
# Applications of Machine Learning in Medicine

Done by: Ghazal Almidanah - 2020001

Course name: Machine Learning - 402331

Date: 14 Jan, 2024

❖ Table of Contents:

# Applications of Machine Learning in Medicine

Applying new technologies in the medicine world could ease the way into fast development for the medicine field leading to a breakthrough in the machine learning field as well.

## ❖ Abstract:

Applications of machine learning has been used in medicine for many different reasons it had high accuracy and more efficiency for the field. The medicine field had always been affected by the technology growth for the reason that it's intertwined in almost every part of the medicine world and such advanced technologies namely Machine Learning and Artificial Intelligence can produce applications that identify disease and diagnoses similar to human way of functioning this would've never been possible before the ML and AI.

## ❖ Introduction:

To discuss the importance of Machine Learning in the medicine field scientists have studied each field in details on it's on for instance machine learning is a subfield of computer science and artificial intelligence (AI) it focuses on using data and algorithms to simulate human learning processes and progressively increase their accuracy. Advances in storage and processing power over the past few decades have made machine learning-based products possible, like Netflix's recommendation engine and self-driving cars.

The algorithms in machine learning are trained to make predictions or classifications and to find valuable information in data mining projects using statistical techniques. Subsequently, these insights inform business and application decisions, which ideally influence important growth metrics. Data scientists will be in greater demand as big data continues to grow and expand. They'll have to assist in determining which business questions are most pertinent and what information is needed to address them.

Typically, frameworks like TensorFlow and PyTorch, which speed up solution development, are used to create machine learning algorithms.

As for the medicine field it's known for a general sense but in more details the study of health and healing is known as medicine. It consists of physicians, nurses, and other specialists. It covers a wide range of topics related to health, including medical research, disease diagnosis, treatment, and prevention the sole aim for it is to promote and maintain health and wellbeing.

Allopathic medicine is another term for conventional modern medicine. It entails the use of medications or surgery, frequently accompanied by lifestyle modifications and counseling. Acupuncture, homeopathy, herbal medicine, art therapy, traditional Chinese medicine, and many more practices are examples of complementary and alternative medicine.

Using such technologies in medicine affected the development of the machine learning and artificial intelligence not only enhance the power of medicine alone because we came to new applications for example Identifying Diseases and Diagnosis, Drug Discovery and Manufacturing, Medical Imaging Diagnosis and many more.

❖ Findings:

**Machine Learning Applications**:

a. Identifying Diseases and Diagnosis

By comparing a patient's symptoms, genetic history, demographics, and other data with the data of other patients in the system, medical professionals used machine learning to ascertain whether a patient has a disease or another, Numerous diseases, such as heart disease, diabetes, kidney disease, breast cancer, Parkinson's disease, Alzheimer's disease, and more, can be detected by machine learning, according to research.

ML Algorithm:

Algorithms depending on the requirements for explain ability versus accuracy and the complexity of the data, a variety of machine learning algorithms can be used to detect diseases like:

- Decision Tree: The attributes in Decision Tree-DT models can have multiple values, or what are called classification trees; leaves represent different classes, while branches represent the combination of characteristics that lead to those class labels. Regression trees, on the other hand, are continuous variables that DT can accept.
- KKN: Preprocessing is done on the provided data to extract all of the metadata. Using all of the training data available, KNN is used to determine who the given data's closest neighbors are. The algorithm terminates if a label is detected; if not, the system classifier is used. The object was identified using the suggested algorithm. The outcomes are contrasted with those acquired using KNN and a single system classifier.

b. Medical Imaging Diagnosis

The innovative field of computer vision is the result of both machine learning and deep learning. This has been accepted by Microsoft's Inner Eye initiative, which develops image diagnostic tools for image analysis. Anticipate the inclusion of more data sources from various medical imagery types in this AI-driven diagnostic process as machine learning grows in accessibility and explanatory power.

ML Algorithm:

Algorithms can analyze medical images to identify specific features that may indicate a particular condition or disease, such as cancer. This can help researchers develop more accurate diagnostic tools and improve treatment outcomes like:

- CNN: is a deep learning algorithm that mimics the visual system of animals by drawing inspiration from the visual cortex of the animal brain [30]. With regards to image classification, segmentation, localization, detection, and other related tasks, CNNs represent a quantum leap in the field of image understanding.

c. Drug Discovery and Manufacturing

Early-stage drug discovery is one of the main clinical applications of machine learning. This also includes research and development technologies that can aid in the discovery of alternate avenues for the treatment of multifactorial diseases, such as precision medicine and next-generation sequencing.

