

APPLICATIONS OF MACHINE LEARNING APPROACHES IN LIFE-SCIENCE AND  
HEALTH CARE

**Prasanth R**

Department of Microbiology, Krupanidhi Degree College, Bangalore

**Khalid Imran**

Department of Microbiology, Krupanidhi Degree College, Bangalore

**Priya Chaubey**

Department of Microbiology, Krupanidhi Degree College, Bangalore

**Siddharth Misra**

Accendere Knowledge Management Services, Bangalore

**Abstract: -**

Machine learning is a method in which the machine is trained first with the programming languages C++, python etc., using these languages, the large datasets such as life science and healthcare data are gathered in linear or in cluster form in the form of algorithms. These fields include sub-fields like ecology, biology, microbiology, and also includes medical, drug discovery and clinical diagnosis. The next step after the trained machine is data mining that takes place in two types of machine learning supervised and unsupervised learning techniques which is implemented and this gives the machine to perform and are able to predict the results from the loaded research questions. This makes the machine to gain knowledge itself and perform actions automatically by comparing all sort of datasets. The development of this technology will improve the quality and quantity of the product. And the growth of machine learning will improve the life of people and also saves peoples from many things such as diseases, food production, climatic change and also predict and gives suggestion for the farmers to improve the soil fertility and crop growth.

**Key words:** Machine learning, supervised, unsupervised learning, datasets, algorithms, C++, python, healthcare, life science, prediction

## Introduction

Several concepts of machine learning were suggested over the years some of the earliest read “A” computer program is supposed to benefit from experience “E” about any class of tasks “T” and performance evaluate “P” if its success at tasks “T” as assessed by “P” increases with experience “E”[1]. Another said learning on machine is the ability of the computer program to produce a new structural data that will be entirely different from an old one to gain output for the numerical or nominal data rules from inputs [2]. Machine learning is a method in which the code is coded using various computer languages and utilizing the same. The language programming is a structured language with a series of commands which may generate specific kind of software outputs for incorporating the algorithms in it. Machine learning falls as a subfield under artificial intelligence and has been officially defined and declared by Arthur Samuel as providing computers the capacity to learn without being directly programmed. The algorithms can be implemented and can correctly predict or measure the results [3]. Machine learning has a very structured approach; the interfered associations are from the databases rather than the fast coding [4]. The life sciences, like medicines and biology, are increasing machine learning technology field [5].

Medicine is essentially a practice dealing with facts, in which vast volumes fairly high-quality data are gathered and processed in databases. Medical data are extremely heterogenous and processed in formats of numeric text, sound, image and video. They provide health data (populations, symptoms, diagnostic studies, diagnosis and different images, critical signals, videos etc.), operations data (expenses, and charges regulations, recommendations, clinical trials, etc.), bibliographic details, and molecular data. Bioinformatics involves all the above category of data conceptualizes molecular biology and applies bioinformatics techniques which is derived from computer science, applied mathematics and statistics to determine and coordinate large-scale knowledge on such molecules [6]. Instead of using the old techniques, such as hand-coding a program routine with a particular collection of instructions to execute a particular operation, this is achieved by teaching the computer utilizing vast quantities of data and algorithms to allow the device or computer to know how to conduct the function. There are usually two key methods used to implement machine supervised learning and unsupervised learning. The use of this supervised method of learning is basically used to develop a training model for predicting future values of such data categories or sustainable factors whereas in the case of another technique, unsupervised methods are used for exploratory purposes [7].

**Machine learning methods:**

There are two methods that are widely used in machine learning as Supervised and Unsupervised learning methods.

Supervised learning:

Supervised learning task or learning a function which deals with input-output. Example input-output pairs. This provides training data which consist of series training examples. Example consist of an input (vector) to get desired output value (supervisory signal). In this method when the training data is analyzed by algorithms, it produces an interfered function that is used for mapping new examples. The correct data can be obtained by the algorithm which are not easily available. For this algorithm learning is must to generalize the training data in a “reasonable” manner to unseen situations.

Following are the steps of supervised leaning: -

1. Preparing the data.
2. Training.
3. Test Step or evaluation.
4. Deployment of the production; step that plays crucial role in supervised learning method.

