

Module 1

1.2. Bioinformatics: applications in biosciences and biohealth

By *Coral del Val Muñoz*

Professor at the University of Granada, Department of Computer Science and Artificial Intelligence (DECSAI) at the Andalusian Interuniversity Institute in Data Science and Computational Intelligence (DasCI).

By *Pedro Carmona Sáez*

Assistant Professor at the University of Granada, Department of Statistics and Operations Research.

INTRODUCTION

We can now find applications for bioinformatics in every life and health sciences specialty, where the development of new experimental technologies continues to generate massive amounts of data. Indeed, bioinformatics is starting to play an increasingly significant role in an extremely broad range of fields of application, including:

- Drug development and discovery
- Pharmacogenomics
- Precision or personalized medicine
- Microbiology
- Biofuels
- Biomedical text mining
- Food technology and nutrition
- Database development in biological fields
- Synthetic and systems biology
- Image processing in medical diagnostics and imaging applications

In this capsule we propose, by way of illustration, some applications in these previously mentioned areas.

2. PRECISION OR PERSONALIZED MEDICINE

If we are looking for one field that is truly representative of the vast array of applications for bioinformatics, it would be precision or personalized medicine. Personalized medicine can be defined as being able to 'serve the right patient with the right drug at the right time;' in other

words, administering the right treatment to the right patient at the right time. In contrast to traditional medicine in which the diagnosis is made mainly based on clinical manifestations of the pathologies in question, personalized medicine is based on understanding the molecular mechanisms that give rise to the underlying disease and establishing the diagnosis and treatment based on these findings.

The field in which this type of approach is having the greatest impact so far is cancer. This is because the same tumor type can be caused by various molecular alterations, or mutations, which may be completely different between two given patients. This means that the same treatment could affect these different patients in completely different ways. Characterization of these alterations at the individual molecular level is already allowing personalized treatments to be administered. This is associated with consequent medical and economic benefits because it allows physicians to avoid administering therapies that will be ineffective.

