Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 22 Issue 1 Version 1.0 Year 2022
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Review on the Application of Machine Learning to Cancer Research

By Eunice C. Chibudike, Henry O. Chibudike, Nwaebuni E. Odega, Emeka E. Njokanma, Olubamike A. Adeyoju & Constance O. Ngige

Federal Institute of Industrial Research

Abstract- This study reviews the application of machine learning through different algorithms in cancer research. In recent years, the introduction of machine learning has been an exciting tool that enhances cancer research which has improved statistical method of speeding up both fundamental and applied research considerably.

The application of machine learning goes around in predicting the future events and outcomes with the available datasets. There is an indication that on yearly bases up to 14 million new cancer patients are diagnosed by Pathologists round the world, and they are people whose conditions are uncertain. Definitely, the diagnoses and prognoses of cancer have been performed by Pathologists. The research on machine learning flourished in 1980s and 1990s and information become digitalized through improved artificial network connectivity and computational power.

Keywords: machine learning, cancer research, cancer diagnoses, cancer predicting, and diagnosis.

GJCST-C Classification: F.1.1

Strictly as per the compliance and regulations of:

Review on the Application of Machine Learning to Cancer Research

Eunice C. Chibudike [α], Henry O. Chibudike [σ], Nwaebuni E. Odega [ρ], Emeka E. Njokanma [ω], Olubamike A. Adeyoju [¥] & Constance O. Ngige [§]

Abstract- This study reviews the application of machine learning through different algorithms in cancer research. In recent years, the introduction of machine learning has been an exciting tool that enhances cancer research which has improved statistical method of speeding up both fundamental and applied research considerably.

The application of machine learning goes around in predicting the future events and outcomes with the available datasets. There is an indication that on yearly bases up to 14 million new cancer patients are diagnosed by Pathologists round the world, and they are people whose conditions are uncertain. Definitely, the diagnoses and prognoses of cancer have been performed by Pathologists. The research on machine learning flourished in 1980s and 1990s and information become digitalized through improved artificial network connectivity and computational power. This shifted the effect of machine learning from artificial intelligence to solving practically natural problems. From there, shortly the potential of machine learning became obvious in medical science by scientists and gained its ground in medical specialties such as radiology, cardiology, mental health and pathology. In health care machine learning is used to interpret data hence speed up workflow, reduce medical error and promote human health. Pathologists are accurate at diagnosing cancer but have an accuracy rate of only 65% when predicting the development of cancer. Computed tomography, mammography, magnetic resonance imaging (MRI), or histopathology have been derived from imaging datasets over decades for diagnoses and staging prognosis of various cancers. The development of novel computational tools for stratification, grading, prognostication of patients with the goal of improving patient care has been achieved through the impact of machine learning.

Keywords: machine learning, cancer research, cancer diagnoses, cancer predicting, and diagnosis.

Author α §: Planning, Technology Transfer and Information Management, Federal Institute of Industrial Research, Oshodi, F.I.I.R.O., Lagos, Nigeria.

Author σ: Department of Chemical, Fiber and Environmental Technology, Federal Institute of Industrial Research, Oshodi, F.I.I.R.O., Lagos, Nigeria.

Author ρ: Nigerian Upstream Petroleum Regulatory Commission (NUPRC).

Author ω: Chevron Nigeria Limited.

Author ¥: Production, Analytical and Laboratory Management, Federal Institute of Industrial Research, Oshodi, F.I.I.R.O., Lagos, Nigeria.

e-mail: henrychibudike@gmail.com

I. Introduction

In recent years, the availability of large datasets combined with the improvement in algorithms and the exponential growth in computing power led to an unparalleled surge of interest in the topic of machine learning *(Khan Academy, 2018)*. Nowadays, machine learning algorithms are successfully employed for classification, regression, clustering, or dimensionality reduction tasks of large sets of especially high-dimensional input data *(Sunil Ray, 2017)*. In fact, machine learning has proved to have superhuman abilities in numerous fields (such as prediction, self-driving cars, image classification, 4 medical diagnoses etc.). As a result, huge parts of our daily life, for example, image and speech recognition, web-searches, fraud detection, email/spam filtering, credit scores, report extraction and many more are powered by machine learning algorithms *(Jonathan Schmidt, et al; 2019)*.While data-driven research and more specifically machine learning, have already a long history in biology or chemistry, they only rose to prominence recently in the field of cancer research. A first computational revolution in cancer research was fueled by the advent of computational methods, especially magnetic resonance imaging (MRI) *(Mandeep Kaur 2019)*. The constant increase in computing power and the development of more efficient codes also allowed for computational high-throughput studies of large samples in order to screen for the ideal experimental candidates.