Types of supervised learning: -

1. Classification
2. Regression

Unsupervised learning

Unsupervised learning is a technique of machine learning which do not have supervised model. The computer is allowed to find out the data of its own unlike the supervised learning method which is already being provided. Natural learning techniques are more useful than the unmonitored learning which is non- predictable. Some of these techniques are useful in: -

Determining the unknown patterns of data.

Finding out the characteristics that can be used for categorization.

Analyzing and labelling of input data in the presence of learners.

Labelled data can be easily obtained unlike the unlabeled data that require manual intervention.

The problems of unsupervised learning are further sub-divided into clustering and association problems. In these the most used type of unsupervised learning is clustering. The concept of unsupervised learning is very important. It's mainly about determining the structure in a non-categorized data which has been collected. Algorithms clustering will process the data if they exist and will find natural clusters (groups).

Among data objects, association rules can be set up which will allow the association of large databases. This unsupervised technique involves the discovery in large databases of interesting relationships between variables. [8, 26]

### **In- life science**

Life science machine learning is one of the most advanced technologies to govern future life science departments as well as healthcare. It can be described commonly as artificial intelligence. Artificial intelligence is a presumed field of technology that makes human work very simple and also the most advanced, broadest-minded technology which is able to do all the work itself (automatically) and also respond quickly. The most important thing is that they have the ability to learn themselves and to predict future results as well. This facilitates or even perfects the whole work process. In many fields, such as the automobile industry, also in some life science healthcare in many countries and this has already been implemented. It is also slowly developing in India. The most important part of artificial intelligence is machine learning and deep learning major role that plays major role and performs combined.

### **What is life science?**

Life science is a vast field of study that looks at every living thing on earth. Life sciences aim to learn everything about life on this planet, from bacteria to begonias to beluga whales. "Life sciences" as the name might suggest, study life in all its forms (past and present). This can include plants, animals, bacteria and viruses, single-celled organisms and even cells. Life sciences are studying the biology of how these organisms live and that is why you can hear this group of specialties called biology. It is estimated 8.7 million animal species, about 400,000 plant species, bacteria, virus species and plenty of different life forms that can be studied. Many life science researchers are specialized in one class or organism and there are even more sub-specialties such as zoology. Life science have more than thirty different

branches but we will review some of the major branches here in which the machine learning plays its major role.

Driven by improved processing capacity, energy, memory and breathe taking data volumes in computers are used with impressive accuracy to carry out a wide variety of dynamic activities. Machine learning (ML) as an academic discipline includes mathematics, informatics features and statistics. ML is the engine which helps progressing in artificial intelligence (AI) and developing this process in both academic and industry. It is impressively employed to accelerate the development of 'intelligent goods' with an accurate prediction capability using different data sources [9]. Though various statistical methods, such as linear and logistic regression, can forecast new data. It uses statistical methodology that is motivated by inferences about the relationships between variables it is usually done by algorithms. There are different types of algorithm languages such languages are C++, etc., and the most trending programming language is python which is used now a days widely and are implemented in some important life science fields namely bioinformatics in ecology botany, zoology, microbiology.

### **What is health care?**

Health care has come in the form of any element, program and tool to take care of health nowadays. Healthcare is not anything to be given, purchased or sold at all, but a whole ecosystem with many specific moving parts linked only by virtue of the patient's life. Each patient with unique needs would have an environment that fits their own health needs, and one that evolves over time. While every people in this world have a need of good health and also, they have the right to choose their ecosystem which suits their own needs. Health care is the one which includes physician offices, hospitals, clinics, radiology facilities, physical therapy departments, pharmaceutical firms, and even health insurance providers, community buying agencies, pharmacy advantage administrators, corporate healthcare networks, and medical/PBM/pharmacy combinations and much more [11].

### **Preparing machine through algorithm: -**

Here the collected data sets are arranged in a suitable format so that these datasets can be programmed using the programming language and it will be easy to feed the machine with all the datasets. The techniques used in this preparation step are

\* Using logistic regressions through Generalized Linear Models

(GLMs) with Shrinkage Operator (LASSO) regularization L1 Least Absolute Selection and

\* Support Vector Machines (SVMs) with a radial basics function (RBF) kernel.

\*Artificial Neural Networks (ANNs) with a single hidden layer. [10]

\*Support Vector Machines (SVMs) with a radial basic function (RBF) kernel.

