Integrative Cancer Science and Therapeutics

Research Article ISSN: 2056-4546

Decoding artificial intelligence and machine learning concepts for cancer research application

Renaud Seigneuric[1-5]* and Isabelle Bichindaritz[1]

[1]Computer Science Department, State University of New York, Shineman Centre 427, Oswego, NY 13126, USA

[2]University Bourgogne Franche-Comté, Faculty of Health Sciences, Dijon, France

[3]INSERM, LNC UMR1231, Dijon, France

[4]TranslationR, Maison Régionale de l'Innovation, Dijon, France

[5]Centre Georges-François Leclerc, Dijon, France

Abstract

Artificial intelligence (AI) and machine learning (ML) are now almost everywhere. Yet, most of us do not have a formal training on this recent topic. Their concepts emerge from several different scientific communities. Thus, deciphering research articles, understanding their underlying assumptions and limits remains quite challenging.

To this end, we propose a succinct unified AI and ML glossary dealing with 70 important concepts in non-technical yet accurate terms to help non-AI or non-ML researchers exposed to or entering this emerging field, to better understand, assess and use these concepts in cancer research.

Introduction

Initiated in 1956 but mostly unknown from the public until only a couple of years ago, artificial intelligence (AI) and machine learning (ML) are now almost everywhere. As we generally do not have a formal training on the topic, do we have a sufficient basic knowledge or understanding of its central concepts?

Research articles rarely define more than 10 AI and ML terms. Blogs devoted to AI and ML are usually highly technical. Wikipedia is easily accessible but lacks clarity due to contributions from several independent authors.

To better assess research articles in cancer research, limit ambiguity, contradictions and selection bias, a short unified AI and ML glossary dealing with 70 concepts in non-technical yet accurate terms was written in order to help non-AI or non-ML researchers entering this emerging field, to better understand, assess and use these concepts in cancer research.

Glossary

Algorithm

A series of ordered instructions in order to perform a specific task [1]. An algorithm encoded into a structured language (e.g.: R, Python) becomes a computer program that transforms the input(s) to output(s). The name algorithm comes from a Persian scholar, Al-Khwarizmi, who worked in mathematics, astronomy and geography. His book "On the Calculation with Hindu Numerals", written about 820, was principally responsible for spreading the Hindu-Arabic numeral system throughout the Middle East and Europe.

Artificial intelligence (AI)

Defined as a field to manufacture devices able to simulate or even exceed the capabilities of humans at performing cognitive tasks such as: reasoning, problem solving, knowledge representation, planning, learning, natural language processing, perception (including vision), motion and manipulation, social and general intelligence. It was coined in 1956 by a computer scientist, John McCarthy [2].

Artificial neuron

First proposed by Warren McCulloch and Walter Pitts in 1943 to model the "nerve net" in the brain. An artificial neuron is the elementary unit in an artificial neural network. Each one receives one or more weighted inputs and the weighted sum is passed through a non-linear function known as an activation function or transfer function (e.g.: having a sigmoid shape). One pioneering study was the perceptron developed by Frank Rosenblatt. It considered more flexible weight values in the artificial neurons and was used in machines with adaptive capabilities [3].

Artificial neural network (ANN)

A type of neural networks performing tasks like pattern recognition, clustering, classification etc. ANN popularity has

Correspondence to: Renaud Seigneuric, University Bourgogne Franche-Comté, Faculty of Health Sciences, Dijon, France, E-mail: renaud.seigneuric@gmail.com

Key words: *artificial intelligence, machine learning, data science, neural networks*

Received: August 27, 2019; **Accepted:** September 12, 2019; **Published:** September 17, 2019

increased a lot recently due to technical advancements resulting in real life feats such as AlphaGo defeating a world champion of the game Go. Major drawbacks are the large volume of data needed for training and the "black box" algorithm type as it is often difficult to interpret the meaning of the underlying weights in the named sake "hidden" layers.

Augmented intelligence (AI)

Augmented intelligence (AI) was coined to replace "artificial" in artificial intelligence that was found to be misleading. The adjective augmented was chosen to highlight that this scientific and technologic endeavour is meant to improve human intelligence rather than to replace it [4].

