancer as a shared trait. Firstly, although cancer is usually found at only one stage of the life of a "host", there is evidence that "although mutations are thought to primarily arise during the development of cancerous tissue, there is a growing body of evidence, including theoretical, histological, and genetic approaches, supporting the idea that somatic mutations occur throughout the entire lifetime of the host organism. Such mutations can be detected at low levels in circulating cells, as well as directly from tissue. In eyelid epidermal cells, for example, it has recently been shown that perfectly functional cells harbor a plethora of mutations that are also found in known cancer genes." Even so, a single snapshot of a genome can only say so much of the actual cancer risk in the patient.

Models and Methods in Computational Biology

Modern sequencing now allows samples from several parts of the tumour, and can also include a time-lapse for different samples to display realistic aftereffects of any treatments undertaken. Although direct sequencing of samples is now routinely carried out, mutation signals from small subsets of cells are difficult to detect. Regarding the models used to practically sort tumour development in a section of the body, "there exist four types of phylogenetic methods used in in biogeography studies (usually related to populations found within a habitat, but equally applicable here): diffusion models, island models, hierarchical vicariance, and reticulate. Excepting reticulate which either a species or individuals of a species can freely move around. Most closely resembling a tumour in that the cells move and spread just as wildlife does, it is of interest to note how relatable methods are in computational biology via use of more traditional forms of data sampling, although for tumours we may say the "habitat" is prone to bend and distort size in unpredictable ways, even jumping randomly to new points far away, making the process somewhat more complex than a simple biogeographic map.

In order to describe the population dynamics of evolving tumors, different models exist, but are differing in their specific usefulness: Population genetics models, via use of the Wright–Fisher or the Moran process, can be used to model the fate of individual cells in a population. However, such a process is usually not exceedingly useful, as tumours tend to evolve along the lines of a cluster or clusters of separate cell divisions, rather than remain as one homogenous mass, especially if and when they tend to develop resistance to efforts to treat them by altering their mutated sequences of genetic material. More generally used are instead

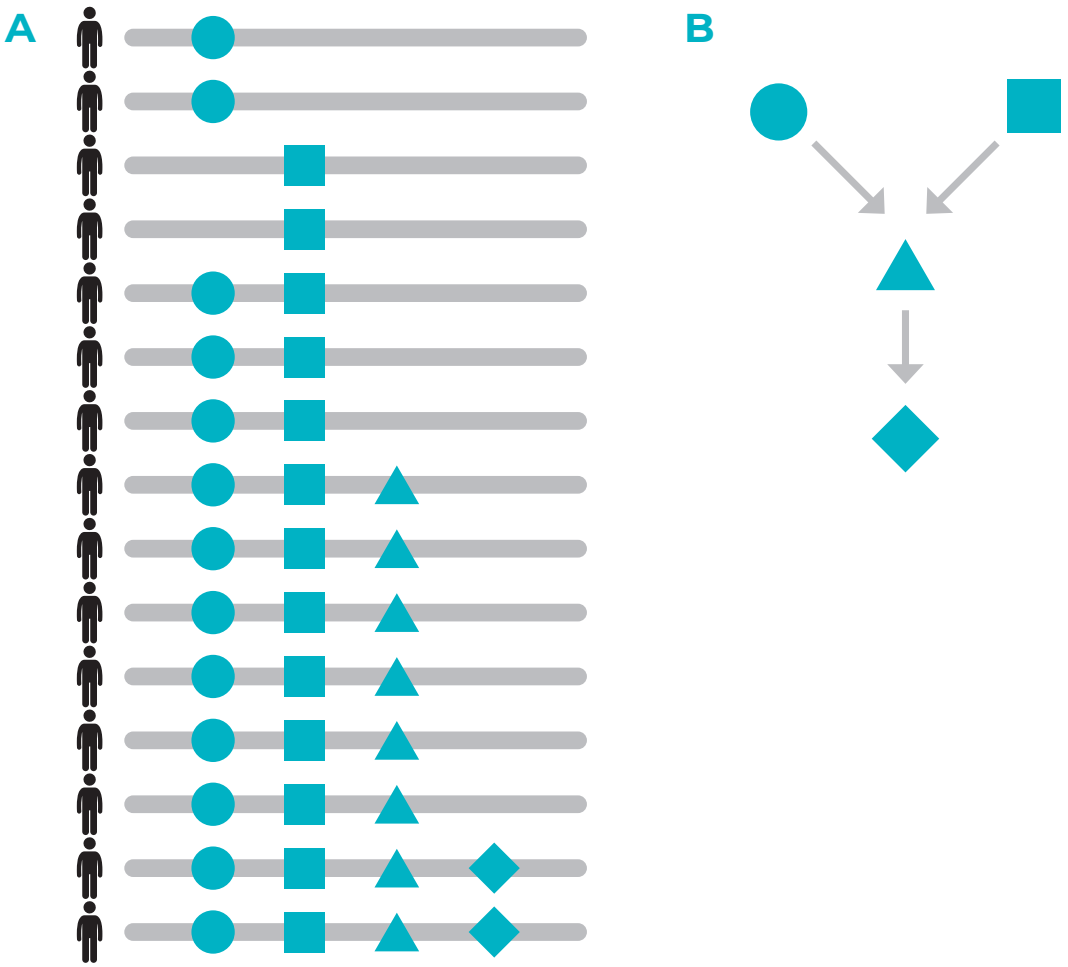


Figure 1.
Each tumor is regarded as an independent realization of the same evolutionary process

branching processes that consider the unpredictable and responsive fluctuations in the growth and composition of the population of mutated cancer cells. These deterministic approximations can often be solved analytically under simplifying assumptions, allowing for factoring in of key points of readily quantifiable interest, including the probability of and time to fixation of a mutant and the size and age of the tumor cell population. By contrast, intricate models tend to quickly collapse into unreliability due to a lack of figures on currently undeterminable variables such as such as population structure or cellular interactions (whether among the cancerous cells or within the healthy vicinity next to the tumour(s)).

Mathematical modeling of tumor cell population dynamics may one day lead to models and software tools that are potentially predictive of disease progression and treatment outcome. However, the majority of current large-scale sequencing that has taken place is of limited depth and geographical scope, and mathematical models based on such information will therefore be most likely rather approximate, not deemed as medically very useful. To construct realistic models of the evolutionary processes taking place, is the need for need for realistic outcomes of the occurring process sequence alongside high-resolution data. One possible eventual solution may become ecological 'habitat' models of the entire tumor micro-

environment, allowing a glimpse into the real time 'cage' of the cells as they live, breed and evolve. Since tumours evolve as individual specimens in each patient, extremely and eerily similar to how each animal or person has its own temperament and individuality, it becomes even more important that the aforementioned case of being able to observe the tumour in its native environment in real-time become a reality. Current models attempt to address this problem via guessing of common features found in each distinct type of tumour of similar 'species' to try to determine general behaviour that will be displayed, as seen in the chart below.

Models of Computational Biology: Advantages and Areas for Improvement

Next we will examine the specific modelling used in the computational environment, and analyze the benefits and issues of each such model as they relate to real-world treatment and progress in keeping track of incidence of cancer. To be truly useful to a biologist or physician, computational modeling should, according to Materi et al., such technology should, at minimum possess the features of being able to: 1) produce useful predictions or extrapolations that match experimental results; 2) permit data to be generated that is beyond present-day experimental capabilities; 3) allow experiments to be performed in silico to save time or cost; 4) yield non-intuitive insights into how a system or process works; 5) identify missing components, processes or functions in a system; 6) allow complex processes to be better understood or visualized and 7) facilitate the consolidation of quantitative data about a given system or process. Current processes are somewhat partially successful at points 1) and 2), and successful at points 3), 6) and 7), but wholly unable to help with the most crucial steps in a disease such as cancer: yielding non-intuitive but needed insights into how seemingly random cell divisions occur due to what triggers (number 4), and identification of unknown elements that, oftentimes, solving simply one unlocks a whole new subsection of research possibilities (number 6); a notable example of this being the infinitely replicating cells of Henrietta Lacks. The cells defy normal logic, but provided a global breakthrough in all existing cancer and mutagenic cell research after that point (and still continue to do so around the world, even today). Continuing onwards to our models of applicable computation options, we are presented with 4 models commonly used among biologists: these are ordinary differential equations (ODEs), cellular automata (DCA) and agent-based models (ABMs). We will cover in summary each method and its existing gaps in supplying knowledge (we have left out Petri nets because they are not generally used in tandem as as parts of hybrid models), followed by a brief mention of hybrid methods that combine one or more of the three main types listed here. Before delving into models, a summary of the various layers of interaction that must be scanned for have been provided, alongside a caption identifying the respective strong suit of each model-to-section match. Also, we include the cautionary information that, "Building models of complex biological processes is an iterative process that requires considerable attention to detail. Quantitatively accurate modeling requires explicit values for many variables including molecular