Over decades, cancer researchers have researched into cancer to identify causes and dive into measures for its prevention, diagnosis, treatment and cure. The epidemiology, molecular bioscience to the performance of clinical trials have been evaluated and compared for the application of their various treatments; *(Susan A. Nadin-Davis, in Rabies (Second Editon), 2007)*. It could be applied in surgery, immunotherapy, hormone therapy, chemotherapy, radiation therapy and combined treatment modalities such as chemo-radiotherapy. In the mid-1990s the clinical cancer research shifted to therapies and this was derived from biotechnology research such as immunotherapy and gene therapy. Cancer research is done in academia, research institutes, and corporate environments, and is largely government funded, according to Martin Stumpe (AI and Data Science, MI, USA 2019), and collaborators

developed a deep-learning system (DLS) 2019. However, the challenges and interesting tasks of physicians are the accurate prediction outcomes of diseases. For this reason, Machine Learning methods have taken over in medical research as a popular tool. This review has an indication of some of the models that have been developed for cancer biopsies and prognoses. For instance, there a model that predicts cancer susceptibility; Craig Mermel (Google AI Healthcare, CA, USA 2019). The model was built to discriminate tumors as either malignant or benign in the midst of breast cancer patients. In this model, the completion of the tasks was done by ANN.

The building of this model was with a large number of hidden layers that could generalize data better. As thousands of mammographic data were fed in the model to obtain and learn the difference between benign and malignant tumors. Before being inputted, all the data was reviewed by radiologists. An approach by Regina Barzilay (MGH, MA, USA) 2019. The causes of cancer have been researched into many different disciplines including genetics, diet, environmental factors (i.e. chemical carcinogens). During the investigation of causes and also potential therapy targets, the route with data derived from clinical observations, basic research commences, and once convinced and independently obtained results are confirmed, proceeds with clinical research, which involves appropriate designed trials on consenting human subjects, with the goal to ascertain safety and efficiency of the therapeutic intervention method; Connie Lehman at Massachusetts General Hospital (MGH, MA, USA) 2019. One of the important parts of basic research is characterization of the potential mechanisms of carcinogenesis, having in mind the types of genetic and epigenetic changes that are associated with cancer development. The use of mouse is like a model for mammalian manipulation of the function of genes that play a role in tumor formation, while basic aspects such as bacteria and mammalian cells are assayed on cultures for tumor initiation, such as mutagenesis.

II. Methodology

Image filtering: In this review we examined a few of the most widely used image processing algorithms, then move on to machine learning implementation in image processing. At a glance is as follows:

- Feature mapping using the scale-invariant feature transform (SIFT) algorithm.
- Image registration using the random sample consensus (RANSAC) algorithm.
- Image Classification using artificial neural networks.
- Image classification using convolutional neural networks (CNNs).
- Image Classification using machine learning.
- Important Terms

Dynamic Contrast enhancement: Conventional contrast-enhanced magnetic resonance imaging (MRI) displays a single snapshot of tumor enhancement after contrast administration; although the anatomical information derived from such images is valuable, it lacks functional information ((National Institute of Health, 2017)). Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), which relies on fast MRI sequences obtained before, during and after the rapid intravenous (IV) administration of a gadolinium (Gd) based contrast agent is analogous to a movie and is an emerging imaging method to assess tumor angiogenesis. To investigate whether a combination of radionics and automatic machine learning applied to dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) of primary breast cancer can non-invasively predict axillary sentinel lymph node (SLN) metastasis

Image Segmentation and Radiomic Feature Extraction: Axial DCE-MRI Digital Imaging and Communications in Medicine (DICOM) images were archived from the Picture Archiving and Communication System (PACS) (Stefan Leger et al 2019). The calculation of time signal intensity curves for tumor lesions in the DCE-MRI images were done using a GE Advanced Workstation ADW4.4 (Jan C. Peeken et al; 2019). Based on these curves, the volumes of interest (VOIs) were delineated on the whole tumor in the images with the strongest enhanced phase. The VOIs were determined manually by a radiologist with 10 years of experience who was blinded to the clinical information of the patients, and all contours were reviewed by another senior radiologist with 20 years of experience (Pan Sun et al 2019). If the discrepancy was ≥5%, the senior radiologist determined the tumor borders. Cohen's kappa method was used to assess inter-reader agreement (lanna Vial1 et al, 2018). In general, the (pre- processing of images are often the first step to later extraction of the features that would be used to train a machine learning classifier. Signal processing can be used to improve or eliminate properties of the image that could enhance the performance of the machine learning algorithm.