Backpropagation

An algorithm for "the backward propagation of errors" was originally introduced by Paul Werbos in 1975. In an artificial neural network, the error is computed at the output. Backpropagation distributes the error term through the layers by modifying the weights at each node. In principle, a learning procedure could repeatedly choose single weights at random, make a small change, and keep this change if it improves the performance of the whole net, but this would be extremely slow. In a neural network with a million weights, backpropagation achieves the same goal about a million times faster than blind trial and error [5]. Its fast implementation by Rumelhart et al in 1986 was a key trigger for renewed interest in neural networks and learning making it possible to use neural nets to solve problems which had previously been insoluble [5]. It is now commonly used to train neural networks [6,7].

Bias

Refers to the tendency of being systematically off target due to the procedure used. In protein-protein interaction studies for instance, current techniques are known to be biased towards stable interactions rather than transient ones.

Big data

Access to and analyses of massive quantities of information produced by and about people, things and their interactions. Originally, it was mostly used to differentiate between analyses ran on desktop versus super-computers. Currently, Big Data generally implicitly means analysing massive quantities of information (e.g.: several gigabytes and above) with a framework (e.g.: Apache Hadoop software library) that allows for the distributed and parallel processing of large data sets across clusters of computers. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage [8-10]. It is estimated that about 90% of the world data has been generated only within the last 2 years.

Breast cancer

The most frequent cancer diagnosed in women worldwide with ~8 million persons currently suffering from it. In the U.S., there are ~270,000 new cases per year and ~41,000 deaths per year due to breast cancer.

Cancer: the 2nd cause of death with ~9 million deaths worldwide. Globally, about 1 in 6 deaths is due to cancer. There are ~100 cancer subtypes.

Case-Based Reasoning (CBR)

A technique, developed by Roger Schank (artificial intelligence) in the late 1970s and early 1980s with contributions from Robert Abelson (social psychology), for modelling how people use memory to solve problems and designing learning machines [11]. CBR is based on two tenets

(a) problems tend to recur and

(b) similar problems have similar solutions.

CBR relies on similar past cases with known solutions to make decisions on new cases (or queries) [12]. A case normally contains a problem, a solution and its result [12,13]. CBR is traditionally split into 4 steps:

(i) retrieve: given a new case to be solved (the target problem), retrieve similar cases (i.e.: trainingexamples) from the case base (or database or datasets or memory) to solve the new problem at hand.

(ii) reuse: adapt the retrieved cases to match to the new case (may require adapting the solution from the previous cases);

(iii) revise: test the new solution and revise if necessary:

(iv) retain: after the solution has been successfully adapted to the target problem, store the resulting solution as a new case in the memory [12,14].

Central processing unit (CPU)

The electronic circuit within a computer that carries out the instructions of a computer program. For machine learning purposes, CPUs tend to be replaced by GPUs and TPUs.

Class

See label.

Classification

The task of generalizing known structures to new data (e.g.: classifying a new e-mail as "legitimate" or as "spam").

Clustering

The task consisting in the assignment of data points to clusters such that items in the same cluster are as similar as possible, while items belonging to different clusters are as dissimilar as possible, thus enabling the emergence of structures within the data. Similarity measures includes Euclidean, correlation and Mahalanobis distances.

Convolutional neural network (CNN)

A type of neural network developed by LeCun in 1989 for processing data having a grid-like topology inspired from animal visual cortex and requiring little pre-processing of the data. Examples include time-series data, which can be thought of as a 1-D grid taking samples at regular time intervals and image data, which can be thought of as a 2-D grid of pixels. Convolutional neural networks use convolution (a linear mathematical operation) in place of general matrix multiplication in at least one of their layers [15].

Data

A piece of information that is selected for an analysis. It is generally categorized as either qualitative or quantitative. In a machine learning perspective, raw (unprocessed) data is encoded according to its type: as a word or a colour (qualitative data) or as a number (quantitative data), collected, pre-processed, visualized, analysed, interpreted and reported in a typical data analysis process.

Dataset

A set or collection of data. A corpus is used for qualitative datasets (e.g.: a set of scientific articles) whereas a dataset tends to refer to quantitative data. A dataset is often organized as a table where each raw corresponds to a variable (or descriptor) and each column an experiment (or patient).