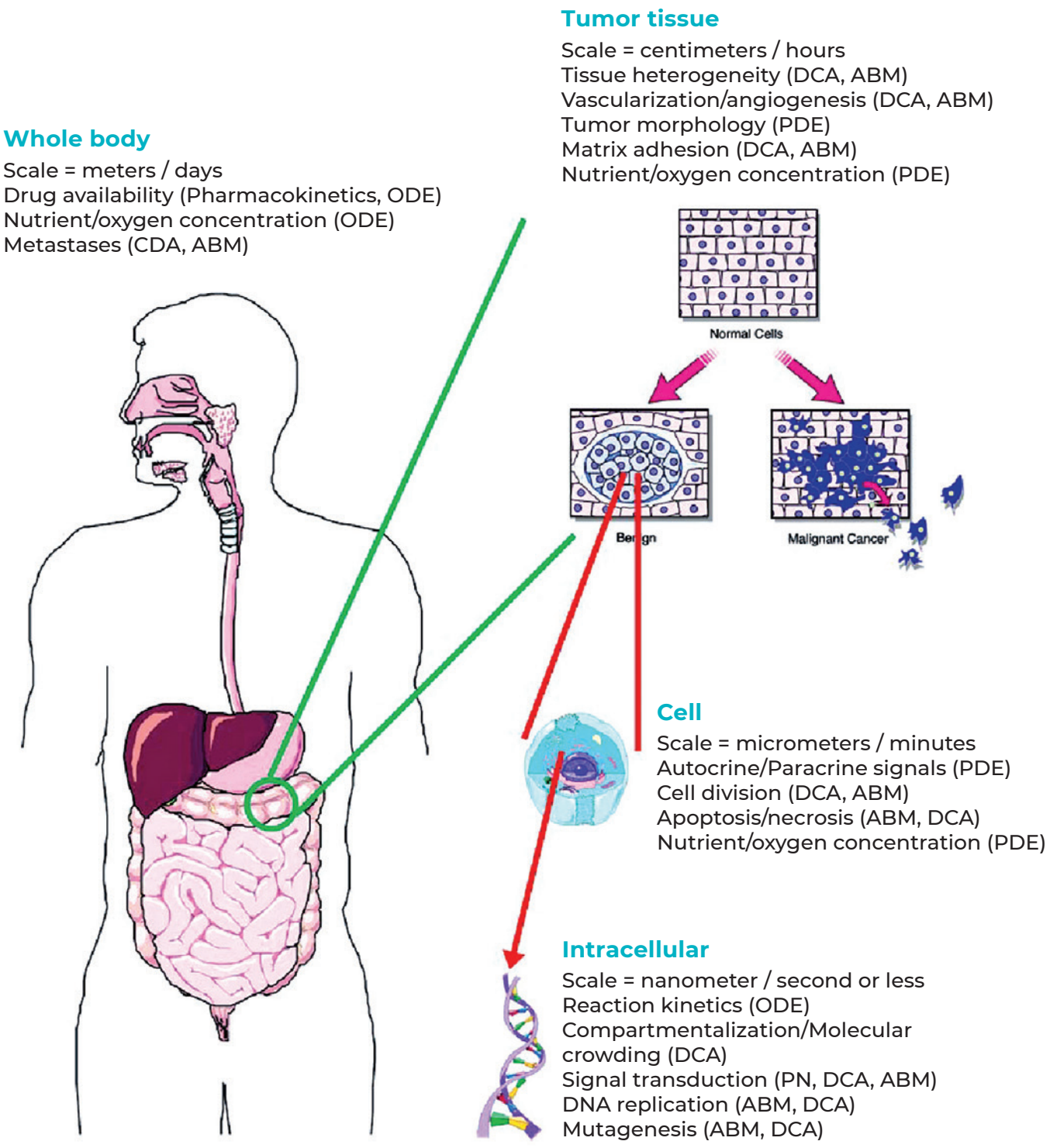


Figure 2.
Figure 1 Issues of scale in modeling cancer. From whole organism to tumor tissue to individual cells to the molecules of replication and metabolism, modeling tumors spans about nine orders of spatio-temporal magnitude. Shown above are some of the modeling issues which need to be addressed at each level of simulation. Each text box includes the relevant spatio-temporal scale and modeling issues encountered at that level. Appropriate modeling approaches to address each issue are shown in brackets. Building hierarchical systems of inter-related models is still a primary challenge to modern researchers. ODE – Ordinary differential equation system, PDE – Partial differential equation system, DCA – Dynamic cellular automaton, PN – Petri net system, ABM – Agent based model.

concentrations, cellular distribution of molecules, reactions rates, diffusion rates, transport rates and degradation rates. While many of these can be estimated from the literature or various online databases, a number of parameters often remain unknown at the start of any simulation. As a result, many modeling processes require that one provide estimates for key parameters. Usually "best guess" first order estimates can be used and then fine-tuned using a well-understood instance of the model as a comparison. Parameters are iteratively adjusted on subsequent simulations until the model accurately reflects the known test case." "...you will never find any biomarker that works in every single trial" says Professor Shirley Liu, in a talk given for thw women of Harvard University.

Ordinary Differential Equations (ODEs)

Biological systems are essentially multicomponent chemical reactors, thus displayable as chemistry equations. This fact permits mathematical analysis: Many standard biochemistry texts provide thorough derivations of ordinary differential equations (ODEs) for both simple and complex reactions. In fact, ODE based modeling is the most common simulation approach in computational systems biology, reflecting both its rigor and adaptability. However, most complex ODEs do not have exact solutions and must be solved numerically instead. Among the problems these lead to, we may briefly mention three:

ODE Issues

Reaction kinetics equations, upon which these mathematical models are formulated, assume steady-states in well-mixed solutions with abundant reactants and few enzymes. On the contrary, even simple bacteria are crowded with macromolecules, having 300 to 400 g/l of macromolecules occupying 20 to 30 percent of cytoplasmic space, compared to the idealistic and 'cleaned' 1 to 10 g/l under which normal reaction kinetics studies happen. Even explicitly modeled, diffusion rates for species of considerably different physical size are often assumed to be identical, although small molecules are much less affected by crowded conditions in regards to diffusion than large ones such as protein complexes. This crowding also impacts significantly the reaction rate, as equilibrium rates constants for reactions under crowded conditions can increase by a magnitude of two or even threefold versus dilute concentrations of macromolecules.

(Dynamic) Cellular Automata and Agent-Based Models (D)CA

An alternative used to model the complex workings of discrete molecules in living organisms is cellular automata (CA): the representation of individual molecules and the rules that govern their interactions as model for research. The tools in question are simple computer simulation tools, used to model both temporal and/or spatio-temporal processes using discrete time and/or spatial steps. (D)CA was Invented in the late 1940's by von Neumann and Ulam, who conceived of an infinite lattice of cells, each capable of a certain limited number of states of being. Each cell is connected to a finite number of neighbors whose collective states at time tn induce it to assume a new state at time tn+1 in a specified manner. In biological systems the lattice represents the equivalent of the 2D or 3D spaces found, wherein each cell can contain

one (or more) molecule or biological cell. Rules of varying complexity govern the interactions between adjacent or nearby molecules, either quite simply or instead , for example, in a complex manner, such as specifying binding of adjacent molecules with a certain probability, or such as the interactions of molecules occurring in an accurately realistic distance-dependent manner.
For these more complicated equations, Dynamic Cellular Automata (DCA) permit "Brownian-like" motion of individual molecules through the incorporation of a random number generator which selects a direction of motion in the next time step, where molecules may move one or more cells in a single time step based on the selected algorithm.

Issues

The main problem with DCA remains that although "chemical reaction rates are emergent properties of DCA models, molecular reaction probabilities derived from conventional reaction rates are usually inaccurate due to problems in deriving biologically relevant reaction rates that take into account macromolecular crowding, extremely low concentrations and compartmentalization-use of a tested model is currently perhaps an acceptable if not wholly accurate solution for the current time.

ABM

Agent Based Modeling resembles Dynamic Cellular Automata, in that ABMs possess genes, proteins, metabolites or cells can all be reactive "agents". Agents are allowed to interact with each other over space and time according to a defined set of rules. The motions may be directed or random (Brownian) and the rules may be simple or highly complex. "Unlike CA models, agent based systems do not formally require spatial grids or synchronized time steps, although practical coding considerations usually force these constraints on ABMs." Space is usually represented in a lattice-free grid.
Although somewhat different, it is easy to spot that ABMs possess more or less the same constraints as DCA models: The extremely high number of agents within a realistic cell model precludes any possibility of completely realistic output. In addition, we face similar problems of non-realistic forces of clustering of molecules, wasted three-dimensional space within the cell, and, not least, the unrealistically sparse environment of most experimental "cells" that are used as models (by containing a very minute amount of internal molecules rather than the gargantuan relative amount held in a real cell).

Hybrids

Of note is that hybrids possess one useful quality that stand-alone models do not: The use of fuzzy parameters to suggest either, roughly, measures of agents as simply "high", "medium" or "low" as a means of more accurately assessing cell function in manner more similar to the looser definitions and requirements that real living cells (and organisms as a whole) tend to use, rather than highly accurate but somewhat non-adjusting numeric situations where a small imbalance may cause a rather significant chain reaction to occur, thereby leading to a 'spread of the cancer or other hypothetical illness'.

Early Lung Cancer Detection Using Nucleus Segementation Based Features

""Lung cancer is the leading cancer killer among both men and women. Based on the statistics by the Ameri-

can Cancer Society, it is believed there are 220,000 new cases, 160,000 deaths per year and the 5-year survival rate for all stages is 15%." Of factors related to survival rates, early detection remains the most important, but usual cases of lung cancer do not display symptoms until the cancer has spread further into the body. This lowers the usual rate of early lung cancer detection to only 24%, requiring alternative methods to ensure better ways of early detection.

Of various available methods currently in use, mass screening via Computed Tomography (CT) scan of the chest area would be a sign of significant improvement over other existing methods. Unfortunately, this method cannot be recommended for general use due to the high cost and lack of safety guarantee due to radiation exposure. As an alternative to this and other costly methods, the study will examine the use of Tetrakis Carboxy Phenyl Porphine (TCPP) for early detection of lung cancer.

The samples study used 15 lung cancer patients and 13 normal patients. Cohort 1 consisted of 15 patients who had recently been diagnosed with lung cancer and had not undergone surgery or received adjuvant therapy for lung cancer. Cohort 2 included 13 subjects who were heavy smokers but did not have a history or diagnosis of lung cancer. (Heavy smoking" was defined as 20 pack-years or greater (i.e. 1 pack/day for 20 years or 2 packs /day for 10 years).