Classification of effectiveness of model: In machine learning, classification models are often used to get a predicted result of population data. Classification is one of the two sections of supervised learning deals with data from different categories (Manojit Chattopadhyay et al; 2017). The training dataset trains the model to predict the unknown labels of population data. There are multiple algorithms, namely, Logistic regression, K-nearest neighbour, Decision tree, Naive Bayes etc. All these algorithms have their own way of execution and different methods of prediction. But, at the end, we need to find the effectiveness of an algorithm (qbal H. Sarker, et al; 2019). To find the most suitable algorithm for a particular problem, there are model evaluation techniques. In this article several model evaluation techniques will be discussed.

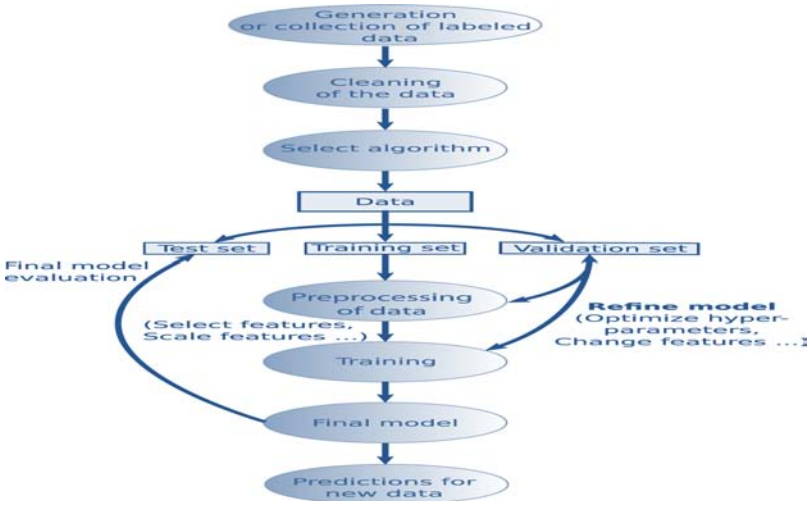

Figure 1: Supervised learning workflow(Jonathan Schmidt et al; 2019)

Figure 1: Depicts the workflow applied in supervised learning. One generally chooses a subset of the relevant population for which values of the target property are known or creates the data if necessary. This process is accompanied by the selection of a machine learning algorithm that will be used to fit the desired target quantity (Jonathan Schmidt et al; 2019).

a) Evaluation

One of the evaluations to conduct during prediction is Confusion matrix in the image. A confusion matrix is a table that describes the performance of a classifier/classification model. It contains information about the *actual and prediction classifications* done by the classifier and this information is used to evaluate the performance of the classifier. Here is the sample of a *Confusion Matrix (Banso D. Wisdom 2017).*

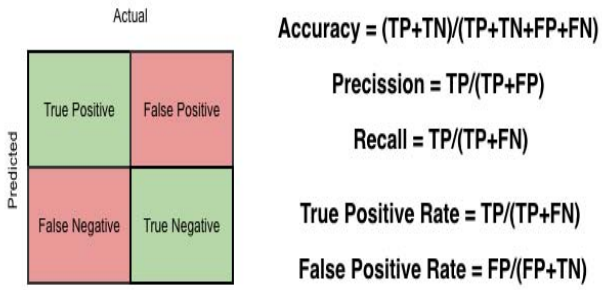

Figure 2: Confusion Matrix (Banso D. Wisdom 2017)

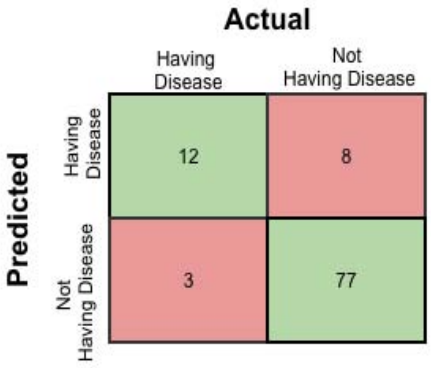

Figure 3: Confusion Matrix of predicting a disease

Confusion matrix is the image given above. This is a matrix representing the results of any binary testing. For example, let us take the case of predicting a disease. You have done some medical testing and with the help of the results of those tests, you are going to predict whether the person is having a disease. So, actually you are going to validate if the hypothesis of declaring a person as having disease is acceptable or not. Say, among 100 people you are predicting 20 people to have the disease. In actual only 15 people to have the disease and among those 15 people you have diagnosed 12 people correctly. So, if I put the result in a confusion matrix, it will look like the following.