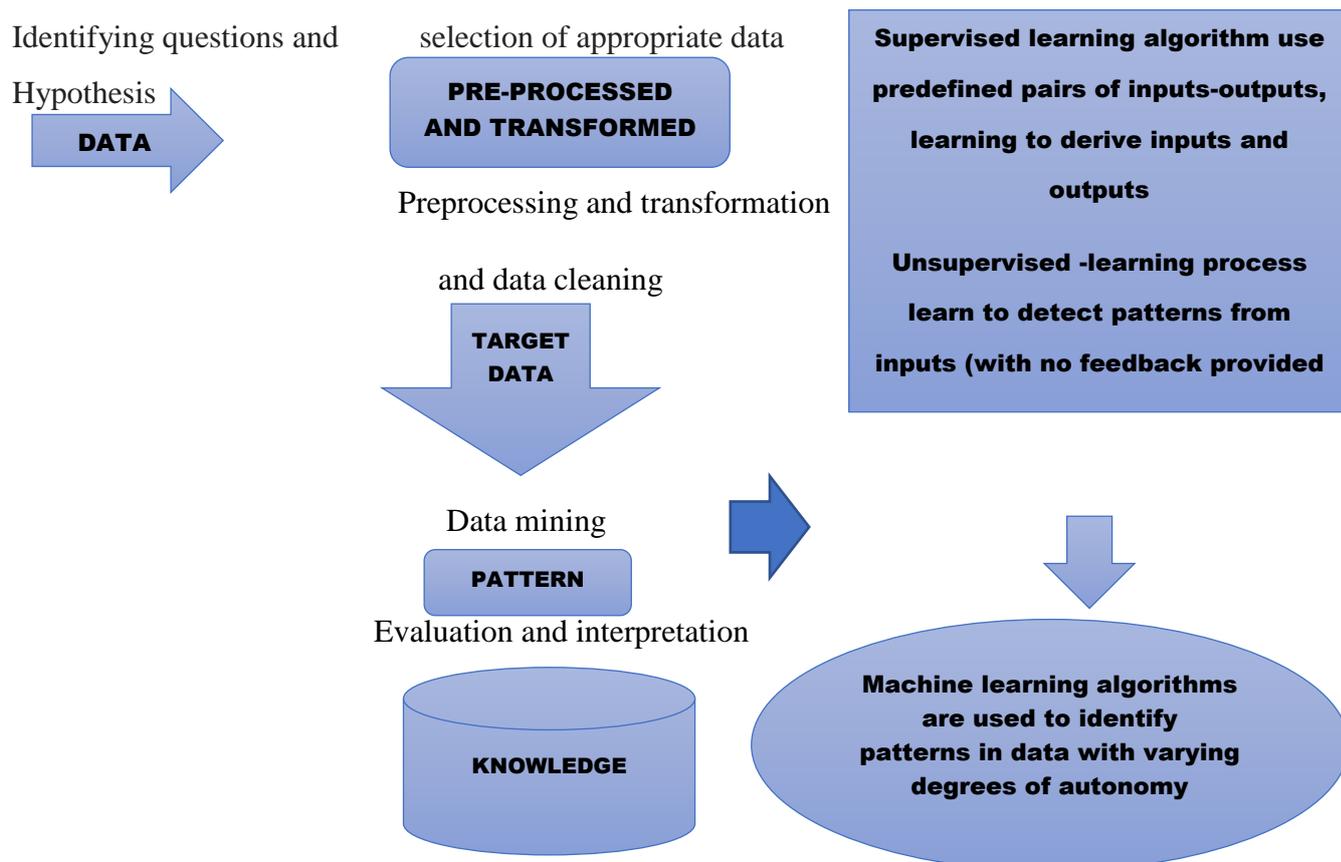
After applying these techniques finally, the data will be uploaded in the machine and then the testing of this trained model drives where the comparison of the predicted data is done. And from this place the machine starts learning itself and developed as artificial intelligence. This is the overview of machine learning technique.

**Machine learning in various fields: -**

### 1) **Machine learning In Ecology**

More number of research-based disciplines are taking a more combining approach to global issues like overall climate change and nutrition stability and human migration [12]. To this end, deep mind works with developing machine learning methods such has environmentalists and conservationists to help study a whole animal community's behavioral dynamic.