Database

A collection of data organized following a predefined data model. A database is generally stored and accessed from a computer using a specific management system. Relational databases were dominant in the 1980s. These model data as rows and columns in a series of tables, and the vast majority use SQL for writing and querying data. In the 2000s, non-relational databases became popular, referred to as NoSQL as they use different query languages.

Data mining

The process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. The data mining process is classically described as a series of 5 steps:

(i) data selection: extracting a target dataset from the full dataset

(ii) pre-processing: enabling the data to be ready for analyses

(iii) transformation (e.g.: normalization)

(iv) data mining (per se): using a tool to identify patterns

(v) interpretation/evaluation: to provide knowledge [16].

Data partitioning

About splitting the available dataset. Indeed, to learn, and to assess the performance of machine learning approaches, the available dataset needs to be split into the training set (largest part) + the validation set (smallest part). Some approaches also use a third set: the confirmation set. The 2 most frequent partitioning setups are: 2/3 (training set) + 1/3 (validation set) and the leave-one-out cross validation. Machine learning approaches typically use the training set to learn and apply what was learned on the validation set (predictions). The accuracy of predictions is then computed.

Data repository

A collection of numeric data sets for secondary use in research, usually part of a larger institution (academic, corporate, scientific, medical, governmental, etc.). Cancer examples include: GEO Omnibus, a public functional genomics data repository hosted by the NCBI, The Cancer Genome Atlas (TCGA) hosted by the National Cancer Institute [17,18]. TCGA collects data over 20,000 primary cancers and matched normal samples spanning 33 cancer types representing over 2.5 petabytes (about 2.5 million of gigabytes) of genomic, epigenetic, transcriptomic, and proteomic data.

Data science

A broad term that encompasses several disciplines including artificial intelligence, machine learning and data mining.

Decision tree

A technique that builds a set of decision rules describing the relationship between selected variables and the outcome. These rules are used to predict the outcome of a new data point. Decision trees are used for both classification and regression. Hence, they are sometimes referred to as classification and regression trees (CART) [19,20].

Deep learning

A machine learning algorithm using a series of successive layers where each layer uses the output from the previous layer as input. Starting in approximately 2006, technical advances and much faster hardware made it feasible to train neural networks with many layers on large data sets, hence the term deep. It was adopted to differentiate this new generation of neural network technology from its progenitors (shallow) [21].

Deep neural network (DNN)

An artificial neural network (ANN) with multiple layers between the input and output layers. Deep learning architectures include recurrent neural networks and convolutional neural networks. They have been applied to fields including: bioinformatics, drug design, medical image analysis but also computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, material inspection and board game programs where they have produced results comparable to and in some cases superior to human experts. In 2019, Yoshua Bengio, Geoffrey Hinton and Yann LeCun were awarded the Turing Award for their conceptual and engineering breakthroughs for deep neural networks [22].

Dimensionality reduction

A means to increase the computational efficiency by reducing the number of "features", or inputs, in a dataset. Reducing the dimensions of a dataset is performed by projecting it into a space of lower dimension in order while trying to retain most of the information. Most popular types of dimensionality reduction techniques are principal components analysis, linear discriminant analysis and t-distributed stochastic neighbour embedding.

Discriminant analysis (DA)

A ML technique to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification. Originally developed by Ronald Fisher in 1936, its current variations include linear, quadratic, mixture and flexible DA [20]. The main application of DA in medicine is the assessment of severity state of a patient and prognosis of disease outcome. For example, during retrospective analysis, patients are divided into groups according to severity of disease - mild, moderate and severe form. Then, results of clinical and laboratory analyses are studied in order to reveal variables which are statistically different in studied groups. Using these variables, discriminant functions are built which help to objectively classify disease in a future patient into mild, moderate or severe form [23].

Elastic net

A regularized regression technique combining the benefits of a set of other regression techniques (including the Lasso). Elastic net is a versatile method that often outperforms the Lasso [24].

Ensemble method

It combines multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. First introduced by mathematician

Condorcet in 1785 (jury theorem), experimented later by Galton, a contemporary extension is the so-called "wisdom of crowds" or random forest combining multiple decision trees to predict a result [19, 25-27].