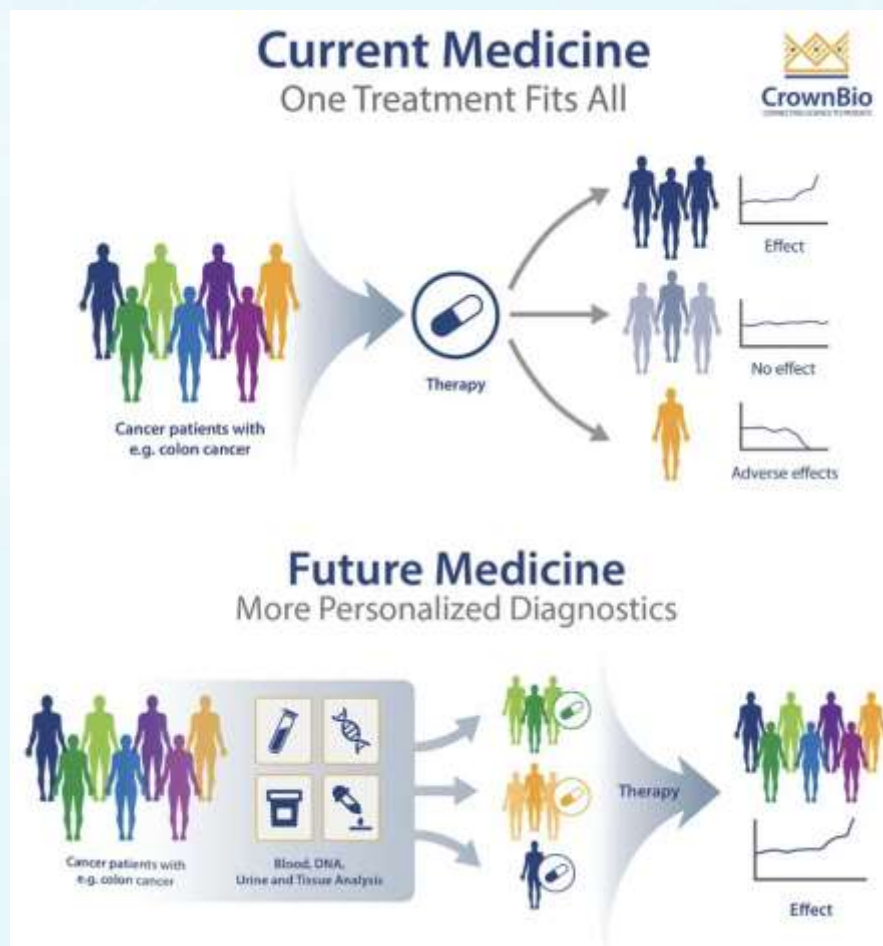


Figure 1. A representation of the potential of personalized precision medicine in the future versus the current paradigm of 'one-fits-all' medicine.

Figura de ejemplo: Fuente <https://blog.crownbio.com/pdx-personalized-medicine>

3. PHARMACOGENOMICS

Pharmacogenomics is becoming an important discipline for the implementation of personalized medicine which is driven by the application of bioinformatics and data analysis techniques. Pharmacogenomics provides information related to an individual's response to a drug based on his or her genetics. This field is based on studying the genome of patients to establish more effective therapies and to predict drug responses or side effects. In this context, polymorphisms (natural variants in the genes of different individuals) in genes encoding enzymes responsible for drug metabolism or that are drug targets are usually studied to establish the individual's likely response to a given treatment or any adverse effects they might have.

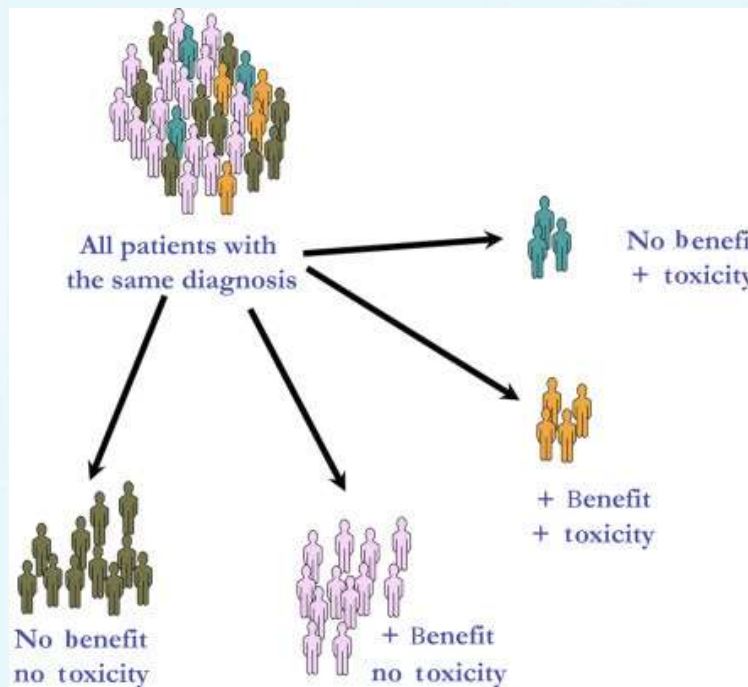


Figure 2. A schematic representation of pharmacogenomics. A drug can have different effects depending on the molecular differences between individuals. Therefore, grouping patients according to these alterations will help physicians to design and administer more effective therapies.

Fuente <https://www.sciencedirect.com/science/article/pii/B9780128006856000047>

4. DRUG DEVELOPMENT AND DISCOVERY

Drug discovery and development is a complex process that requires elevated levels of investment by pharmaceutical companies, both in terms of cost and time. Over the last few decades, the main strategy for drug development has been based on mass screening techniques to simultaneously examine the activity of thousands of molecules to identify any

with promising activity. The more compounds evaluated, the more likely that useful results will be obtained.