As observed in figure 3

True Positive: 12 (You have predicted the positive case correctly!)

1. *True Negative:* 77 (You have predicted negative case correctly!)
2. *False Positive:* 8 (You have predicted these people as having disease, but in actual they do not have. There is no course for alarm; this can be rectified during further medical analysis. So, this is a low risk error. This is type-II error in this case.)
3. *False Negative*: 3 (You have predicted these three poor fellows as fit. But actually they have the disease. This is dangerous! Be careful! This is type-I error in this case.)

Now, this is the accuracy of the prediction model was followed to get this results; i.e the ratio of the accurately predicted number and the total number of people which is (12+77)/100 = 0.89. There is need for you to study the confusion matrix thoroughly so as to find the following things.

b) Test for specificity and sensitivity

In medical diagnosis, the term test sensitivity is the reliability of a test to correctly identify those affected with the disease (true positive rate), while test specificity is the ability of the test to correctly identify those that are not affected with the disease (true negative rate).

Table 1: Test for specificity and sensitivity (Dr. Aaron Swanson, 2011).

	Sensitivity	Specificity
Definition	Proportion of patients with a disease who test positive	Proportion of patients without the disease who test negative
100% (1.0) Means	The test correctly identify every person who has the target disorder	The test correctly identify every person who does not have the target disorder
Statistical Outcome	True Positive	True Negative
Ideal Test Result	Negative Test Result	Positive Test Result
Test Interpretation	They are definitely not positive → They DON'T have it	They are definitely not negative → They DO have it
The Rule	Rule Out (SnOut)	Rule In (SpIn)

Sensitivity and specificity are great values to lead you in your fair clinical examination. It gives more information regarding the patient and guide to a better assessment and authentic diagnosis. Keep in mind that there is always the possibility of false positives and negatives. Special tests should never be the only sign to determine a patient's pathology. It is merely a piece of the clinical examination and assessment (Dr. Aaron

Table 2: Attribute Information Swanson, 2011).

Sample code number	Id number
Clump Thickness	1 – 10
Uniformity of Cell Size	1 – 10
Uniformity of Cell Shape	1 – 10
Marginal Adhesion	1 – 10
Single Epithelial Cell Size	1 – 10
Bare Nuclei	1 – 10
Bland Chromatin	1 – 10
Normal Nucleoli	1 – 10
Mitoses	1 – 10
Class	(2 for benign, 4 for malignant)

III. Results and Discussion

a) Parameters for cancer dictation

In the development of metastases there is a negative prognostic parameter for the clinical result of breast cancer. Bone consists of the first site of distant metastases for several affected women. The idea of this attribute information is to perform an exploratory analysis of the information contained in the dataset, figuring out ways of making the dataset tidier. The ultimate objective is to, in the end, build and compare models to predict if a given tumor is benign or malignant (breast cancer) using the information available on the dataset in Table 2 below.

Table 3: A Sample of Analysis and Modeling of Breast Cancer Data (Random Forest model) from (*ml-repository* '@' *ics.uci.edu)*.

The analysis shows that, with a Random Forest model, we can predict if a given tumor is malignant with 97.86% of Accuracy. This result is 1.96% higher than the Accuracy of 95.90% reported in the UCI Machine Learning as the highest for this dataset (*ml-repository* '@' *ics.uci.edu*). We also conclude that the most important information for this prediction is the 'uniformity of the cell size'. The idea is to perform an exploratory analysis of the information contained in the dataset, figuring out ways of making the dataset tidier. The ultimate objective is to, in the end, build and compare models to predict if a given tumor is available on this dataset. The analysis show that, with a Random Forest model, we can predict if a given tumor is malignant or benign for (breast cancer) using the information (*ml-repository* '@' *ics.uci.edu*).