The ecosystem is globally unparalleled in its biodiversity, hundreds of motion sensitive cameras have been installed inside the heart of the protected area by Research program. The cameras are triggered by passing through wildlife, frequently capturing animal images across vast spatial scales, allowing researchers to study animal behavior, distribution and populations with high spatial and temporal resolution [13]. This can be implemented by a specialized work process such as follows



Step by step process which has been mentioned above is followed to train the machine with the data mining, where the machine learning languages with large data sets have been uploaded. Hence the machine is able to capture the animals and categorize them and the data are stored and statistical evaluations are done. There are some instruments which is used to predict the climate changes and pollutions and some are also used to predict the fossil fuel available in the earth. So, ML has a major role in predicting and protecting the environment for future.

### Machine learning in Biological fields

Agriculture is a field with very unique conditions and curb on its work. This is due not only to weather dependence but also to the labor market. The workload includes prominent peaks during periods for highly concentrated farming (harvest) operations so that one can predict the short-term weather and the seasonality. Most of the time, agriculture is a routine, routine seeding, pruning, harvesting, weeding, sorting, , picking and feeding, etc. agricultural robots automate these slow, reputable and slow tasks to enable farmers to focus more on strategic issues and improve overall output [14]. The most important and well-defined applications of

machine learning in biological field (botany). Here the large varieties of botanical data are gathered and then they are programmed in to a machine using programming language and the model tests are done. Due to this the machine is capable of predicting the plants fruiting time and it is able to identify the diseases by just scanning the plant and tells which diseases is caused and also it is capable of suggesting the medicine. These all are done automatically by comparing a large variety of datasets by machine learning. In India the technique of (ML) and (AI) is developing in every area. Another most important application of (ML) is to suggest the farmers at which period the crop should be cropped. It also evaluates the soil and the climate and suggest the time periods for using fertilizers. By this the yield is increased and workload is reduced [27].

In the field of machine learning two major paradigms exist, supervised and unsupervised learning. Both have potential biologic applications. They are used in bioinformatics which is able to compare a wide range of genetic information's or genetic sequences of the plants and identify the specific one and gives the total details of that plant family. Widely used in genetic engineering field where machine learning helps to plot the genetic mapping and also to sequence a biological data itself. There are some more important applications of machine learning in botany. This helps the researchers to study further and also to predict what will happen in future accurately.

### **Machine learning in healthcare:**

Machine learning also plays a wide range in health care fields in terms of medicine. Machine learning algorithm has been grown by Google to help classify mammograms with cancerous tumors. Stanford has used deep machine learning algorithm for tracking skin cancer. A recent JAMA article published findings of a deep (ML) algorithm which is capable of diagnosing retinopathy in retinal photographs.

Machine learning can be targeted, efficient, and patient-centered in reducing readmission. Clinicians Amy provides regular advice to the patients so that they can take the necessary step to get readmitted reduce the risk. Machine learning helps to minimize HAIs, such as central-line related bloodstream infections (CLABSIs) and every year 40 percent of CLABSI patients die due to this. Machine learning comes into play where it can predict which patients may develop a CLABSI with central line. Clinicians can monitor and intervene the patients who are at high risk to this disease at. It also helps the healthcare systems to classify the patients with chronic disease that has not been identified or misdiagnosed, in predicting the

probability of developing the chronic disease, and provide specific treatment to the patients. These are some important applications of machine learning in clinical fields [15]. It also has a role in medicinal and drug discovery fields many pharmaceutical companies and startups concentrate on decoding genetic data and on personalized medicine. Knowing the genetic profile of the patient helps to deliver the best medications and care. Helps to deliver the best medication and care. Machine learning may advance the building of computational approaches to analyze genetic data and develop novel therapies. Machine learning methods include a range of resources with plentiful, high quality data that can facilitate exploration and decision taking for well-specified questions. Chances of applying machine learning occur at all phases of drug development, it includes target validation, prognostic biomarker recognition, and clinical trial analyzes of the digital pathology results. The application of machine learning has the ability to decrease the health risk and failures of treatment and discovery of drug [16].