Nonetheless, these approaches are still associated with huge costs and low efficiency, or put another way, their efficiency-to-cost ratio remains extremely low. However, obtaining prior information on biological systems as well as data mining using bioinformatic techniques presents enormous possibilities to make the drug discovery processes cheaper and faster. Bioinformatics offers solutions in relation to several different steps in this process, as described in the sections below.

4.1. STRUCTURAL AND BIOPHYSICAL ANALYSIS OF PROTEINS

Bioinformatics can be applied to automatically search databases to identify potential drug candidates that may interact with the protein of interest.

4.2. DRUG REPOSITIONING

This is the science of finding new applications for already known drugs. In this sense, bioinformatics techniques have been used to search for drugs that reverse gene expression profiles. This is achieved by analyzing the information already available in databases, such as connectivity maps, which contain tens of thousands of gene expression signatures for different compounds and drugs (see <https://clue.io/cmap> for more information). This area of research is now highly topical because of the potential of this approach for the discovery of drugs that could help treat COVID-19, allowing the identification of effective compounds that have already been approved for the treatment of other pathologies.

4.3 IDENTIFICATION OF THERAPEUTIC TARGETS

One of the main objectives in the analysis of omics data is the identification of biomarkers (genes, proteins, and metabolites, etc.) that are altered in pathological conditions. These elements can serve as diagnostic biomarkers, but are also potential therapeutic targets that can be explored for the development of specific therapies.

5. MICROBIOLOGY AND METAGENOMICS: FROM BIOBATTERIES TO BIOREMEDIATION

The sequence of more than 250,000 microbial genomes is already known and a multitude of large metagenomics research projects are underway. Given the implications of microbes in health, energy, the environment, and industry, studying their genetic material will allow us to understand them at a very fundamental level. This knowledge is the first step towards isolating

the genes that give them their unique abilities to survive in extreme conditions or to perform various processes of interest. Here we describe two such examples.

5.1. NEW ENERGY SOURCES

In this novel field of study, several projects are leveraging the power of bioinformatics to understand and improve the metabolism of microorganisms that produce electrical currents. For example, microbes capable of generating energy from light, such as *Chlorobium tepidum*, or by metal ion reduction processes via metabolically active biofilms, as in the case of *Geobacter sulfurreducens*, have been identified. Moreover, new advances in the understanding of the time required for electron transport to take place in different bacteria is bringing us ever closer to the creation of bio-batteries that could help increase the production of green energy.