The process began via collection of three days worth of morning sputum from the lungs via the triple morning cough method. The sputum was dyed with TCPP using Biomoda CyPath® assay and slides made accordingly for holding of samples. Sample slides are observed with a fluorescent microscope to acquire images of cells. These images are then parsed via segmentation to yield pictures of single cells for further look at 71 specific features in the initial phase, such as shape, color and texture of cell. These first results obtained a fairly viable accuracy of 81%. The results were interpreted based on the characteristics of cancer cells, normal cells and necrotic cells. One discriminator used for differentiating between lung cancer and normal cells is that cancer cells glow bright red when TCPP is added.

The other main factor used in determination of cancer cells was the nucleus of the cell, which also plays a separately needed role in determination of cancer cells via its observable size and florescence-the process was added to the initial results as additive progress via using nucleus segmentation to extract features of size and the individual (nucleic) intensity from each cell. The modified feature set consists of 79 features including Nucleus size, Nucleus perimeter, Ratio b/w nucleus size and cytoplasm, mean, variance, skewness and kurtosis of intensity values of nucleus and shape parameters. The addition of nucleic factors boosted accuracy to 88% (notably higher than any existing methods) and ensures consist monitorability of treatment effectiveness, to detect the recurrence of lung cancer, and to identify patients who may need an invasive diagnostic procedure.

As seen from the results of one small sample study in detecting early cancer, the use of more cost-effective and quicker methods of biological treatments via computation is the melting of traditional treatment with newer forms of technological efficiency to ensure better clinical results and Lower mortality rates. Using the evolutionary pathways of the individual tumour

found in the patient as a guideline for which method or mixed methods to use for a certain context, allowing for a certain small margin of error to account for currently improbably complex calculations of a completely realistic cell structure, we may deem that current usage of computational biology serves an efficient role in the experimental laboratory, but requires greater application within more mainstream applications and areas of medical progress where it matters most, such as cancer, via changing attitudes that view it, not as a simple skewering of traditional biology via a computer model, but as a modern way to speed up and predict bodily process, especially of cells and other small-scale models needed in common diseases.

References

[1] Beerenwinkel, N., Greenman, C. D., & Lagergren, J. (2016). Computational cancer biology: an evolutionary perspective. PLoS computational biology, 12(2), e1004717.

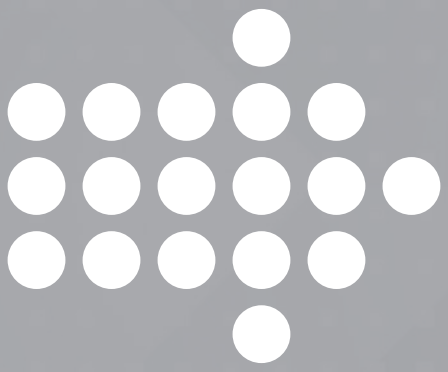
[2] Kancherla, K., & Mukkamala, S. (2013, April). Early lung cancer detection using nucleus segementation based features. In 2013 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (pp. 91-95). IEEE.

[3] Markowetz, F. (2017). All biology is computational biology. PLoS biology, 15(3), e2002050.

[4] Materi, W., & Wishart, D. S. (2007). Computational systems biology in cancer: modeling methods and applications. Gene regulation and systems biology, 1, 117762500700100010.

[5] Prof. Liu S., Computational Biology in Translational Cancer Research". (2020). Retrieved August 13, 2021, from "Computational Biology in Translational Cancer Research".

[6] Statistics on preventable cancers. (2020, December 31). Retrieved September 22, 2021, from https://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/preventable-cancers



Granular Metamaterials for Soft Robotic Applications

Defne Ciliz[1] and Cory S. O'Hern[2]

[1] Uskudar American Academy, Istanbul 34664 Turkey

[2] Department of Mechanical Engineering & Materials Science, Yale University, CT 06520

November 22, 2021

Abstract

This summary presents our current research studies of granular metamaterials for soft robotic applications. Soft robotics show great promise for improving the manipulation and grasping of objects with complex shapes. Traditional robotic grippers use rigid, multi-fingered mechanical manipulators. However, these approaches do not perform well for gripping brittle, complex-shaped, and deformable objects. Recent research has shown that soft robotic grippers made of granular materials confined within an elastic bag have dramatically improved performance. When the soft gripper presses on an object, it takes the shape of the object because the enclosed grains can rearrange. Then, a vacuum

can be created within the elastic bag, which causes the granular material to jam and become attached to the object, and thus the object can be picked up by the gripper without any sensory force feedback. Proof-of-principle studies of such soft robotic grippers have been performed, but there is currently no fundamental understanding of these systems that would allow us to predict their performance for a wide range of tasks. In this research, we carry out experimental and computational studies of model granular metamaterials (granular cylinders confined by elastic bands) to better understand the types of jammed configurations that can occur in these systems.

Keywords: Granular metamaterials, soft robotics, jamming.

1. Introduction

With their complex nature and wide range of pattern forming behaviors, granular materials are used in a variety of applications in industry. Their reversible behavior, where they transition from a fluid-like state to a jammed one, makes them suitable for numerous applications in engineering, such as soft robotics. Applications in soft robotics include grasping objects with complex shapes, as well as delicate or sharp objects, locomotion on variable and complex terrains, and the ability to change shape to enter confined spaces. During natural disasters when buildings collapse and entrance to buildings is limited, soft robots made up of granular materials (such as soft robotic snakes) can change their shape to enter collapsed buildings for search and rescue. Such robots can be used for underwater applications as well [1]. There have been several recent reviews of soft robotic applications [2, 3, 4, 5, 6, 7].

One specific and rather new application of granular metamaterials is soft robotic grippers, where robotic manipulators use a flexible membrane filled with granular materials. The working mechanism of jamming grippers allows them to grip onto different shaped objects by suction (by applying vacuum) while their elastic membrane allows them to mold around any shape. First, the gripper approaches an object in a soft state, then it deforms around the object, and air is sucked out of the membrane so that the gripper can hold onto the object [8]. See Figure 1.

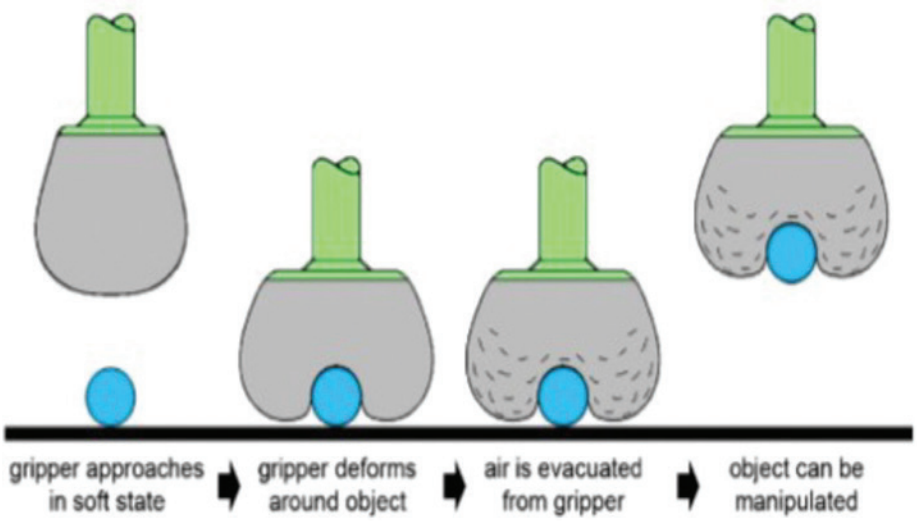


Figure 1.
Jamming-based gripper mechanism from Ref. [8].

When the vacuum is established, the grains become jammed and apply a range of forces around the object they grip onto, holding onto it strongly [8]. The arrangements of the grains inside the grippers as the vacuum is established are unknown. There have been previous computational and experimental studies characterizing the jamming transitions in systems with particles confined by rigid boundaries [9, 10, 11, 12, 13, 14]. Such jamming transitions occur when granular materials reach a certain packing fraction, preventing particle motion.

Yet there have been no studies analyzing jamming in systems where particles are confined within elastic bound- aries. The aim of our research is to carry out experimental and computational studies of granular cylinders confined by deformable boundaries to determine t h e types of grain configurations that can occur.

2. Methodology

To analyze the grain arrangements and characterize the jamming of the granular metamaterials inside the soft robotic grippers, we carry out simulations and experiments on granular cylinders confined by a rubber band. We carry out experiments, as well as Matlab simulations of frictionless disks to determine if friction between the granular cylinders plays an important role in determining the particle configurations within the elastic band.

For the experiments, we analyze the configurations for N = 8, 9, and 10 granular cylinders. We use a 3D printer to generate uniform cylinders. The cylinders have following dimensions: Height: 3 cm, Radius: 0.95cm. The 3D printer we used in this study and the printed cylinders are shown in Figure 2.