### **Machine learning in microbiology:**

Microbiology is a vast field where it includes many species. It is generally classified into prokaryotes, eukaryotes, protozoa and viruses. Usually microorganisms can be isolated manually by preparing the media and inoculating and then incubating. There are lots of challenges that has to be faced because each and every step should be sterile, and any sort of contamination should be avoided. In some cases, researchers are pushed to work on pathogenic microorganisms and viruses which is very harmful and pathogenic for the handling persons. Chances of getting infected is 100% by the microorganisms. The only way to stop or to solve this problem is implementing the machine learning technique in microbiology. This will not only reduce the contamination but will also reduce the workload and researchers can also predict the results using the machine learning. The work will be done automatically and will rapidly increase the productivity and source of failure will be down. Some of the applications of machine learning in microbiology are: -. The SVM is a generalized linear classifier based on the maximum-hyperplane learning set, which can perform binary classification of data using a decision basis. The SVM can sort non-linear data using the kernel methods [17] and used in bioinformatics to predict the protein sequence, DNA sequence and their structure.

Machine learning enables vast and complex data sets to be analyzed and has the potential to improve healthcare. The clinical laboratory for microbiology is of particular importance for

the development of machine learning systems at the clinical practice and diagnostic interface. Some machine learning programs targets to reduce the time between collection of sample and evaluation of sample in microbiology. Recently, researchers reported using Raman optical spectroscopy in which they used the colony of bacteria in machine learning to recognize a specific kind of microorganism among all the 30 species of bacteria & yeast [18]. Machine learning is also used to identify HPV genotypes in virology [19]. Advances in metagenomic sequencing technologies that produce several WGS time-points in combination with machine learning technologies providing a better knowledge of the virus and viral relationship with their host and/or environment. Machine learning applications used microbiological data to identify, classify, and forecast clinical effects of infectious or non-infectious diseases. Microbiota (obtained by selective gene sequencing or metagenomics shotgun sequencing) which collaborated with machine learning for the prediction of periodontitis dental caries or bacterial vaginosis. [20, 21 and 22]. Machine learning will cause the development of creative diagnostic point-of-care devices that have the ability to be utilized in remote areas. Co-evolutionary neural networks have been used to treat malaria or stool samples to identify *Helminths* on mobile images of dark blood stain [23].

### **Conclusion:**

Machine learning has a high approach in life science and healthcare. The paper gives a basic detailed idea of the autonomous technology that may dominate the world in the future. Developing machine learning in healthcare and life science can minimize the workload as well as decrease the loss and maximize the productivity in less amount of time.

If it is implemented in India the technology will save many poor people's life by decreasing the infections, diseases etc., and hence each and every people in India will be benefitted. Role in microbiology has a great impact. Many life processes include microorganisms which impact their natural ecosystem and other microorganisms. Machine learning in microorganisms plays significant roles in agriculture, livestock, human health, livestock farming, environmental protection, food production and industrial chemicals manufacturing.

### **Acknowledgement**

The satiation and euphoria that accompany the successful completion of this research would be incomplete without the mention of the people who made it possible. We thank the research team of Accendere Knowledge Management Services, CL Educate Ltd. for their unflinching

guidance, continuous encouragement and support to successfully complete this research work.

## Reference

- (1) Mitchell TM. Does machine learning really work? *AI magazine*. 1997 Sep 15; 18(3):11-.
- (2) Cios KJ, Pedrycz W, Swiniarsk RM. Data mining methods for Knowledge discovery: Neural networks. *IEEE transactions*. 1998 Nov; 9(6):1533-4.
- (3) Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media; 2009 Aug 26.
- (4) Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*. 2016 Sep 29; 375(13):1216.
- (5) Aliferis CF, Statnikov A, Tsamardinos I. Challenges in the analysis of mass-throughput data: a technical commentary from the statistical machine learning perspective. *Cancer Informatics*. 2006 Jan; 2:117693510600200004.
- (6) Lefèvre C, Jobard F, Caux F, Bouadjar B, Karaduman A, Heilig R, Lakhdar H, Wollenberg A, Verret JL, Weissenbach J, Özgüc M. Mutations in CGI-58, the gene encoding a new protein of the esterase/lipase/thioesterase subfamily, in Chanarin-Dorfman syndrome. *The American Journal of Human Genetics*. 2001 Nov 1; 69(5):1002-12.
- (7) Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*. 2019 Jun; 18(6):463-77.
- (8) Shalev-Shwartz S, Ben-David S. *Understanding machine learning: From theory to algorithms*. Cambridge university press; 2014 May 19.
- (9) Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Sci (NY)*. 2015; 349(6245):255–60. <https://doi.org/10.1126/science.aaa8415>.
- (10) Sidey-Gibbons and Sidey-Gibbons *BMC Medical Research Methodology* (2019) 19:64 <https://doi.org/10.1186/s12874-019-0681-4>
- (11) Brown EC. Irrational hospital pricing. *Hous. J. Health L. & Pol'y*. 2014; 14:11.