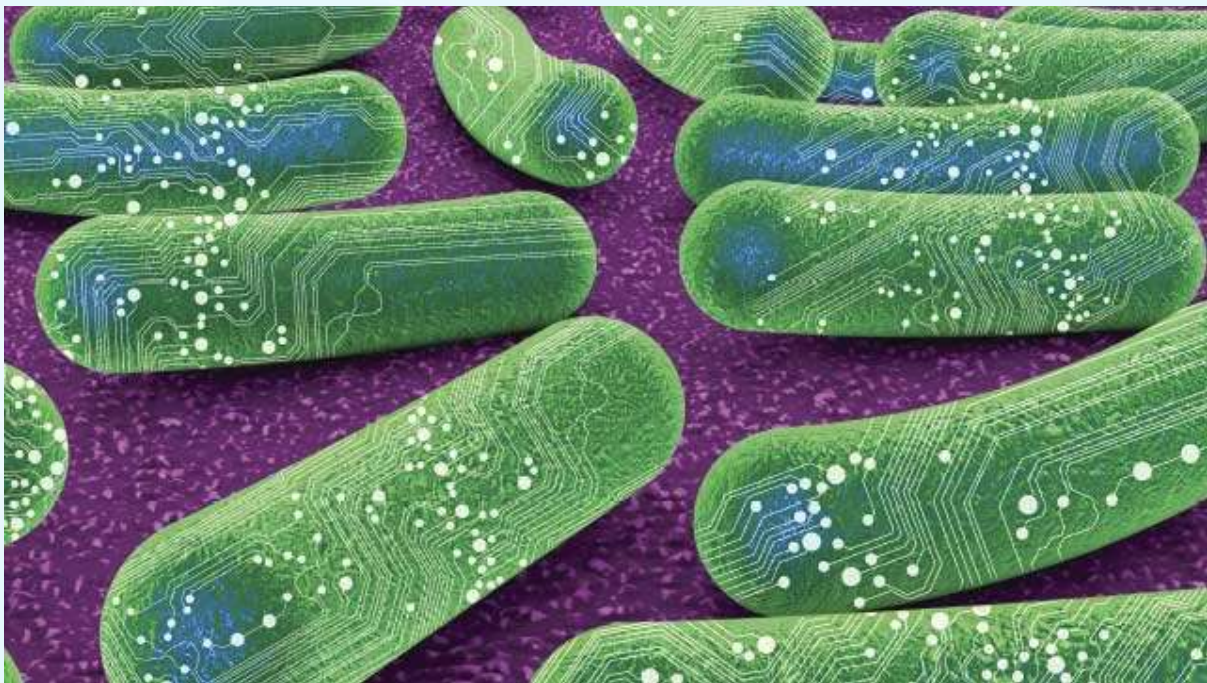


Figure 3. Bioinformatic analysis of the genetic material and metabolic pathways of organisms is opening up the possibility of creating natural batteries that could use wastewater or sludge to generate green energy.

Fuente <https://www.the-scientist.com/features/synthetic-biology-comes-into-its-own-33460>

5.2 BIOREMEDIATION

Bioremediation explores the potential use of different (mainly prokaryotic) organisms in the biodegradation of xenobiotic compounds and waste materials. This is because many of them can degrade contaminants and help restore original environmental conditions. Bioinformatics has been used in this field to determine the structures and biodegradation routes utilized by xenobiotic compounds. One example is the case of *Deinococcus radiodurans*, which is being studied for its potential use in the cleanup of sites contaminated with radiation and/or toxic chemicals. Other examples are *Escherichia coli* and *Pseudomonas aeruginosa* which may be useful in the degradation of petroleum derivatives.