Table 4: A sample of Dataset (ml-repository '@' ics.uci.edu)

ID	ID number	Radius mean	Texture mean	Perimeter mean	Smoothness mean	Compactness mean	Concavity mean	Concave points mean	Symmetry mean	Fractal dimension mean	Diagnosis
3	842302	19.69	10.38	122.8	0.118	0.2776	0.3001	0.1471	0.2419	0.07871	1
4	842517	20.57	17.77	132.9	0.085	0.07864	0.0869	0.07017	0.1812	0.05667	1
5	84300903	19.69	21.25	130	0.11	0.1599	0.1974	0.1279	0.2069	0.05999	1
6	84348301	11.42	20.38	77.58	0.143	0.2839	0.2414	0.1052	0.2597	0.09744	1
7	84358402	20.29	14.34	135.1	0.1	0.1328	0.198	0.1043	0.1809	0.05883	1
8	843786	12.45	15.7	82.57	0.128	0.17	0.1578	0.08089	0.2087	0.07613	1
9	844359	18.25	19.98	119.6	0.095	0.109	0.1127	0.074	0.1794	0.05742	1
10	84458202	13.71	20.83	90.2	0.119	0.1645	0.09366	0.05985	0.2196	0.07451	1
11	844981	13	21.82	87.5	0.127	0.1932	0.1859	0.09353	0.235	0.07389	1
12	84501001	12.46	24.04	83.97	0.119	0.2396	0.2273	0.08543	0.203	0.08243	1
13	845636	16.02	23.24	102.7	0.082	0.06669	0.03299	0.03323	0.1528	0.05697	1
14	84610002	15.78	17.89	103.6	0.097	0.1292	0.09954	0.06606	0.1842	0.06082	1
15	846226	19.17	24.8	132.4	0.097	0.2458	0.2065	0.1118	0.2397	0.078	1
16	846381	15.85	23.95	103.7	0.084	0.1002	0.09938	0.05364	0.1847	0.05338	1
17	84667401	13.73	22.61	93.6	0.113	0.2293	0.2128	0.08025	0.2069	0.07682	1
18	84799002	14.54	27.54	96.73	0.114	0.1595	0.1639	0.07364	0.2303	0.07077	1
19	848406	14.68	20.13	94.74	0.099	0.072	0.07395	0.05259	0.1586	0.05922	1
20	84862001	16.13	20.68	108.1	0.117	0.2022	0.1722	0.1028	0.2164	0.07356	1

The diagnosis of breast tissue(1 = malignant, 0 = benig)

b) Datasets and their Features

In table 4 above, when it comes to classification, there is a need of dataset to classify. Dataset is a statistical matrix which represents different features. It is a matrix where all the information about different features is given. Each column of the dataset represents the feature of the tumorous tissue and each row represents the number of instances. Table 4 is the details of attributes found in *WDBC dataset (19) (Vania V Estrela et al; 2019):* ID number, Diagnosis (M=Malignant, B=Benign) and ten real real-valued features are computed for each cell nucleus: radius, Texture, Perimeter, Area, Smoothness, Compactness, Concavity, Concave points, Symmetry and Fractal dimension *(20) (Anirban Banerji, 2013*). These features are computed from digitized image of a fine needless aspirate (FNA) of a breast mass *(ml-repository* '@' *ics.uci.edu).* They described characteristics of the cell nuclei present in the *image (21)(Tula Neilson 2012).* When the radius of an individual nucleus measured by averaging the length of the radial line segments, it is defined by the centroid of the snake and the individual snake points. The Nuclear Perimeter constitutes of the total distance between consecutive snake points.

c) *Exploratory Analysis*

To explore this data and later also be able to create models correctly, we need to separate our data into *train* and *test* data. This is to achieve a simulated real world dataset (test) that have class information that has not been used in anyway during the analysis (instead we use train). This ensures that our test dataset is really simulating real world data, since it has not been seen during exploration or modeling (*Prasad Patil 2018)* For this purpose, the R package caTools, as displayed below

```
library (caTools)
set. Seed(1000)
split=sample. Split (cancer$Class, Split Ratio=0.80)
train=subset(cancer, split==TRUE)
test=subset(cancer,split==FALSE)
```

d) *Cancer Research*

Cancer research is a research into the cause of cancer, prevention, diagnosis, treatment and cure which involves many diverse disciplines including genetics, diet, environmental factors (i.e. chemical carcinogens). The ranges of cancer research are from epidemiology, molecular bioscience to the performance of clinical trials to make evaluation and comparison of the application of various cancer treatments (Douglas Hanahan et al 2011). Cancer research has been on for ages. In the early years of research, the focus was on the causes of cancer. The first identification of environmental trigger (chimney soot) for cancer was PercivallPott in 1775 and identification of cigarette smoking as a cause of lung cancer in 1950. The treatment of cancer was early focused on enhancing surgical techniques for removing tumors. Radiation therapy took hold in the 1900s (Douglas Hanahan et al; 2011.) The development and definition of Chemotherapeutics were done throughout the 20th century. Cancer research involves various types and interdisciplinary areas of research. Scientists in cancer research may get their trainings in areas such as epidemiology, chemistry, biomedical engineering, molecular biology, medical physics, physiology and biochemistry. Research principles and mechanisms were always clarified at basic research level. Translational search aims to discover the mechanisms of cancer development and progression and convert n basic scientific results into ideas that can be applied to the treatment and prevention of cancer (The Hallmarks of Cancer, published in 2000). The development of pharmaceuticals, surgical procedures, and medical technologies for the eventual treatment of patients are achieved through clinical research

