International Journal of E-Health and Medical Communications
Volume 12 • Issue 1 • January-March 2021

Rummage of Machine Learning Algorithms in Cancer Diagnosis

Prashant Johri, SCSE, Galgotia's University, Greater Noida, India
https://orcid.org/0000-0001-8771-5700

Vivek sen Saxena, INMANTEC Institutions, Ghaziabad, India
https://orcid.org/0000-0001-6107-5506

Avneesh Kumar, SCSE, Galgotia's University, Greater Noida, India
https://orcid.org/0000-0001-5860-3689

ABSTRACT

With the continuous improvement of digital imaging technology and rapid increase in the use of digital medical records in last decade, artificial intelligence has provided various techniques to analyze these data. Machine learning, a subset of artificial intelligence techniques, provides the ability to learn from past and present and to predict the future on the basis of data. Various AI-enabled support systems are designed by using machine learning algorithms in order to optimize and computerize the process of clinical decision making and to bring about a massive archetype change in the healthcare sector such as timely identification, revealing and treatment of disease, as well as outcome prediction. Machine learning algorithms are implemented in the healthcare sector and helped in diagnosis of critical illness such as cancer, neurology, cardiac, and kidney disease as well as with easing in anticipation of disease progression. By applying and executing machine learning algorithms over healthcare data, one can evaluate, analyze, and generate the results that can be used not only to advance the prior health studies but also to aid in forecasting a patient's chances of developing of various diseases. The aim in this article is to present an overview of machine learning and to cover various algorithms of machine learning and their present implementation in the healthcare sector.

KEYWORDS

AI-Enabled Support System, Artificial Intelligence, Clinical Decision Making, Computer-Aided Diagnosis, Healthcare, Machine Learning Algorithm

1. INTRODUCTION

Machine learning, a subfield of Artificial Intelligence, provides algorithms to learn from past experiments while performing a particular task and measuring performance. By working continuously on a task, the performance of the task gets improved and the user experience as well. A Machine learning system has a training data set working as knowledge base and rules for decision making (Blum 2007). Machine learning is the building and exploring of methods in a computer programming language and making them "learn". The program developed using machine learning algorithms accesses the data, trains the machine and tests it again for performance evaluation. The most important characteristic of machine learning is its ability to forecast. In machine learning, a model for prediction

DOI: 10.4018/IJEHMC.2021010101

is build by existing information and it is further used for predicting the data. The major aspect of learning is the features selection from the data set as all the features cannot be used in learning. The data set may have multiple fields and perspectives. The selection of features is done according to their relevance, implication and scenario (Du & Swamy 2013). The primary goal of machine learning is to produce and enhance the learning algorithms and models in order to facilitate their easy application in various disciplines such as agriculture (Patrício & Rieder 2018), banking (Erdogan 2013), cyber security (Buczak & Guven 2016), economics (Einav & Levin 2014), finance (Lin et al., 2012), insurance (Gan 2013), natural language processing (Collobert & Weston 2008), online & traditional marketing (Tripathy et al., 2006), healthcare (Crown 2015), network & telecommunication (Richter & Khoshgoftaar 2018) and others as well.

2. MACHINE LEARNING PARADIGMS

Machine learning algorithms are categorized according to their design, required input, produced output and applications. The majorly of machine learning algorithms are categorized as supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, evolutionary learning and deep learning as shown in Figure 1 (Fatima & Pasha 2017). Supervised Learning algorithms construct a model that uses a set of labeled data and desired output. The machine has been trained with the sampled data having essential features and preferred output. Once the machine is trained, testing is performed against the test data and the results are matched with desired results and the accuracy of the machine and algorithm gets measured. With the availability of huge amount of unlabeled data, supervised learning is not possible. Hence it is necessary to use unlabeled data to train a machine. The learning process using unlabeled data is called unsupervised learning. Unsupervised learning accomplishes the training process of an algorithm, with unlabeled data that are grouped on basis of relationship, variations and patterns within the data. Based on similarities and differences a hieratical structure is formed up to the point that similar objects are grouped together (Goyal & Kishnan 2019). Unsupervised learning algorithms are also termed as clustering algorithms as a cluster of data is formed from large amount pf unlabeled data while considering resemblance and variation (Mohssen & Eihab 2017). Clustering algorithms are further categorized as Density based methods; Connectivity based methods, Hierarchy Based Methods, Centroid Based Method or Partitioning Methods and Distribution Based Methods. Semi supervised learning uses labeled data (in small amount) and unlabeled data (in large amount) (Zhu 2005). Reinforcement learning is quite different from supervised learning and unsupervised learning. There is neither a knowledgeable expert supervisor nor an input/output pair. Reinforcement learning is a goal oriented technique of machine learning in dynamic environments. Reinforcement learning is machine learning process to produce intelligent programs or agents through learning and adopting from environmental changes. Learning can be done even though the information about the environment is not completely known. Agents get the feedback about the action performed and reward/punishment immediately. The methods for reward/punishment signal are Finite- horizon model, receding-horizon control, infinite-horizon discounted model, and average-reward model (Kaelbling et al., 1996). Evolutionary learning is a learning process where an algorithm learns from its past result and improves its performance (Zhang et al., 2011). Deep learning is the improvement over Multi layer perceptron model where the hidden layers can be increased, to a given computational level (Litjens et al., 2017). In deep learning the minimum number of hidden layers must be more than two. Categorization of machine learning algorithms is given in Figure 1.

3. STUDY OF MACHINE LEARNING ALGORITHMS IN HEALTHCARE SECTOR

Healthcare sector is one of the most emerging sectors, which present lifesaving to millions of people; it is also becoming one of the top revenue-earning sectors in various countries. Today in India, approximately 4.7% of total GDP is spent on the healthcare sector per year (Global Health Observatory