(12) Baziliana, M., Rogner, H., Howells, M., Hermann, S., Arent, D., Gielen, D., Steduto, P., Mueller, A., Komor, P., Tol, R.S.J., Yumkella, K.K., 2011. Considering the energy, water and food nexus: Towards an integrated modelling approach. *Energy Policy* 39, 7896–7906. <https://doi.org/10.1016/J.ENPOL.2011.09.039>.

(13) Willcock S, Martínez-López J, Hooftman DA, Bagstad KJ, Balbi S, Marzo A, Prato C, Sciandrello S, Signorello G, Voigt B, Villa F. Machine learning for ecosystem services. *Ecosystem services*. 2018 Oct 1; 33:165-74.

(14) Edan, Y., Han, S., and Kondo, N. (2009). Automation in agriculture,” in Springer Handbook of Automation, ed. Y Shimon (Berlin: Springer), 1095–1128. doi: 10.1007/978-3-540-78831-7\_63

(15). Corbett E. The real-world benefits of machine learning in healthcare. *Health Catalyst...* 2017.

(16). Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*. 2019 Jun; 18(6):463-77.

(17) Drucker, H., Wu, D., and Vapnik, V.N. (2002). Support vector machines for spam categorization. *IEEE Transactions on Neural Networks* 10(5), 1048-1054

(18). Ho C-S, Jean N, Hogan CA, Blackmon L, Jeffrey SS, Holodniy M, et al. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat Commun*. 2019 Dec; 10(1):4927

(19) Tanchotsrinon W, Lursinsap C, Poovorawan Y. A high performance prediction of HPV genotypes by Chaos game representation and singular value decomposition. *BMC Bioinformatics*. 2015 Dec; 16(1):71.

(20) Chen W-P, Chang S-H, Tang C-Y, Liou M-L, Tsai S-JJ, Lin Y-L. Composition Analysis and Feature Selection of the Oral Microbiota Associated with Periodontal Disease. *Bio Med Research International*. 2018 Nov 15; 2018:1–14.

(21) Wang Y, Li R, Zhou Y, Ling Z, Guo X, Xie L, et al. Motif-Based Text Mining of Microbial Metagenome Redundancy Profiling Data for Disease Classification. *Biomed Res Int*. 2016; 2016:6598307.

- (22) Beck D, Foster JA. Machine Learning Techniques Accurately Classify Microbial Communities by Bacterial Vaginosis Characteristics. White BA, editor. PLoS ONE. 2014 Feb 3; 9(2):e87830.
- (23) Quinn JA, Nakasi R, Mugagga PKB, Byanyima P, Lubega W, Andama A. Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics. arXiv:160802989 [cs] [Internet]. 2016 Aug 9 [cited 2019 Oct 18]; Available from: <http://arxiv.org/abs/1608.02989>
- (24) Ascher J, Sartori G, Graefe U, Thornton B, Ceccherini MT, Pietramellara G, Egli M. Are humus forms, mesofauna and microflora in subalpine forest soils sensitive to thermal conditions? *Biology and Fertility of Soils*. 2012 Aug 1; 48(6):709-25.
- (25) DiMucci D, Kon M, Segre D. Machine learning reveals missing edges and putative interaction mechanisms in microbial ecosystem networks. *M systems*. 2018 Oct 30;3(5):e00181-18.
- (26) Singh Y, Bhatia PK, Sangwan O. A review of studies on machine learning techniques. *International Journal of Computer Science and Security*. 2007 Jun; 1(1):70-84.
- (27). Royal Society of London. Series B, Biological sciences 360(1462), 1935-1943. doi: 10.1098/rstb.2005.1725