e) *Genes involved in cancer*

The aim of *oncogenomics* is to discover new *oncogenes* or *tumor suppressor genes* that may provide new knowledge into diagnosing cancer, predicting clinical outcome of cancers, and an update targets for cancer therapies *(John Carpten et al on RNASEL: M. Sprinsky/AACR 2002)*. As the Cancer Genome Project stated in a 2004 review article, "a central aim of cancer research has been to identify the mutated genes that are causally implicated in oncogenesis (cancer genes). The project of Cancer Genome Atlas is a related effort which focused in investigating the genomic changes that relates to cancer, while the genetic mutations from hundreds of thousands of human cancer samples were acquired from COSMIC cancer database documents. In the cause of several literature reviews, there is an indication that projects have been carried out, involving about 350 different types of cancer, have identified ~130,000 mutations in ~3000 genes that have been mutated in the tumors. The majority occurred in 319 genes, of which 286 were tumor suppressor genes and 33 oncogenes *(American Association for Cancer Research, Databases for oncogenomic research)*. Some hereditary factors can shoot up the chance of cancer-causing mutations that includes activating oncogenes or inhibiting tumor suppressor genes. The functioning of various oncogenes and tumor suppressor genes can be interrupted at different levels of tumor progression. Gene's mutations can be used to classify the malignancy of a tumor. In some stages, tumors can form a resistance to cancer treatment. The understanding of tumor progression and treatment success is achieved when identification of oncogenes and tumor suppressor genes done. The function of a given gene in cancer progression may differ tremendously, as it depends on the stage and type of cancer involved (*the National Cancer Institute 2017)*

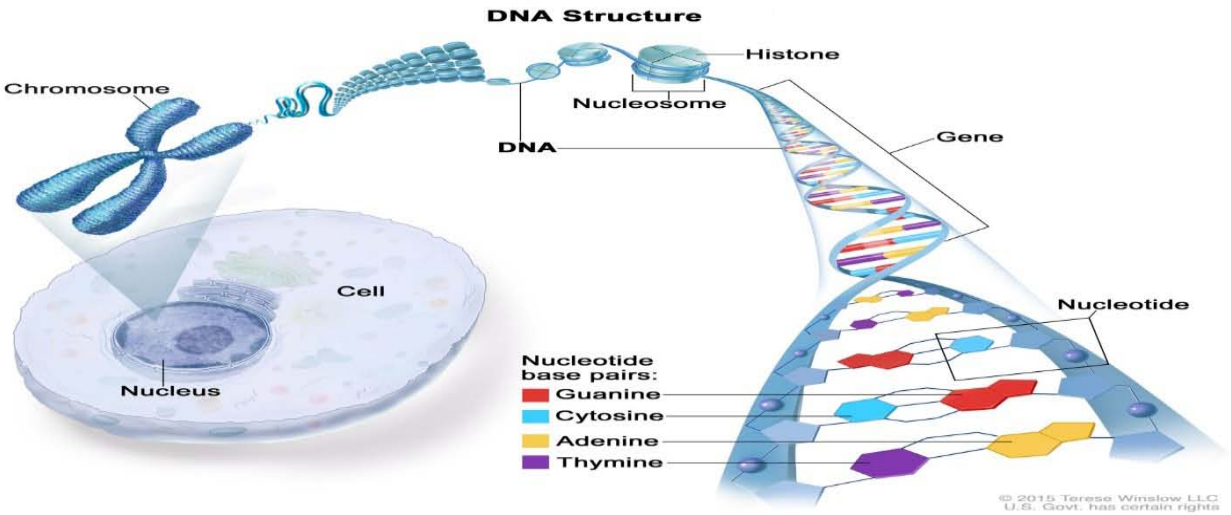

Diagram: Structure of DNA (*the National Cancer Institute 2017*)