Feature

A feature is a characteristic (also called attribute, characteristic, descriptor or variable) in a dataset. In transcriptome analyses, a feature is a gene (i.e.: a raw of data). Collectively, features are the inputs fed to the computer program for generating outputs. In the context of images (e.g.: immunohistochemistry, X-ray, anatomical computed tomography etc.), typical features in breast cancer research are based on tumour intensity, texture and shape [28,29].

Feature selection (or feature extraction or featurization or segmentation)

A process aiming at reducing the complexity of a dataset by decreasing the number of features (for computational purposes or for the sake of interpretation). In oncology, instead of working on a full dataset of 1,000 patients and 20,000 gene transcripts, one may only work on features (e.g.: gene transcripts) exhibiting a large expression change between cancer and healthy subjects. If only 2,500 gene transcripts meet the chosen criteria, the dataset after feature selection would only be 1,000 by 2,500.

Feedforward neural network

A class of artificial neural networks that do not have cycles nor loops (e.g.: convolutional neural networks). In contrast, recurrent (feedback) neural networks exhibit cycles or loops in their architecture. It gives them the ability to 'memorize' parts of the inputs and use them to make accurate predictions for sequential data [30].

Fuzzy C

An algorithm assigning data point membership to one or more cluster(s), in contrast to hard clustering where data points belong completely to just one cluster. In fuzzy C-means, the smaller the distance between the data point and the cluster centre, the stronger the association to the cluster. The most widely used fuzzy clustering algorithm is Fuzzy C-means, partitioning the data into c fuzzy clusters. It was developed by Dunn in 1973, improved by Bezdek in 1981 and is frequently used in pattern recognition tasks (e.g.: image segmentation) [31,32].

Genetic algorithm (GA)

A bio-inspired ML technique that intends to mimic natural selection in order to search a very large solution space efficiently and find an optimal solution to a given problem. GA was first used by John Holland, a pioneer in the study of complex adaptive systems in 1975. The first step is to mutate, or randomly vary, a given collection of a binary strings ("chromosomes"). The second step is a selection step, which is often done through measuring against a fitness function. The evolutionary process is repeated until a suitable solution is found [33,34].

Gigabytes (GB)

Gigabytes ($10e^9$ or GB); Terabytes ($10e^{12}$ or TB); Petabytes ($10e^{15}$ or PB) and Exabytes ($10e^{18}$ or EB) are multiples of the unit byte for digital information. A few orders of magnitude include [35]:

7 GB = How much data we are using per hour when streaming Netflix Ultra HD video,

10 TB = Amount of data produced by the Hubble Space Telescope per year,

24 TB = Amount of video data uploaded to YouTube per day in 2016,

1.5 PB = 10 billion photos on Facebook,

20 PB = The amount of data processed by Google daily in 2008,

15 EB = Total estimated data held by Google.

Graphics processing unit (GPU)

A specialized electronic circuit designed to accelerate the creation of images for output to a display device (e.g.: mobile phones, personal computers, workstations and game consoles). Their highly parallel structure makes them more efficient than general-purpose central processing units (CPUs) for algorithms.

Greedy algorithm

A strategy for reaching quickly a solution to a computer-intensive problem. A greedy algorithm makes the choice that seems to be the best at that moment (selecting a series of locally optimal solutions without going back) in the hope that it will lead to a globally optimal solution. In many problems, a greedy strategy does not produce an optimal solution. Yet, it may approximate a globally optimal solution in a reasonable amount of time.

Hierarchical clustering (HCL)

A clustering method which seeks to build a hierarchy of clusters. The two main are:

i) agglomerative: this is a "bottom-up" approach where each observation starts in its own cluster and pairs of clusters are merged as one moves up the hierarchy

ii) divisive: this is a "top-down" approach where all observations start in one cluster and splits are performed as one moves down the hierarchy. The results of hierarchical clustering are usually presented in a dendrogram [36,37].