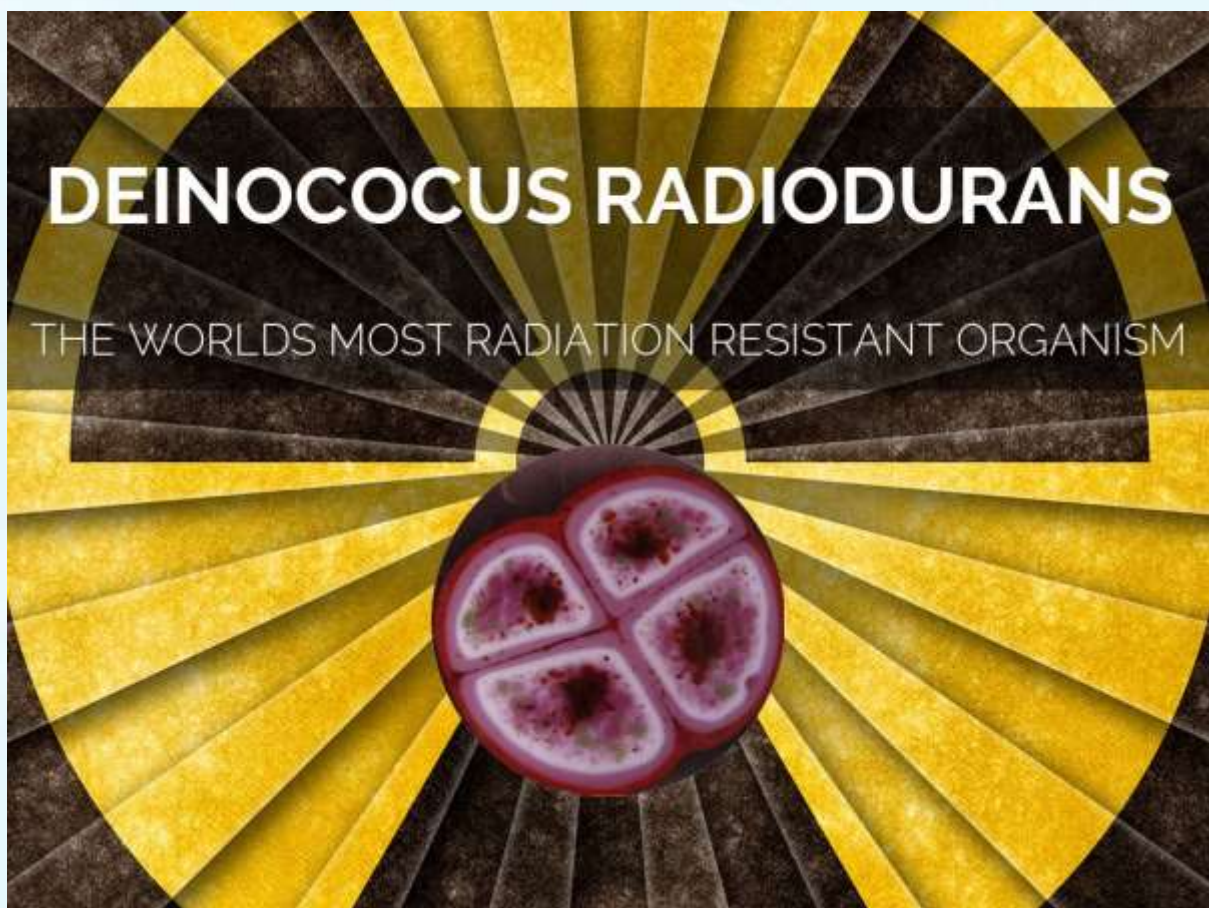


Figure 4. Bioremediation describes the use of organisms such as *Deinococcus radiodurans* to eliminate radioactive contaminants and/or toxic chemicals.

Fuente: <https://www.haikudeck.com/deinococcus-radiodurans--education-presentation-sg6oBMZXEn>

6. SYNTHETIC BIOLOGY: FROM BIOMEDICINE TO BIOFUELS

Synthetic biology is an emerging discipline that aims to design new biological systems or to redesign existing ones for specific purposes; it is an interdisciplinary field that integrates physical, chemical, computational, and biomedical sciences. Apart from integrating bioinformatics techniques, generating knowledge to better understand biological systems and genomes through the analysis of large amounts of data is at the heart of this discipline. Indeed, synthetic biology has been experiencing considerable progress in several key areas in recent years, as described in the examples provided below.

6.1. BIOMEDICINE

Organisms with some interesting capabilities have been engineered to acquire other specific abilities. One example is the modification of a genus of bacteria, *Yersinia*, to acquire the capacity to invade tumor cells. In fact, these bacteria have already been successfully programmed to trigger the expression of a drug that suppresses tumor growth in mice.

6.2. PRODUCTION OF SUBSTANCES OF INTEREST

Microorganisms have been used for decades for this purpose and the advancement of synthetic biology has provided new tools and strategies to improve the efficiency and capabilities of these organisms. For example, metabolic engineering and the introduction of genes from other microorganisms has been used to design novel yeasts with new metabolic pathways capable of synthesizing specific compounds.