It has been ascertained that most DNA is seen inside the nucleus of a cell, where it forms the chromosomes. Chromosomes acquire proteins called histones that join to DNA. DNA has two strands that fickle into the shape of a spiral ladder called a helix. DNA is made up of four building blocks called nucleotides: adenine (A), thymine (T), guanine (G), and cytosine (C) *(the National Cancer Institute 2017)*. The nucleotides attach to each other (A with T, and G with C) to form chemical bonds called base pairs, which connect the two DNA strands *(the National Cancer Institute 2017).* Genes are short pieces of DNA that carry specific genetic information *(the National Cancer Institute 2017).*

IV. Cancer Detection

It is advisable to dictate cancer early so as to avert the difficulty of treating it in later stages. Accuracy in detection of cancer is paramount because false positives can cause harm owing to unnecessary medical procedures *(Joensuu, Heikki et al; 2013)*. Some screening procedures are not accurate currently (such as prostate-specific antigen testing). In some other cases like a colonoscopy or mammogram are unpleasant and gives room for some patients to opt out. Active research is to address all these problems *(Andrew Mckeon et al; 2016).*

Three main ways cancer cells can spread.

1. Through the blood vessels: This is known as hematogenous spread. Cancerous cells invade blood vessels and use the flow of blood cells as transportation.
2. Through nearby tissue: This is known as transcoelomic spread. Cancerous cells penetrate the surfaces of peritoneal cavities in the body.
3. Through the Lymphatic system: This is known as lymphatic spread. Cancerous cells invade the lymph nodes and use the lymphatic system to travel.

V. Application of Machine Learning to Cancer Research

There are two ways to cancer, Prediction/ Prognosis and Detection/Diagnosis. *In cancer Prediction/ Prognosis there are three core points:*

- *Prediction of cancer susceptibility (i.e. risk assessment)*
- *Prediction of cancer recurrence*
- *Prediction of cancer survivability*
- Risk assessment is predicting the probability of developing a type of cancer prior to the occurrence of the disease *(Wolters Kluwer Health, Inc. 2003).* The prediction of cancer recurrence is about trying hard to discover the likelihood of re-developing cancer after to the apparent resolution of the disease. The predicting of cancer survivability is determining outcomes like life expectancy, progression, survivability, tumor-drug sensitivity after the diagnosis of the disease *(Joseph A. Cruz, 2006)*. The quality of the diagnosis and other factors determines the success of the prognostic prediction. However, a medical diagnosis and a prognostic prediction must take into account more than just a simple diagnosis before disease prognosis can takes place.

Experts in cancer research have already compiled a list of features to dictate cancer cells, which is preferably to adding chemicals to blood samples that destroys cells *(Cancer Informatics 2006: 2 59–78)*. Refractive indices is an example of what data is used to help the machine predict and diagnose cancer and by using that, it tells us how much light slows down when passing through cells. It helps in light absorption, scattering properties as well as morphological features *(Joseph A. Cruz et al; 2006)*. The input is an image, then the neural networks help identify the cancer cells by learning the relationships of what values of the features

leads to cancer cells. The deep learning algorithm makes use of these features to classify cells based on learning the values of each feature that leads to a cancerous cell. Metastasized detection requires highly.

a) *Artificial Intelligence*

Artificial Intelligence manages more comprehensive issues of automating a system. This computerization should be possible by utilizing anything any field such as image processing, cognitive science, neural systems, machine learning etc. Most recent updates in Artificial Intelligence (AI) are due to application of machine learning to very large data sets. Artificial Intelligence is when computer algorithm does intelligent work. Artificial intelligence is the superset of machine learning i.e all the machine learning is artificial intelligence but not all the AI is machine learning. *Machine learning (ML*) manages and influences user's machine to gain from the external environment. This external environment can be sensors, electronic segments, external storage gadgets and numerous other devices. Machine Learning enables computers to learn by themselves. With the aid of modern computers, it is easier manipulating large data sets *(Japan AI Experience in 2017)*. The algorithms detect patterns and learn ways to make predictions and recommendations by processing data and experiences, instead of being explicitly written in program. *Machine Learning* is made up of three major types: *Supervised* In which data is labeled. The model is to identify the labels and put them in groups accordingly. In other words, the input is provided to the model and the desired output is offered. This process is done countless times until the desired output is obtained. *Unsupervised* In which data is not labeled *(B.J. Copeland; 2019)*. Different features and classifications have to be identified based on the distinct characteristics through the model. In this case, the input is given, but there is no expected output. The logical classifications or groupings are made by computer. *Reinforcement*: This learning treats the problem of finding optimal or sufficiently good actions for a situation in order to maximize a reward. In other words, it learns from interactions *(JAKE FRANKENFIELD; 2019).*