K-means (clustering)

One of the most commonly used unsupervised machine learning algorithm for partitioning a given data set into a set of k groups (i.e. k clusters), where k represents the number of groups pre-specified by the data scientist. K-means clustering classifies objects in multiple groups (i.e., clusters), such that objects within the same cluster are as similar as possible (i.e., high intra-class similarity), whereas objects from different clusters are as dissimilar as possible (i.e., low inter-class similarity). In k-means clustering, each cluster is represented by its centre (i.e, centroid) which corresponds to the mean of points assigned to the cluster [38,39].

k-nearest neighor (kNN)

A simple ML technique used for both classification and regression. The kNN algorithm predicts the outcome of the new observation by comparing it to k similar cases in the training dataset, where the value of k is chosen by the data scientist [1,20].

Label

A class where a sample belongs to (synonyms: category, class, tag, sometimes called concepts). In a cancer versus healthy design, a patient will be assigned to one of the 2 possible labels: cancer or healthy.

Learning

Humans learn from previous experience to formulate decisions [2].

Least absolute shrinkage and selection operator (Lasso or LASSO)

A regularization technique for performing linear regression. Lasso was introduced in order to improve the prediction accuracy and interpretability of regression models to reduce the number of predictors in a regression model rather than using all of them. It was developed independently in geophysics and in statistics [40,41].

Long Short-Term Memory networks (LSTM): are a subtype of recurrent neural networks (RNN) introduced by Hochreiter and Schmidhuber in 1997. LSTM are capable of learning long-term dependencies and tend to perform well on a large variety of tasks such as handwriting recognition, speech recognition and generating image descriptions [42,43].

Machine learning (ML)

An interdisciplinary field focused on the study and construction of computer systems that can learn from data without being explicitly programmed. While existing for decades, it is only recently that computing power and data storage improved enough to make it readily accessible. The data is split into training data and test data. Initial data is used to develop the model. The model is a set of rules to predict the dependent variable (y) based on selected independent variables (X) from the dataset. Forms of machine learning are diverse and include regression analysis, clustering, dimensionality reduction, support vector machines, artificial neural networks and decision trees.

Meta

(from Greek, meaning "after" or "beyond") is a recursive prefix implying a circular definition or self-reference. For instance, meta-data are data about data (who has produced them, when, what format the data are in and so on).

Meta-learner

literally “learning to learn” is a learning approach dealing with:

(i) the combination of more than one learning technique to improve learning performances on a specific problem (e.g.: breast cancer diagnosis based on X-ray data

(ii) strategies aiming at enabling a learning program to master more than one problem [44,45].

Model

A simplified, theoretical version of a real phenomenon. A model is constructed with logic rules and/or variables, parameters and constants. Like statistics, construction of the mathematical model relies on two types of variables: independent variables, x, that affect y, the dependent variable one wishes to predict. Based on existing data, machine learning is about crafting a predictive mathematical model that must be accurate, unbiased and robust to provide actionable insights when handling previously unseen data.

Model evaluation

The task of evaluating the performance of a classification model. To this end, a set of metrics are available. They include average classification accuracy (the proportion of correctly classified data points); confusion matrix (a 2x2 table counting the number of true positive, true negative, false positive and false negative cases); precision, recall and specificity; ROC (receiver-operator) curve [20].

Naive Bayes (classifier)

A classification algorithm using Bayes theorem on probabilities, that is the probability of something to happen, given that something else has already occurred [20]. A Naive Bayes classifier computes the probability of an event if every feature being classified is independent from other features. Since features may in fact not necessarily be independent, this algorithm is considered as “naive”. Yet, Naive Bayes classifiers can often outperform more sophisticated algorithms. They are widely used in common applications like spam detection and document classification [46,47].

Neural network (NN)

A network or circuit of neurons. Biological neural networks are made up of real biological neurons whereas artificial neural networks (ANN) are composed of artificial neurons (or nodes) for solving artificial intelligence problems. In the latter, connections between neurons are modelled as weights. A positive weight reflects an excitatory connection whereas negative values mean inhibitory connections. Artificial neurons were proposed in 1943 by Warren McCulloch, a neurophysiologist, and Walter Pitts, a logician as well as by Alan Turing, widely considered to be the father of theoretical computer science and artificial intelligence, in 1948. Although ANN were originally inspired from their biological counterpart, ANN tend to be static and symbolic, while the biological brain of most living organisms is dynamic (exhibits plasticity) and analog [2,20,48].