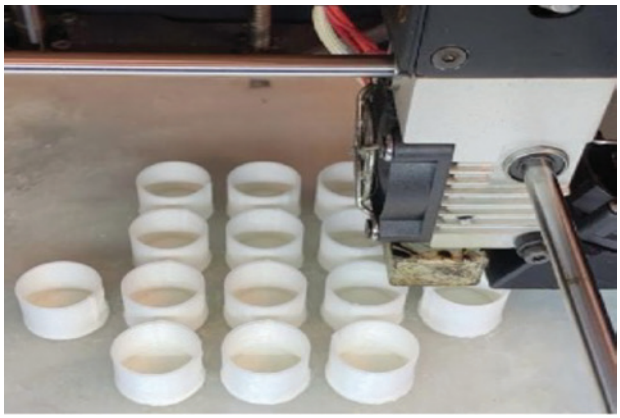
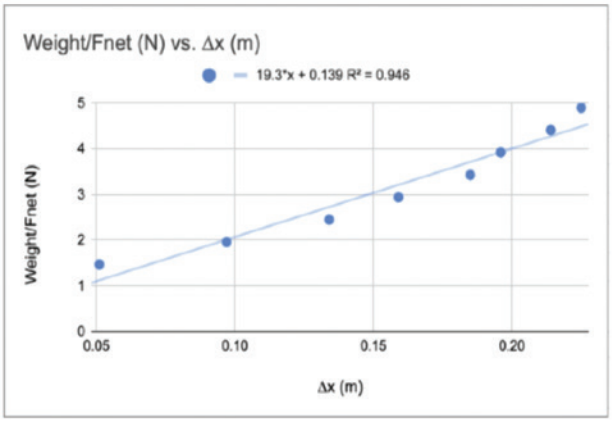


Figure 2.
3D Printer and 3D-Printed Granular Cylinders.

The rubber band we used for the experiments is 19.8 cm in length (unstretched) and has a spring constant of k = 19.3 N/m, which we calculated from experiments using Hooke's Law. See Figure 3.

Δx (m)	weight (N)	mass (kg)
0.051	1.4715	0.150
0.097	1.9620	0.200
0.134	2.4525	0.250
0.159	2.9430	0.300
0.185	3.4335	0.350
0.196	3.9240	0.400
0.214	4.4145	0.450
0.225	4.9050	0.500

a. Experimental Data



b. Estimation of the spring constant of the rubber band

Figure 3.
a) Experimental data table for the amount of stretch of the rubber band for different test masses. b) The force required to stretch the rubber band to a given change in length.

As can be seen from the Figure 3, the force required to stretch the rubber band is roughly linear in the change in length of the rubber band (the amount of stretch) with a slope of k = 19.3N/m.

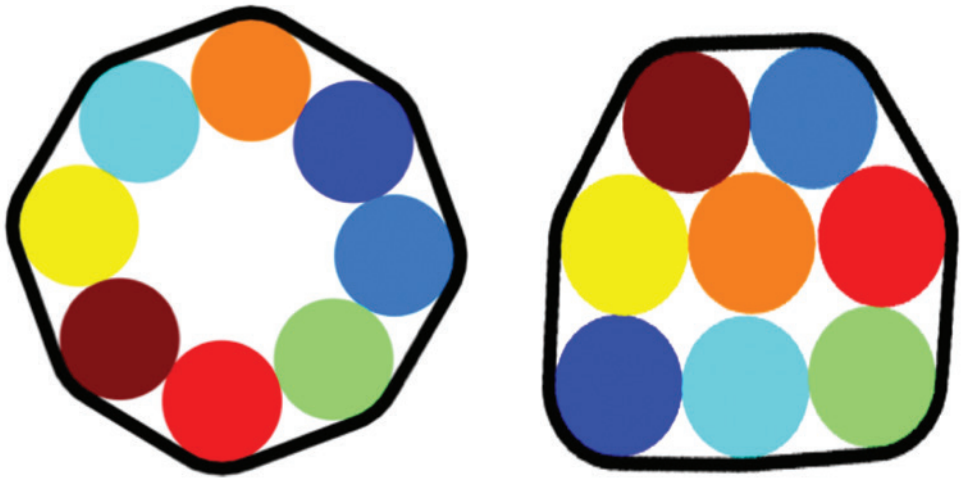
The rubber band we used with an unstretched length = 19.8 cm required a minimum of 8 cylinders to stretch it. For N = 8 cylinders, we put the rubber band around the cylinders and applied different forces to the outside of the rubber band to obtain different configurations. We recorded the images of 10 - 15 configurations and repeated these steps by adding more cylinders until N = 10. We analyzed more than 50 distinct configurations, eliminating configurations that we observed more than once.

After obtaining all of the images of these configurations, we calculated the shape parameter for each configuration of granular cylinders. The shape parameter A for each configuration is given by

$$A = \frac{p^2}{4\pi a} \quad (1)$$

where a is the area enclosed by the configuration, and P is the perimeter of elastic band surrounding the configuration. Note that from Eq. (1), A = 1 for circles and A > 1 for non-circular shapes. The shape parameters for two different N = 8 particle configurations are shown in Figure 4 to demonstrate the computation of A for both cases. Note that the more circular configuration has a shape parameter A closer to 1, (A = 1.055), whereas the second configuration with noncircular shape has a shape parameter of A = 1.238.

When calculating the shape parameter A for different configurations, we exclude the gaps formed between the cylinders and the rubber band and focus on polygons formed from the centers of the cylinders. This calculation w a s performed manually for each case and an example of this method is shown in Figure 5 for 3 different N=8 configurations.



a. b.

Figure 4.
a) Configuration with N=8 Particles with Shape Parameter A =1.055. b) Configuration with N=8 Particles with Shape Parameter A=1.238.

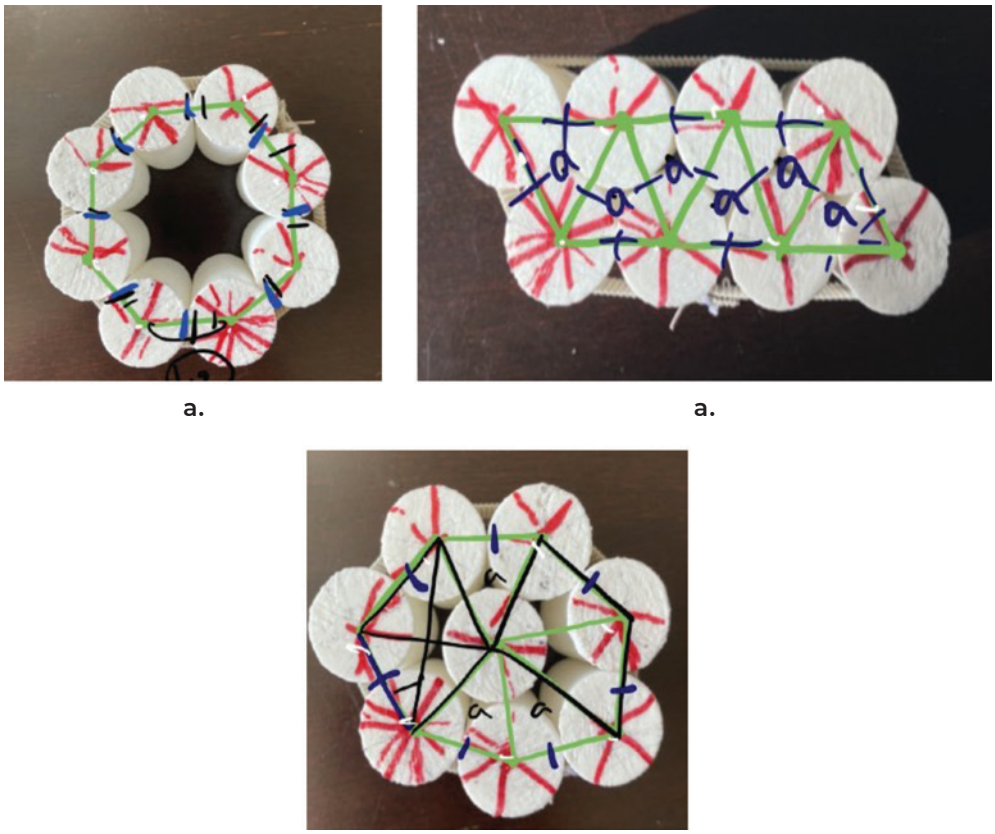


Figure 5.
Figure 5: (a) - (c) Polygons formed by connecting the centers of the cylinders in three configurations for N=8.

After calculating and recording the shape parameters using the same method as shown in Figure 5 for different number of particles, N= 8, 9, and 10, we also recorded the number of contacts (*Nc*) with the boundary (elastic band) for each configuration. The results for these experiments are given in Figure 6 (where *Nc* represents the number of contacts and *a8, a9* and *a10* represent shape parameters in this Figure).

N=8		N=9		N=10	
Nc_8	a_8	Nc_9	a_9	Nc_10	a_10
7	1.2248	7	2.125	8	1.171
7	1.2446	6	2.126	8	1.1077
8	1.96	7	1.47	8	1.1763
7	1.1876	8	1.1403	8	1.142
7	1.188	8	1.273	8	1.1077
7	1.158	8	1.0548	8	2.177
8	1.0548	7	1.273	7	1.862
6	1.9898	7	1.346	7	2.009
6	2.038	7	1.280	7	1.654
6	2.117	7	1.388	8	1.123

Figure 6.
Experimental Results

We then carried out computer simulations for frictionless granular disks confined by an elastic band in 2D. We compare the simulation results with the experimental results. We obtained at least 10 different configurations for each system size, *N* = 8, 9 and 10. For each simulation, we start the simulation with different initial positions of the disks, so that we end up with different configuration for each case. We then calculated the shape parameters and recorded the number of contacts for each configuration and the results are displayed in Figure 7.

To better analyze the experimental and simulation results, we plot the shape parameters *(A)* versus the number of contacts *(Nc)* for each configuration on the same graph for different particle numbers N=8, 9, and 10. Comparison of the experimental results with the simulation results for the three different numbers of particles N=8 ,9 and 10, are shown in Figures 8,9 and 10, respectively.