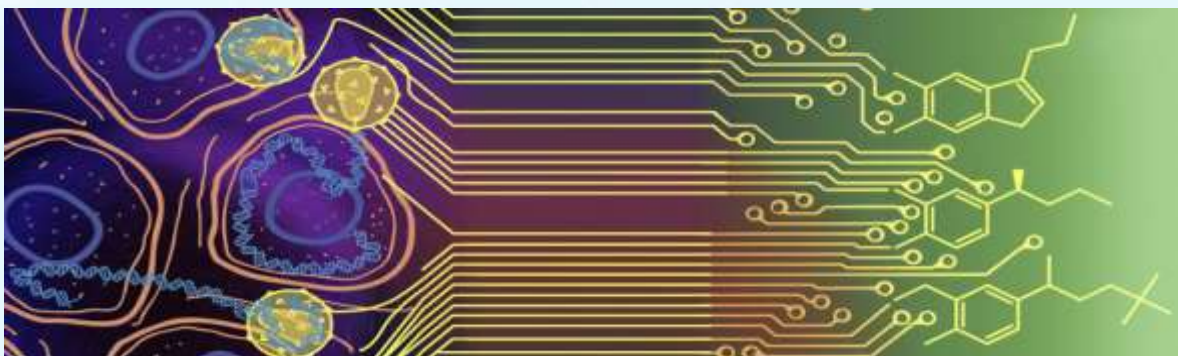


Figure 5. Synthetic biology is used to design new metabolic pathways or gene regulation systems that give cells or microorganisms the ability to produce compounds of interest.

Fuente: <https://www.forbes.com/sites/johncumbers/2020/06/29/pharmas-dirty-secret-and-the-synthetic-biology-startup-offering-a-new-approach/#431641e3436e>

6.3. PLANT BREEDING BY SELECTING MICROORGANISMS ADAPTED TO STRESSORS

Climate change is a major threat to global agricultural production which especially impacts the poorest countries. Bioinformatics is now playing a key role in the integration of genomic, proteomic, lipidomic, and metabolomic data resulting from plant–microorganism interactions. This allows phenotypes resistant to abiotic stressors, such as increased temperature and drought, and/or biotic stressors, such as resistance to fungi or insects, to be selected for in order to maintain agricultural production and soil diversity.

6.4. BIOFUELS

This is one of the fields in which great advances can be achieved through synthetic biology and bioinformatics. Thus, several projects are being developed to improve the metabolism of microorganisms, mainly microalgae, for the production of hydrogen or ethanol as energy sources, the transformation of waste into energy, or the conversion of solar energy into hydrogen.



Figure 6. In the near future, synthetic biology will likely have fundamental applications such as in the production of biofuels and in the design of more resistant and efficient crops.

Ejemplo de imagen:

<https://www.euractiv.com/section/agriculture-food/news/electricity-and-biofuels-needed-in-tandem-to-meet-climate-goals-un-report-says/>

The understanding of the structure and organization of the genome of different organisms thanks to bioinformatics techniques is even making it possible to delve into the synthesis of complete organisms to understand the basis of life at a previously unimaginable level. In this context, in 2019 an article was published in the journal Nature in which a group of researchers created a variant of E. coli with a synthetic genome of four megabases.

7. FOOD TECHNOLOGY AND SPORTS SCIENCE

The use of proteomics is becoming more widespread in the characterization and standardization of raw materials, process development, detection of batch-to-batch variations, and quality control of final products. In addition, genomics is now being used alongside the production of genetically modified foods and to monitor biological and microbial safety. In the latter case, a good example is the application of these techniques to locate the origin of infection in the *Listeria monocytogenes* outbreak in Andalusia (Spain) in 2019.

Bioinformatics also plays a fundamental role in the context of nutrigenomics. Nutrigenetics can be defined as the study of nutrition and heredity, while nutrigenomics is the study of the mutual interactions between molecules, genes, and diet in terms of gene function. The importance of bioinformatics in this area is determined by the accumulation of substantial amounts of data, especially because numerous databases and bioinformatics software packages are dedicated exclusively to this content.

In this context, the study of complex diseases such as obesity also particularly stands out. Interactions between dietary pattern and genetic factors (genetic variations) that condition the way people respond to diet are now being studied with the aim of preventing or reducing obesity and improving sports performance. Many studies have also been conducted to try to improve the nutritional value of products derived from animals and plants. For example, gene transfer into rice strains to increase the resulting levels of vitamin A, iron, and other micronutrients.



Figure 7. Understanding the interactions between genetics and nutrition will help physicians to prevent pathologies and better understand the individual characteristics of the effect of diet and eating patterns.