Occam’s razor

A problem-solving principle that stating that "Entities are not to be multiplied without necessity". Occam’s razor implies that "simpler solutions are more likely to be correct than complex ones.". Thus, having different models that can solve a problem, one should select the solution with the fewest assumptions. The idea is attributed to English Franciscan friar William of Ockham (circa 1287-1347) [49].

Omics

Meaning all, in Greek. Omics deals with techniques, data or studies generated by high-throughput techniques. Examples of omics are genomics (study of the genome), transcriptomics (study of the transcriptome), proteomics (study of the proteome), etc. Although omics intend to monitor all possible biological entities at a given level (e.g.: all proteins at the protein level), only a fraction of them are monitored due to technical limitations and biases.

Ontology

`A controlled vocabulary. This is central for creating categories and for users who want to share information. In Biology, it was largely used in 18th century by Carl Linnaeus, a Swedish botanist, physician, and zoologist who formalized the modern system of naming organisms as well as diseases [50]. More recently, a Gene Ontology (GO) endeavour was launched where characteristics from yeast are transferred to human or mouse [51]. It is widely used in omics to better understand gene function. In computer science, an ontology is a data model representing domain knowledge by describing a set of concepts within a domain and relationships between them. Most of them are based on XML syntax with OWL being a recent version providing many features for ontology development [12,52-54].

Overdetermined

Situation in which there are more equations (constrains) than unknowns. An overdetermined system is almost always inconsistent (it has no solution) when constructed with random coefficients. However, it will have solutions for example if a few equations are linear combinations of the others.

Overfitting

Consists in not being able to generalize beyond what was learned. Overfitting is a major concern and limitation in learning. It occurs when the training dataset is too small and/or not representative enough regarding all possible cases, and/or the complexity of the approach used is too important and should be reduced (following Occam's razor) [49].

Prediction

The expected result concerning an outcome of interest based on a chosen model.

Principal component analysis (PCA)

A linear transformation that projects the data into a new coordinate system such that the greatest variance of the data comes to lie on the first coordinate (called the first principal component or PC1), the second greatest variance on the second coordinate (PC2), the third greatest variance on the third coordinate (PC3), and so on. In practice, for visualization purposes, most analyses use the first two principal components (2D) or the first three (3D) if the data is too complex. PCA is a data reduction technique invented in 1901 by Karl Pearson.

Random forest (RF)

A ML algorithm made of several decision trees (i.e. an ensemble method) developed by Tin Kam Ho in 1995. In a RF, every tree is built from a random selection within the training dataset. A frequent way to obtain the result is by selecting a majority vote of all trees. RF can be used for classification and regression [19].

Recurrent (feedback) neural networks (RNN)

A class of neural networks implementing the fact that understanding is based on previous knowledge. To allow information to persist (dependency), the architecture of RNN has repeating modules with cycles or loops forming a directed graph along a temporal sequence. Cycles or loops allow RNN to exhibit temporal dynamic behaviour and to process sequences of inputs. RNN include Long Short-Term Memory networks (LSTM) [42,43].

Regression

An analysis aiming at fitting a curve (not necessarily a straight line) through a set of data points spread across a continuous scale using a goodness-of-fit criterion. The earliest form of regression was the method of least squares, which was published by Legendre in 1805 and by Gauss in 1809 for determining orbits from astronomical observations. The term "regression" was coined by Francis Galton in the nineteenth century to describe a biological phenomenon: the heights of descendants of tall ancestors tend to regress down towards a normal average (a phenomenon also known as regression toward the mean). The simplest and most common type of regression is linear regression. Until recently, regression was by far the most utilized learning approaches.

Regularization

The process of adding information in order to solve an ill-posed problem or to prevent overfitting. The least-squares method can be viewed as a simple form of regularization to approximate the solution

Sample

In Statistics, a sample is a set (or collection) of data points (or records, cases, observations, statistical unit). In computer science and machine learning, a sample often refers to a single record.

Support vector machine (SVM)

A supervised ML technique that finds the optimal boundary (with margins) separating data points from different groups and then predicts the class of new data points. SVM, developed by Vapnik in 1982, can handle linear or non-linear boundaries, two-class and multi-class classification problems [20].