VI. Algorithm for Cancer Cell Development Prediction

Machine learning algorithms have already revolutionized other fields, such as image recognition. However, the development from the first perception up to modern deep convolutional neural networks was a long and tortuous process. In order to produce significant results in cancer research, one necessarily has not only to play to the strength of machine learning techniques but also apply the lessons already learned in other fields (*Konstantina Kourou et al; 2015).*

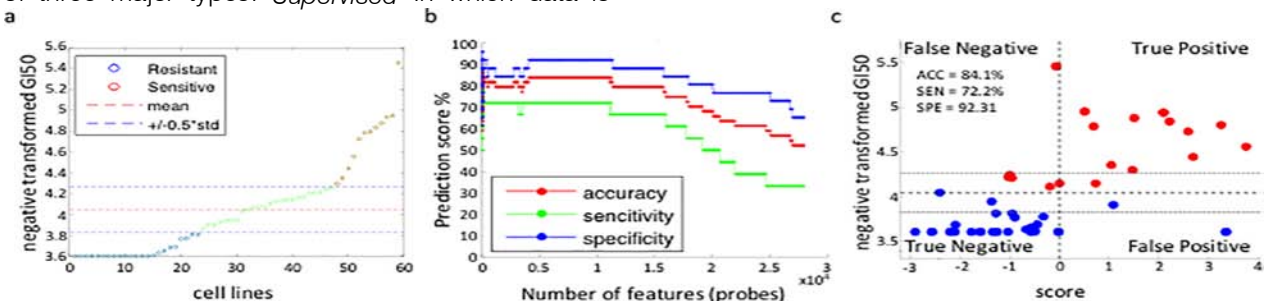

Figure 1: A Sample Algorithm for Cancer Prediction (*Konstantina Kourou[a] et al; 2015).*

VII. Conclusion

In conclusion, there has been an estimated 100 plus types of cancerous cells. This imposes difficulty in curing cancer. For example, if a certain group of similar cancer cells accepts a particular drug or treatment, it could have a peculiar different effect on another group.

- Skilled pathologists or radiologists that perform manual segmentation, which is time-consuming and prone to error, particularly in cases where tumors are few or there are no tumors. Deep learning networks have significantly enhanced accuracy on a wide range of computer vision tasks such as object detection, image recognition, and semantic segmentation.
- Machine learning is now a veritable tool used in cancer research labs to classify tumors based on growth characteristics; features such as where they grow, how fast they grow and size etc. and they are classified into groups based on similar range of predictive outcomes. The reason being that, one can create a controlled environment by picking a classified group and perform desired experiments to see the effect.

References Références Referencias

1. W.H. Wolberg, &, O.L. Mangasarian `(1990). In Proceedings of the National Academy of Sciences, 87, 9193--9196

2. J. Zhang, (1992). Selecting typical instances in instance-based learning. (pp. 470--479). Aberdeen, Scotland: Morgan Kaufmann.
3. PA. Futreal, L. "Early Theories about Cancer Causes – American Cancer Society".Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, Stratton MR (2004). "A census of human cancer genes". Nat. Rev. Cancer. 4(3): 177–83. doi:10.1038/nrc1299. PMC 2665285. PMID 14993899.
4. Forbes S, Clements J, Dawson E, Bamford S, Webb T, Dogan A, Flanagan A, Teague J, Wooster R, PA. Futreal, MR Stratton (2006). "COSMIC 2005". Br J Cancer. 94 (2): 31822. doi:10.1038/sj.bjc.6602928. PMC 2361125. PMID 16421597.
5. Gavin Brown. Diversity in Neural Network Ensembles. The University of Birmingham. 2004.
6. BabackMoghaddam and Gregory Shakhnarovich. Boosted Dyadic Kernel Discriminants. NIPS. 2002.
7. Krzysztof Grabczewski and Wl/odzisl/aw Duch. Heterogeneous Forests of Decision Trees. ICANN. 2002.
8. AndrásAntos and BalázsKégl and Tamás Linder and Gábor Lugosi. Data-dependent margin-based generalization bounds for classification. Journal of Machine Learning Research, 3. 2002. [View Context].
9. P. Kristin Bennett and AyhanDemiriz and Richard Maclin. Exploiting unlabeled data in ensemble methods. KDD. 2002.
10. P. Kristin Bennett and Erin J. Bredensteiner. A Parametric Optimization Method for Machine Learning. INFORMS Journal on Computing, 9. 1997.