N=8		N=9		N=10	
Sim_Nc_8	Sim_a_8	Sim_Nc_9	Sim_a_9	Sim_Nc_10	Sim_a_10
7	1.539	7	1.466	8	1.1318
7	1.238	7	1.489	8	1.109
7	1.122	7	1.512	8	1.109
7	1.1469	7	1.512	8	1.1263
8	1.055	7	1.466	8	1.171
7	1.074	7	1.481	9	1.0925
8	1.055	7	1.353	8	1.144
7	1.0677	8	1.0548	8	1.1604
7	1.0907	8	1.0548	8	1.119
7	1.0846	9	1.0427	8	1.146

Figure 7.
Simulation Results

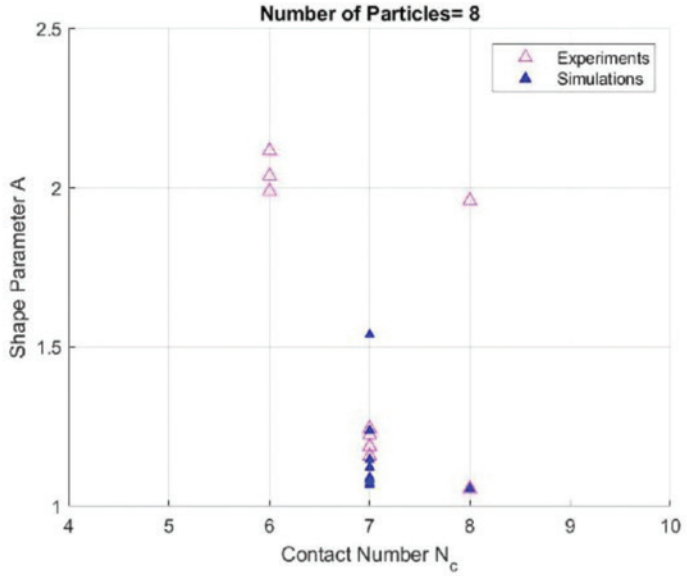


Figure 8.
Shape Parameters for Experiments and Simulations for N=8

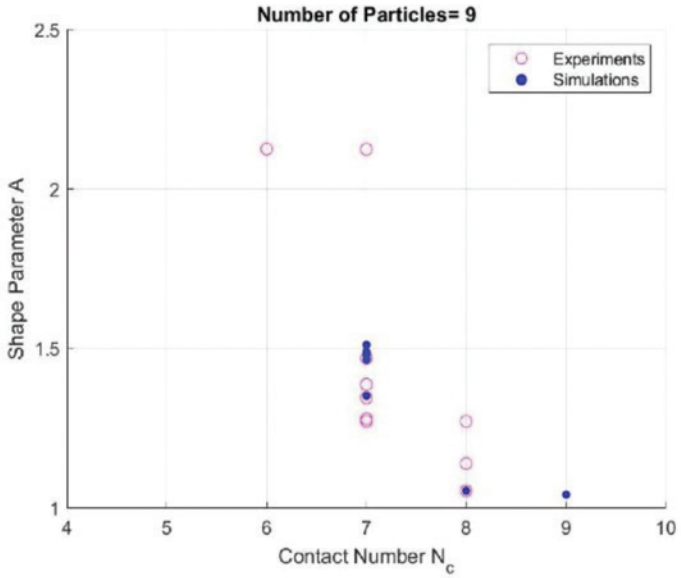


Figure 9.
Shape Parameters for Experiments and Simulations for N=9

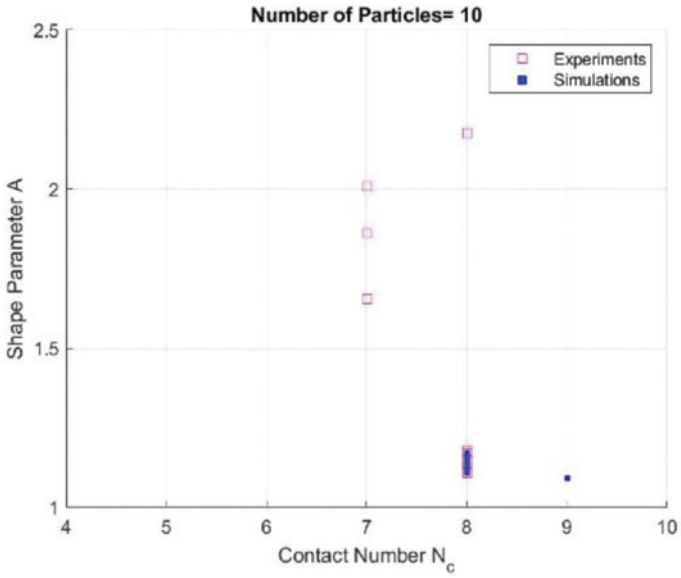


Figure 10.
Shape Parameters for Experiments and Simulations for N=10

Note that for the simulation results for N = 10, we observed a contact number of *Nc* = 8 for all cases except for one case which had a contact number of 9. This is why we have highly populated data points at *Nc* = 8 for simulation tests in Figure 10. It is clearly seen from the above Figures 8-10 that the simulation tests resulted in configurations with significantly smaller shape parameters *(A)* as compared to the shape parameters of experimental tests. We also note that the number of contacts *(Nc)* for configurations obtained via simulations are also greater than the number of contacts obtained in experimental tests. This is mainly due to the unmodeled friction between the cylinders in the simulation studies, whereas friction is present in the experimental tests.

3. Results and Discussion

We find that the granular cylinders can take on a range of different configurations with different shapes within the elastic band, all of which are jammed. When comparing the experimental and simulation results, we observe that con-

figurations obtained from the experiments have fewer contacts and a larger range of shape parameters compared to configurations obtained from the simulations. The configurations from the simulations have greater contact numbers and more circular shapes.

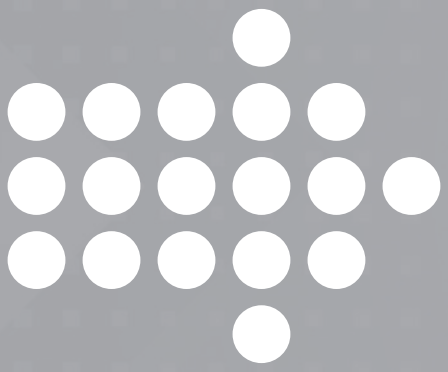
The reason behind this difference is that without the presence of friction, there are only normal forces between the cylinders and between the elastic band and cylinders, hence the granular cylinders can easily slide. However, the presence of frictional contacts of the cylinders in the experimental studies stabilizes the jammed packings and non-circular shapes can easily be formed with fewer contacts between the cylinders and the cylinders and the elastic band.

4. Conclusions and Future Directions

In future work, we propose to car ry out similar studies in three dimensions, where an elastic bag, such as a balloon is filled with spherical grains. Using experiments, we can calculate the shape parameters of the jammed configurations of spherical particles by taking multiple images from different views and using image processing. Using computer simulations, we can determine the granular configurations that correspond to each shape of the elastic bag. We will also include frictional interactions between the spherical particles and between the particles and the elastic container.

References

1. S. Licht, E. Collins, M. L. Mendes, and C. Baxter, “Stronger at depth: Jamming grippers as deep sea sampling tools,” Soft Robotics, vol. 4, no. 4, 2017.
2. J. Amend, N. Cheng, S. Fakhouri, and B. Culley, “Soft robotics commercialization: Jamming grippers from research to product,” Soft Robotics, vol. 3, no. 4, 2016.
3. S. G. Fitzgerald, G. W. Delaney, and D. Howard, “A review of jamming actuation in soft robotics,” Actuators, vol. 9, no. 4, 2020.
4. J. R. Amend, E. Brown, N. Rodenberg, H. M. Jaeger, and H. Lipson, “A positive pressure universal gripper based on the jamming of granular material,” IEEE Transactions on Robotics, vol. 28, no. 2, pp. 341–350, 2012.
5. J. Shintake, V. Cacucciolo, D. Floreano, and H. Shea, “Soft robotic grippers,” Advanced Materials, vol. 30, no. 29, p. 1707035, 2018.
6. T. Mitsuda and S. Otsuka, “Active bending mechanism employing granular jamming and vacuum-controlled adaptable gripper,” IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 3041–3048, 2021.
7. S. Terrile, M. Argu¨elles, and A. Barrientos, “Comparison of different technologies for soft robotics grippers,” Sensors, vol. 21, no. 9, 2021.
8. E. Brown, N. Rodenberg, J. Amend, A. Mozeika, E. Steltz, M. R. Zakin, H. Lipson, and H. M. Jaeger, “Universal robotic gripper based on the jamming of granular material,” Proceedings of the National Academy of Sciences, vol. 107, no. 44, pp. 18809–18814, 2010.
9. Q. Wu, C. Cui, T. Bertrand, M. D. Shattuck, and C. S. O’Hern, “Active acoustic switches using two- dimensional granular crystals,” Phys. Rev. E, vol. 99, p. 062901, Jun 2019.
10. G.-J. Gao, J. Blawzdziewicz, C. S. O’Hern, and M. Shattuck, “Experimental demonstration of nonuniform frequency distributions of granular packings,” Phys. Rev. E, vol. 80, p. 061304, Dec 2009.
11. M. Miskin and H. Jaeger, “Adapting granular materials through artificial evolution,” Nature Materials, vol. 12, p. 326–331, 2013.
12. G.-J. Gao, J. Blawzdziewicz, and C. S. O’Hern, “Geometrical families of mechanically stable granular pack- ings,” Phys. Rev. E, vol. 80, p. 061303, Dec 2009.
13. P. J. Tuckman, K. VanderWerf, Y. Yuan, S. Zhang, J. Zhang, M. D. Shattuck, and C. S. O’Hern, “Geometrical families of mechanically stable granular packings,” Soft Matter, vol. 16, p. 9443, 2020.
14. J. Gomez–Paccapelo, A. Santarossa, H. Bustos, and L. Pugnaloni, “Effect of the granular material on the maximum holding force of a granular gripper.,” Granular Matter, vol. 23, no. 4, 2021.