Tensor

Conceived in 1900 by 2 Italian mathematicians, a tensor is an extension of the concept of scalar (number), vector (column of numbers) and matrix (2-dimensional table of numbers). This multidimensional extension of a matrix accommodates the layers of complexity in deep learning network operations [55].

TensorFlow

Developed by the Google Brain team, TensorFlow is a library for using ML algorithms including neural networks. It was released as open source on November 9, 2015. TensorFlow can run on multiple CPUs, GPUs and TPUs.

Tensor processing unit (TPU)

An application-specific electronic circuit built specifically for machine learning and tailored for TensorFlow. TPU is a programmable AI accelerator that delivers a better-optimized performance per watt for machine learning applications.

t-Distributed Stochastic Neighbor Embedding (t-SNE)

An unsupervised dimensionality reduction technique that minimizes the divergence between two distributions: the pairwise similarity of the data points in the higher-dimensional space and the pairwise similarity of the data points in the lower-dimensional space. Developed by Laurens Van Der Maaten 2008, it has been recently applied for single-cell RNA-sequencing data [56].

Triple negative breast cancer (TNBC)

A breast cancer subtype representing about 15% to 20% of breast cancers. TNBC is a breast cancer subtype lacking expression of estrogen receptor (ER) and progesterone receptor (PR), and that does not overexpress human epidermal growth factor receptor 2 (HER2). TNBC remains a clinical challenge with poor prognosis since no therapeutic targets have been identified.

Underdetermined

When there are fewer equations than unknowns (in contrast to an overdetermined situation). Each variable (e.g.: gene) is providing a degree of freedom. On the other hand, each introduced equation can be viewed as a constraint that restricts one degree of freedom. Ideally, the number of equations and the number of variables is equal. Thus,

to every variable giving a degree of freedom correspond a constraint removing a degree of freedom. In the underdetermined case, the unknowns outnumber the equations so that many possible solutions exist. To prevent overfitting, it is then recommended to use the simplest possible solutions (so called Occam's razor) [49].

Conclusion

Applications of AI and ML are becoming quite popular, including in the cancer research community. However, at least 2 major limits apply to AI and ML. One, they are brain children of humans and are thus by definition limited in their scope. Their limits and pitfalls need to be clearly assessed. Two, most AI and ML approaches currently require large energetic costs that are well beyond the 20 Watt per hour of our human brain. In a world with limited resources, efforts intending to decrease energetic costs are also greatly needed.

Funding

This work was supported by SUNY EIPF grant #172, the Région Bourgogne Franche-Comté PARI (grant number 9201AAO050S01716), Ligue contre le Cancer (grant number R18032MM) and Nano2Bio and FEDER (grant number BG0005900). No funding sources were involved in the study design, collection, analysis, interpretation of data, writing or in the decision to submit the manuscript for publication.