8. MEDICAL IMAGE PROCESSING

Medical imaging data is one of the richest sources of patient information, for example in X-rays, computed tomography, magnetic resonance imaging (MRI), and positron emission tomography scans. However, these results are often some of the most complex and challenging to interpret, even for the most experienced clinicians. There is therefore a growing interest in transforming medical images into useful data that can be used to improve clinical decision making. Thus, the machine learning and artificial intelligence implemented in many image analysis bioinformatics tools has found a broad niche in biomedical diagnostics. Applications include improving the diagnoses of cancer (e.g., in skin cancer; Esteva, 2017) or retinal diseases. In addition, in a highly topical context, deep learning (DL)-based algorithms are currently being developed for the diagnosis of COVID-19 by analyzing lung radiography images. Specific examples of some of these current applications are described below.

8.1. CANCER DIAGNOSIS THROUGH IMAGING ANALYSIS

Medical imaging is often used in routine preventive examinations for different tumors such as breast, skin, or colon cancer. One of the most widely used techniques is digital breast tomography (DBT) also known as 3D mammography. Indeed, the use of DBT combined with MRI and ultrasound has improved the accuracy of breast cancer diagnoses by minimizing error and aiding in the early detection of cancer. Another example is the timely detection of melanoma—a cancer type that is highly curable if diagnosed early and treated properly; its survival rates are as high as 65% if diagnosed in its initial stages. Proper treatment can even result in a 5-year survival rate of more than 98 percent. In this context, DL algorithms are one of the most promising short-term methods of detecting melanoma in images with a degree of accuracy exceeding that of most experts.

8.2. CARDIOVASCULAR IMAGE ANALYSIS

The use machine learning and artificial intelligence in cardiovascular image analysis has minimized the use of other invasive diagnostic tests such as angiography and catheterization. The ability of this technology to measure the various structures of the heart in images can reveal an individual's risk of cardiovascular disease and identify problems that need to be addressed by surgery or with pharmacological treatment.

8.3. DIAGNOSIS OF NEUROLOGICAL DISEASES

Many degenerative neurological diseases, including amyotrophic lateral sclerosis (ALS) and primary lateral sclerosis (PLS), present similar neurological conditions, although some are much more devastating than others. Increasing the diagnostic accuracy of such diseases could help avoid unnecessary misdiagnoses and can help patients plan for long-term care or to express their end-of-life wishes. Algorithms that are currently being developed attempt to simplify the process of cataloging and annotating images by highlighting suspicious findings and calculating risk indices for images that might contain evidence of the presence of ALS or PLS.



Figure 8. New machine learning methodologies are producing very promising results in terms of medical imaging diagnostics.

REFERENCES

- **Ayers, D., and Day, P.J. (2015).** Systems Medicine: The Application of Systems Biology Approaches for Modern Medical Research and Drug Development (Hindawi).
- **Baumann, N. (2016).** How to use the medical subject headings (MeSH). *Int. J. Clin. Pract.* 70, 171–174.
- **El Karoui, M., Hoyos-Flight, M., and Fletcher, L. (2019).** Future Trends in Synthetic Biology—A Report. *Front. Bioeng. Biotechnol.* 7.
- **Katara, P. (2013).** Role of bioinformatics and pharmacogenomics in drug discovery and development process. *Netw. Model. Anal. Health Inform. Bioinforma.* 2, 225–230.
- **Krallinger, M., Erhardt, R., and Valencia, A. (2005).** Text-mining approaches in molecular biology and biomedicine. *Drug Discov. Today* 10, 439–445.
- **Malkaram, S.A., Hassan, Y.I., and Zemleni, J. (2012).** Online Tools for Bioinformatics Analyses in Nutrition Sciences. *Adv. Nutr.* 3, 654–665.
- **Rigden, D.J., and Fernández, X.M. (2020).** The 27th annual Nucleic Acids Research database issue and molecular biology database collection. *Nucleic Acids Res.* 48, D1–D8.
- **Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., et al. (2017).** A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 171, 1437–1452.e17.
- **Xia, X. (2017).** Bioinformatics and Drug Discovery. *Curr. Top. Med. Chem.* 17, 1709–1726.
- **Gillies R, Kinahan P, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016; 278: 563–77.**