Breast Cancer Imaging and Artificial Intelligence

Mert Albeyoglu

https://orcid.org/0000-0001-9999-1382

Cancer has been the number one cause of death since its ancient start. Cancer treatment and diagnosis has played a significant part in cancer research for decades. Modern medicine, with all its advances and uses, has yet to be improved upon. Artificial Intelligence applications integrated together with human ability in medicine can be the resolution we have been awaiting to improve breast cancer diagnosis using imaging. Artificial intelligence programmed into imaging can 1-help reduce the amount of wrong diagnosis, 2-help reduce the number of cases where a diagnosis would have been missed, and 3-help speed up diagnosis time through the double-screening process through imaging.

Introduction

"Cancer is a broad term for a class of diseases characterized by abnormal cells that grow and invade healthy cells in the body. Breast cancer starts in the cells of the breast as a group of cancer cells that can then invade surrounding tissues or spread (metastasize) to other areas of the body."[i] We have been trying to create and treat cancer with multitude therapies and drugs and worldwide efforts in this regard have meant trillions of dollars in contributions to science. With the introduction of artificial intelligence into the cancer research arena it could be argued that drug development will be easier and more precise. Studies conducted, like that of Professor Tate at Imperial College London, who works on protein mapping for potential new cancer drugs could mean something new can be added to the equation.[ii] He mentions the process of tagging which could be applied using artificial intelligence. "1 in 2 people will develop some form of cancer during their lifetime. In the UK, the 4 most common types of cancer are: Breast, lung, prostate, and bowel cancer" out of the potential 200 types (Figure 1 and 2).[iii] When cancer cells over reproduce and form a mass this is known as a tumour and is usually where the cancer has started to form. When the cancer spreads to other parts of the body and forms a mass, it is known as a "secondary tumour or metastasis." Moreover, it is not only the formation of a tumour that can cause symptoms in a cancer patient, but symptoms may also arise from problems with "blood circulation, lymphatic and immune systems, and the hormone system" (Figure 3).[iv]

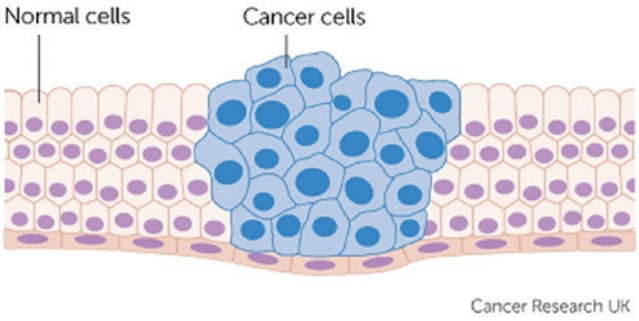


Image 1.
Cancer cell identification, Cancer Research UK[v]

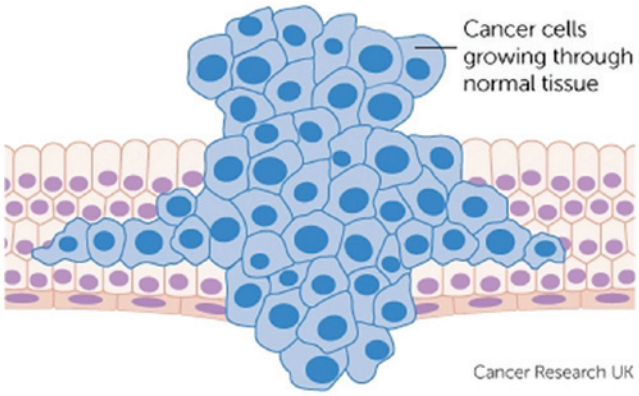


Image 2.
Cancer cells reproducing, Cancer Research UK[vi]

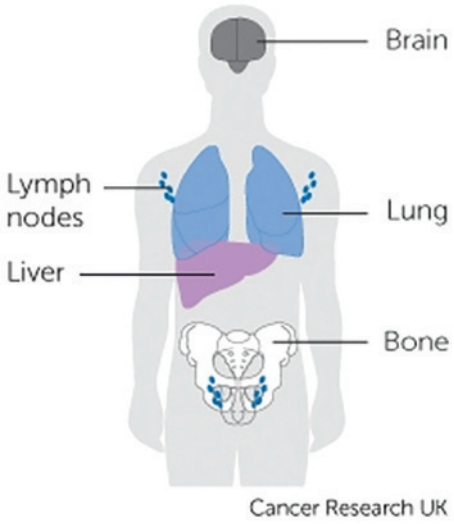


Image 3.
Cancer spreading to other parts of the body, Cancer Research UK[vii]

Breast Cancer, as the focus of this paper, is the most common type of cancer that affects women. "Most women diagnosed with breast cancer are over the age of 50, but younger women can also get breast cancer."[viii] The most important aspect of treating cancer comes with its early diagnosis and this is where artificial intelligence (AI) enters cancer research. While the symptoms and causes of breast cancer cannot be ignored it is the aim of this paper to address the importance of detecting and diagnosing all factors that may come into question.

Artificial Intelligence for Breast Cancer Screening

Screening for breast cancer is the most effective way to prevent cancer from becoming life-threatening and to treat patients earlier for a more effective recovery. "Mammographic screening, where X-ray images of the breast are taken, is the most commonly available way of finding a change in your breast tissue (lesion) at an early stage."[ix] Perhaps AI could be more effective and offer less invasive testing for women and offer fewer tests for women aged 50-70 than the required one every three years.[x]

Artificial intelligence, or AI, was developed in the 1950s while imaging dated back earlier to the late 1800s when German professor Wilhelm Rontgen invented medical imaging (The concept of medical imaging began in 1895 with the invention of the x-ray by a German professor of physics, Wilhelm Rontgen. [xi] Since the 1970s we have had CT and MRI scanners (Computed Tomography and Magnetic Resonance Imaging).[xii] It could be argued that AI has been used with medical imaging since the 1970s since the theoretical ideas from John McCarthy, from Stanford University, existed prior. He called AI "...the science and engineering of making intelligent machines, especially intelligent computer programs."[xiii] Moreover, "AI is a field, which combines computer science and robust data sets, to enable problem solving."[xiv] As subsets of AI, machine learning and deep learning have a primary focus in the literature about cancer. These two subsets primarily function on algorithms that are programmed by computer scientists to perform certain tasks such as data processing through images and prediction making. Applications of Artificial Intelligence can be primarily useful in imaging with the use of computer vision "technology [that] enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, and based on those inputs, it can take action."[xv] Computer vision "is powered by convolutional (complex) neural networks" and can be applied to "photo tagging... [in] radiology imaging in healthcare" and other areas.[xvi] Machine Learning is the most cited application of AI for cancer imaging. "Machine learning is a subfield of AI that gives computers the ability to learn without explicitly being programmed."[xvii] Professor Aleksander Madry, of the Massachusetts Institute of Technology, says that "machine learning is changing, or will change, every industry, and leaders need to understand the basic principles, the potential, and [its] limitations."[xviii] Created by Arthur Samuel in 1950 machine learning is about enabling machines to learn (with algorithms) by themselves. [xix] Machine learning has the potential to make breakthroughs in cancer research especially used for mammograms. Thomas Malone argues that machine learning is the best type of AI because it can process big data.[xx]

This could be a great response to people who think that using AI in the

mammogram screening process will make a lot of people lose their jobs. While unemployment is an issue as we further develop AI technologies, many companies are using machine learning in several ways to promote healthcare innovations. These AI applications are readily found in medical imaging and diagnostics. Machine learning programs can be trained to examine medical images or other information and look for certain markers of illness, like a tool that can predict cancer risk based on a mammogram.

Artificial intelligence entered the field of medicine in the early 1970s. Today, AI can be used for multitude purposes in medicine and the healthcare industry such as for "online scheduling of appointments, online check-ins in medical centres, digitization of medical records, reminder calls for follow-up appointments and immunization dates for children and pregnant females to drug dosage algorithms and adverse effect warnings while prescribing multidrug combinations." (See pie chart)

Radiology is primarily where AI can be in medical applications to treat and diagnose diverse diseases and illnesses. So much so that they have become an "indispensable component of the work environment with the origin of picture archiving and communication systems."[xxii] CAD, or computer aided design, is useful in screening of breast cancer, but is known for its errors in terms of its uses in diagnostics, prediction of tumours and accuracy. It tends to output "false positives" in diagnosis and create inaccurate information for physician treatment as a result.[xxiii] AI should increase clinical administrative productivity, decrease the human labour involved with primary care and increase "productivity, precision, and efficacy."[xxiv]

Specifically, imaging using AI for cancer can reduce the amount of wrong diagnosis, to help reduce the number of cases where a diagnosis would have been missed by a physician and to help speed up diagnosis time through the double-screening process. AI in