References

1. Bhargava A (2016) Grokking Algorithms: An illustrated guide for programmers and other curious people. Manning Publications.
2. Theobald O (2017) Machine Learning for Absolute Beginners: A Plain English Introduction.
3. Beach, CSUL.
4. Julia L (2019) L'intelligence artificielle n'existe pas. 1st edition.
5. Hinton G (2018) Deep Learning-A Technology with the Potential to Transform Health Care. *JAMA* 320: 1101-1102. [Crossref]
6. https://en.wikipedia.org/wiki/Paul_Werbos
7. Rumelhart D, Hinton G, Williams R (1986) Learning representations by back-propagating errors. *Nature* 323: 533-536.
8. https://hadoop.apache.org/
9. Ghemawat S, Gobioff H, Leung ST (2003) The Google File System. SOSP Bolton Landing.
10. Dean J, Ghemawat S (2004) MapReduce: Simplified Data Processing on Large Clusters. OSDI.
11. Kolodner J, Hmelo C, Narayanan NH (1996) Problem-Based Learning Meets Case-Based Reasoning. *J Learn Sci* 12: 188-195.
12. Aamodt A, Plaza E (1994) Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *IOS Press* 7: 39-59.
13. Begum S (2011) Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments. IEEE Transactions on systems, man, and cybernetics.
14. Bahls D, Roth-Berghofer T (2007) Explanation Support for the Case-Based Reasoning Tool myCBR. Association for the Advancement of Artificial Intelligence, 1844-1845.
15. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press.
16. Usama MF, Gregory PS, Padhraic S (1996) From data mining to knowledge discovery: an overview, in Advances in knowledge discovery and data mining. American Association for Artificial Intelligence pp. 1-34.
17. https://www.ncbi.nlm.nih.gov/geo/
18. https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
19. Smith C (2017) Decision Trees and Random Forests: A Visual Introduction for Beginners. Blue Windmill Media.
20. Kassambara A (2018) Machine Learning Essentials: Practical Guide in R. CreateSpace Independent Publishing Platform.
21. Naylor CD (2018) On the Prospects for a (Deep) Learning Health Care System. *JAMA* 320: 1099-1100.
22. http://neuralnetworksanddeeplearning.com/
23. https://en.wikipedia.org/wiki/Linear_discriminant_analysis
24. Zou H, Hastie T (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society* 67(Series B): 301-320.
25. Polikar R (2006) Ensemble based systems in decision making. IEEE Circuits and Systems Magazine 6: 21-45.
26. Rokach L (2010) Ensemble-based classifiers. *Artif Intell Rev* 33: 1-39.
27. Opitz D, Maclin R (1999) Popular ensemble methods: An empirical study. *J Artif Intell Res* 11: 169-198.
28. Lambin P (2012) Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48: 441-446. [Crossref]
29. Aerts HJ (2014) Decoding tumour phenotype by non-invasive imaging using a quantitative radiomics approach. *Nat Commun* 5: 4006.
30. https://towardsdatascience.com/learn-how-recurrent-neural-networks-work-84e975feaaf7
31. https://reference.wolfram.com/legacy/applications/fuzzylogic/Manual/12.html
32. Jipkate B, Gohokar V (2012) A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms. *Int J Comput Eng Res* 2: 737-739.
33. http://mathworld.wolfram.com/GeneticAlgorithm.html
34. Mitchell M (1998) An Introduction to Genetic Algorithms. MIT Press.
35. https://www.howtogeek.com/353116/how-big-are-gigabytes-terabytes-and-petabytes/
36. Eisen MB (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95: 14863-14868.
37. Spellman PT (1998) Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell* 9: 3273-3297. [Crossref]
38. https://www.datanovia.com/en/lessons/k-means-clustering-in-r-algorith-and-practical-examples/
39. MacQueen J (1967) Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, (University of California Press): 281-97.
40. Santosa F, Symes W (1986) Linear Inversion of Band-Limited Reflection Seismograms. *SIAM J Sci Comput* 7: 1307-1330.
41. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Royal Stat Soc* 58: 267-288.
42. https://colah.github.io/posts/2015-08-Understanding-LSTMs/
43. http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/
44. Finn C, Abbeel P, Levine S (2017) Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. arXiv:1703.03400.
45. Schmidhuber J (1987) Evolutionary principles in self-referential learning, or on learning. (Diploma thesis).
46. Morris D (2017) Bayes' Theorem Examples: A Visual Introduction for Beginners. Blue Windmill Media.
47. https://www.datasciencecentral.com/profiles/blogs/naive-bayes-for-dummies-a-simple-explanation
48. Sullivan W (2018) Machine Learning Algorithms for Supervised and Unsupervised Learning: The Future Is Here, (2nd edn), CreateSpace Independent Publishing Platform.
49. Seigneuric R (2009) Systems Biology Applied to Cancer Research, in Handbook of Research on Systems Biology Applications in Medicine. In: Daskalaki A (eds) Medical Information science reference, Hershey, New York, USA pp.339-353.
50. https://www.nytimes.com/2008/05/06/health/research/06dise.html

51. Ashburner M (2008) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25-29. [Crossref]

52. Draghici S (2002) Statistical intelligence: effective analysis of high-density microarray data. *Drug Discov Today* 7: 55-63. [Crossref]

53. Supek F (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6: e21800. [Crossref]

54. Eden E (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10: 48. [Crossref]

55. http://mathworld.wolfram.com/

56. van der Maaten L, Hinton G (2008) Visualizing High-Dimensional Data Using t-SNE. *J Machine Learning Res* 9: 2579-2605.