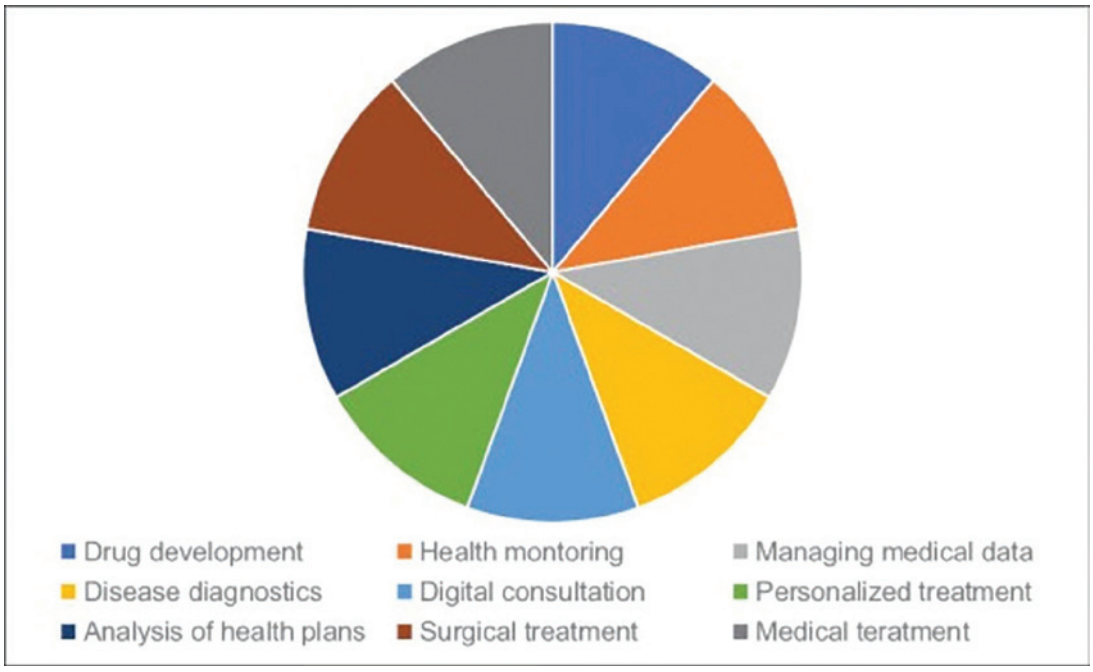


Image 4.
The Uses of AI in healthcare.[xxi]

mammograms, according to Professor the Lord Ara Darzi of Imperial College London argues that "Screening programmes remain one of the best tools at our disposal for catching cancer early and improving outcomes for patients, but many challenges remain – not least the current volume of images radiologists must review."[xxv] In the UK, cancer risks increase with women's age, and it is clinically proven that one out of eight women will be diagnosed with cancer and the early detection of such cancers remains problematic.[xxvi] AI, while it offers potential increase in prediction of tumour size and location for breast cancer and better imaging quality using its quantitative process, it still gives "similar level of accuracy to human doctors," "reduction in incorrectly identified cases (5.7% in UK)," and "identified cases where cancer was missed (9.4% in UK)."[xxvii] In this area of imaging what is known as the "double reading process" revolves around the double interpretation of imaging results by two physicians, and in this case, AI was found to decrease interpretation through workload by 88%.[xxviii]

AI screening integrated together with human screening can 1. Help reduce the amount of wrong diagnosis, 2. Help reduce the number of cases where a diagnosis would have been missed, 3. Help speed up the double-screening process (which in turn speeds up the diagnosis time of a patient, which could be crucial in the treatment process. Using AI for mammogram detection is still a new and developing technology. Therefore, as the researchers themselves have pointed out, there still needs to be a lot of research done to move this technology into our everyday healthcare practices.

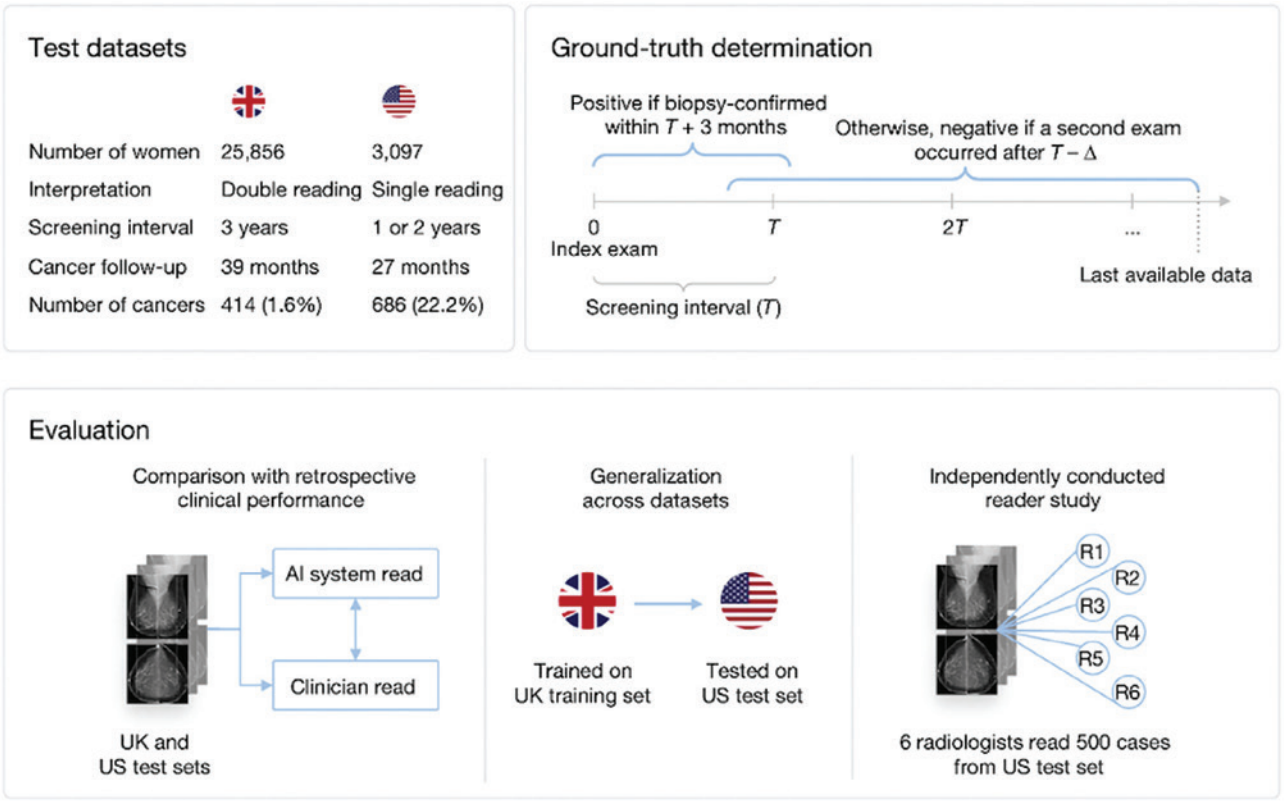


Image 5.
Development of an AI system to detect cancer in screening mammograms.[xxix]

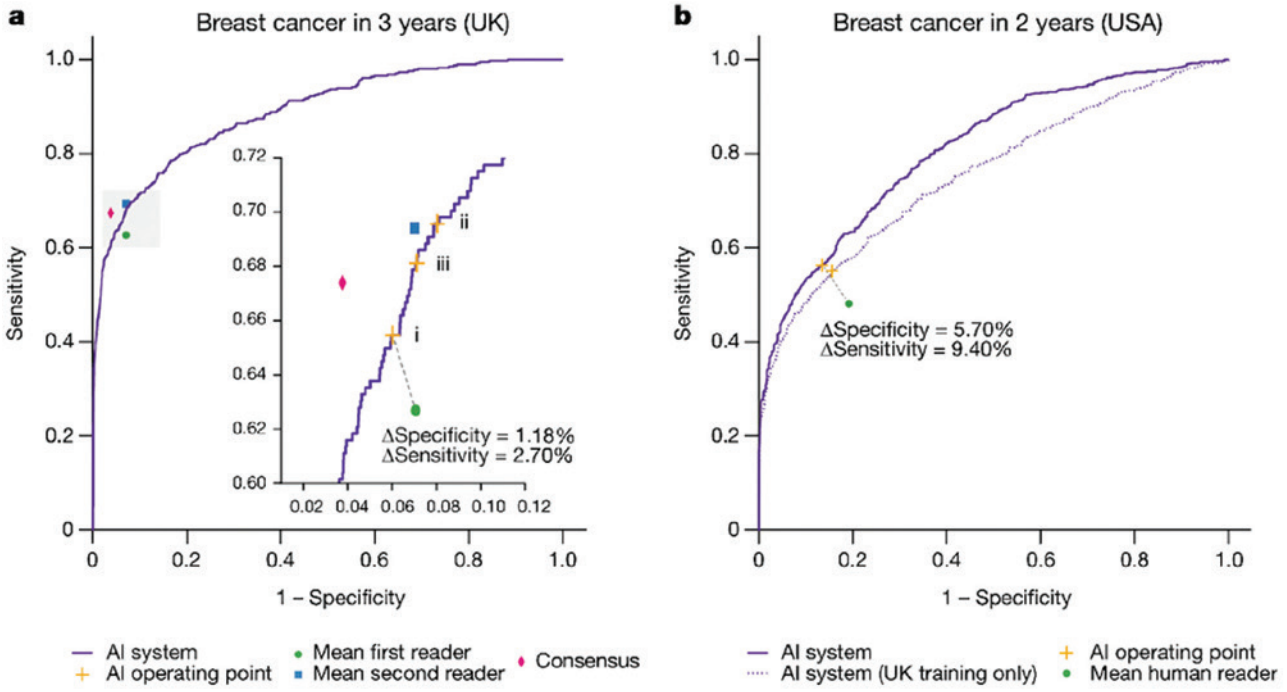
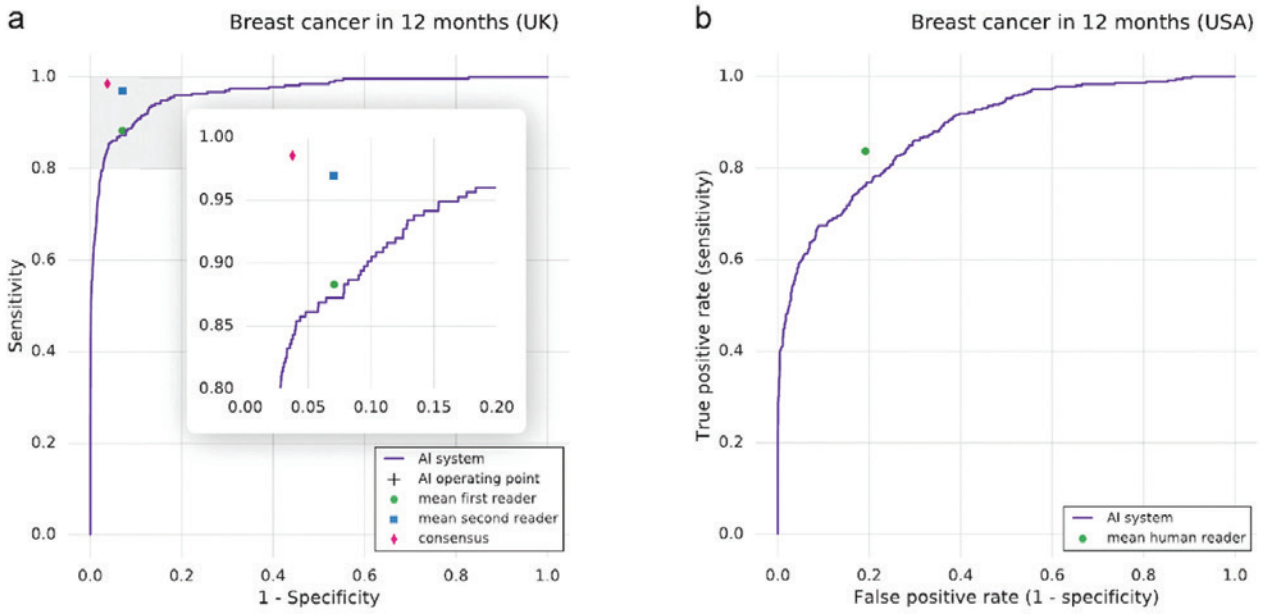


Image 6.
Performance of the AI system and clinical readers in breast cancer prediction. A. The ROC curve of the AI system on the UK screening data. B. The ROC curve of the AI system on the US screening data.[xxxii]

Image five depicts a study that evaluates the performance of AI technology in breast cancer detection in the UK and the US. The results of the test reveal that reading of imaging results were more successful with the AI than with physician alone.[xxx] "As the data sets below show, the ai system achieves superiority, if not, non-inferiority for both sensitivity and specificity from the mean human reader."[xxxi]

One important note to make is that these results also reveal that "no difference in the distribution of cancers detected by the AI system and human readers" were made[xxxiii] In the analysis of the US data the focus seems to be on "the identification of invasive cancers rather than in situ cancer."[xxxiv] However, as was reported in the US case study of imaging for breast cancer detection, there were cases when the AI and human interpretation got mixed up in terms of accuracy and suggested that these roles, between AI and the human physician in radiology, were complimentary.[xxxv] Limitations in both these country-based studies prove that AI is still far more superior to human interpretation and offers doctors with less labour around the meanings of imaging while also exceeding their human capabilities at times, but they both prove that improved AI technologies from diverse manufacturers would be beneficial in increasing the statistical rates of interpreting imaging.[xxxvi]

Questions posed around the accuracy and use of AI could include why is it that we cannot program one AI using UK data and another AI using the US data and run them coexistent? Why not create an AI that is able to double read just as humans can? In this way the graphical data could show the reduction of false diagnosis (see image 7).



c

test dataset	human benchmark	metric	clinical decision (%)	*n*
UK	first reader	sensitivity	88.3	265
		specificity	92.9	25,116
	second reader	sensitivity	97.0	265
		specificity	93.0	25,114
	consensus	sensitivity	98.5	274
		specificity	96.2	25,443
USA	reader	sensitivity	83.7	359
		specificity	80.8	2,185

Image 7.
Quantitative evaluation of reader and AI system performance with a 12-month follow-up interval for ground-truth cancer-positive status.[xxxvii]

Artificial Intelligence Ethics in Breast Cancer Processing

The figure below stands for the advantages and disadvantages of using AI for breast cancer screening. While the use of AI for patients, in terms of early detection may improve disease treatment and outcomes, it creates problems when it comes to giving patients the empathy, sense of the human touch and emotional intelligence. The advantages, however, are noteworthy in that AI supplies efficiency, accuracy and precision through machine learning and deep learning programs. Moreover, in the clinical administrative area, AI offers decreased workload for staff, and it increases patient face-to-face time while also saving patients on costs and offering them better monitoring. What is excluded from the figure below is the aspect of the ethics of consent and security of health information which is a problem that continues to challenge the health care industry with the growing data and resultant need for AI applications.

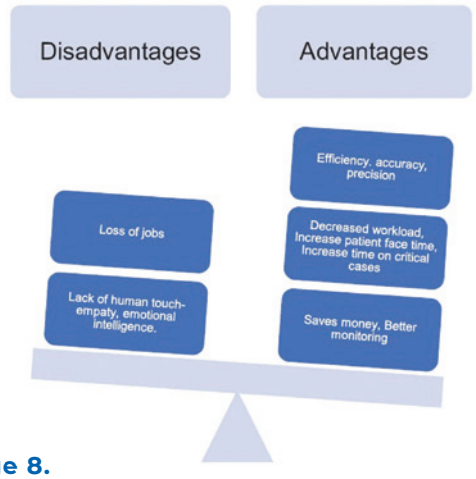


Image 8.
Advantages and disadvantages of artificial intelligence in medicine.[xxxviii]

Studies conclude that AI will never replace the human doctor around mammography:

> "In 2016, the Digital Mammography DREAM Challenge was done where several networks of computers were connected, and the goal was to establish an AI-based algorithm by reviewing 640,000 digital mammograms. The best which was achieved was a specificity of 0.81, sensitivity of 0.80, area under receiver operator curve was 0.87, which is approximated to bottom 10% radiologists. In conclusion, AI has potential, but it is unlikely that AI will replace doctors out rightly.[xxxix]

AI would be an integral part of medicine in the future. Hence, it is important to train the new generation of medical trainees about the concepts and applicability of AI and how to function efficiently in a workspace alongside machines for better productivity along with cultivating soft skills like empathy in them.

Conclusion

Cancer has been a part of our lives for decades and breast cancer has affected many women globally. In terms of its diagnosis and treatment, cancer has been one of the most challenging diseases to cure. While no cure may be in sight soon it is with much optimism that we use AI, for its powerful classifications and prediction purposes, to detect tumours early to help physicians treat the disease faster to save many lives. AI integrated with machine and deep learning have offered the most promising results in imaging to reduce the numbers of wrong diagnosis, reduce number of patients who were missed and to help speed up the process of imaging results that are also more correct. It is important that medical staff are more literate and comfortable with AI for the future of health care. Even if AI could physically and mentally replace humans, as developments are continually be made in affective AI, emotion sensing AI, it is not likely that the near future will be able to address fully the needs of humans in health care. Even if AI were to become fully autonomous and conscious, more human-like, if humans exist, we will always have a relationship with our technologies in every part of our lives, and this includes health.

References

(n.d.). Retrieved from https://www.nhs.uk/conditions/breast-cancer/

[Photographer], T. A., Myers, M., & Producer], M. S. (2017, December 14). Experts explain the potential for new cancer drugs and treatments: Imperial News: Imperial College London. Retrieved from https://www.imperial.ac.uk/news/183706/experts-explain-potential-cancer-drugs-treatments/

About Cancer. (n.d.). Retrieved from https://www.cancerresearchuk.org/about-cancer/what-is-cancer.

Amisha, Malik, P., Pathania, M., & Rathaur, V. K. (2019, July). Overview of artificial intelligence in medicine. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6691444/

Brown, S. (2021, April 21). Machine learning, explained. Retrieved from https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

By: IBM Cloud Education. (n.d.). What is Artificial Intelligence (AI)? Retrieved from https://www.ibm.com/cloud/learn/what-is-artificial-intelligence

OHare, R. (2020, January 01). Artificial intelligence could help to spot breast cancer: Imperial News: Imperial College London. Retrieved from https://www.imperial.ac.uk/news/194506/artificial-intelligence-could-help-spot-breast/

Rubin, A., & Rubin, A. (2017, April 06). History of Medical Imaging - A Brief Overview. Retrieved from https://www.flushinghospital.org/newsletter/history-of-medical-imaging-a-brief-overview/

What Is Cancer? (2019, September 30). Retrieved from https://www.nationalbreastcancer.org/what-is-cancer/

Endnotes

i https://www.nationalbreastcancer.org/what-is-cancer/
ii https://www.imperial.ac.uk/news/183706/experts-explain-potential-cancer-drugstreatments/
iii https://www.cancerresearchuk.org/about-cancer/what-is-cancer.
iv Ibid.
v Ibid.
vi Ibid.
vii Ibid.
viii https://www.nhs.uk/conditions/breast-cancer/
ix Ibid.
x Ibid.
xi https://www.flushinghospital.org/newsletter/history-of-medical-imaging-a-brief-overview/
xii See endnote 5.
xiii https://www.ibm.com/cloud/learn/what-is-artificial-intelligence
xiv Ibid.
xv Ibid.
xvi Ibid.
xvii https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained
xviii Ibid.
xix Ibid.
xx Ibid.
xxi https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6691444/
xxii Ibid.
xxiii Ibid.
xxiv Ibid.
xxv https://www.imperial.ac.uk/news/194506/artificial-intelligence-could-help-spot-breast/
xxvi Ibid.
xxvii Ibid.
xxviii Ibid.
xxix Ibid.
xxx Ibid.
xxxi Ibid.
xxxii Ibid.
xxxiii Ibid.
xxxiv Ibid.
xxxv Ibid.
xxxvi Ibid.
xxxvii Ibid.
xxxviii See endnote 21.
xxxix See endnote 21.

NEXT FRONTIER

JOURNAL OF NEXT FRONTIER FOR LIFE SCIENCES AND AI

2021 Volume 4 Issue 4

/NextFrontier