ML Algorithms:

AI systems can analyze diverse data types, such as genetic, proteomic, and clinical data, to identify potential therapeutic targets. By uncovering disease-associated targets and molecular pathways using functions like:

- SVM: modeling, the hyperplane represents a classification boundary. The *margin* of the hyperplane is the distance between two object classes in feature space separated by the hyperplane for SVM classification. *Support vectors (SVs)* represent data samples of one class that are closest to the other class and thus used to define the margin of the hyperplane. *Kernel function* is a similarity function that takes as input vectors in original feature space and calculates a modified inner product in a higher-dimensional space. The *kernel trick* refers to a strategy for generating a non-linear SVM.

d. Clinical Trial and Research

They are two areas where machine learning may find several uses. Clinical trials are expensive in terms of both time and money, and they often take years to finish, as anyone in the pharmaceutical business will attest. Researchers can select participants for clinical trials by using machine learning (ML)-based predictive analytics to combine data from social media, medical records, and other sources. Additionally, machine learning has been applied to determine the ideal test sample size, guarantee real-time monitoring and data access for trial participants, and harness the power of electronic records to minimize data-based errors.

ML Algorithms:

The conditions for artificial intelligence (AI) to succeed have been created by increased access to vast amounts of electronic data (publicly available datasets in medicine, such as the Cancer Genome Atlas and the UK Biobank), Additionally, ongoing advancements in algorithms have made it possible for machines to perform tasks with ever-increasing complexity and nuance using algorithms like:

- Logistic regression: Independent identically distributed (iid) observations are necessary for logistic regression. When case-control status is the binary outcome, this assumption is broken in certain studies, where cases are matched to controls. In this case, the data might be fitted to a conditional logistic regression model that takes the matched study design into consideration.

- Random Forest: An ensemble of decision trees trained concurrently is called a random forest. Using bootstrapping and aggregation, the training process for individual trees iterates over all the features to choose the best features that divide the spaces. Using distinct feature subsets, the decision trees are trained on different subsets of the training datasets. The outcome of a random forest (RF) is determined by combining the output of each decision tree and choosing the prediction with the highest number of votes.

e. Radiotherapy

Radiology is one of the fields in which machine learning is most sought-after in the healthcare industry. Numerous discrete variables that can appear at any time can be found in medical image analysis. Numerous lesions, cancer foci, and other conditions are too numerous to be adequately represented by intricate equations. It gets simpler to diagnose and identify the variables because ML-based algorithms learn from the vast array of available samples. The most widely used application of machine learning in medical image analysis is the classification of objects, like lesions, into normal and abnormal, lesion and non-lesion categories.

ML Algorithms:

Building systems that automatically get better through experience—and encoding that experience into statistical models derived from historical examples of input-output data—is the aim of machine learning. By modeling the underlying statistical patterns that appear in the input-output observations, these systems aim to automate the generation of outputs given future inputs with the goal of creating decision rules requiring little to no human intervention using algorithms like:

- Bayesian Networks: (BN) are models that describe dependencies relationships between variables in a dataset using a Directed Acyclic Graph and Conditional Probability Distributions. They are very effective tools because they can deal with uncertainty in the modeling and take into account prior knowledge. They are used in many multidisciplinary application fields, such as healthcare or the social sciences, because their graphical component makes them more intuitive to non-Machine Learning users.

- Ensemble methods: Higher accuracy was obtained using ensemble methods, which are known for their ability to handle large and incomplete datasets. To better comprehend the significance of the data, ensemble models should be investigated in more detail. By enabling a more flexible structure in the models, ensemble methods outperformed other single algorithms in terms of predictive performance. It might be more efficient to extract knowledge by straying from the conventional CART models and other single models, particularly when dealing with sizable and frequently incomplete datasets. The outcome of a random forest is determined by combining the output of each decision tree and choosing the prediction with the highest number of votes.

❖ Analysis:

**Challenges**:

There is many benefits to using machine learning algorithms as stated briefly before, many methods have high accuracy level and can handle big data as well as incomplete data in a remarkable way regardless machine learning methods still face many hard challenges that is very hard to pull through from them.

For in instance lack of quality data is one of the most challenging aspects as in the medical field images is quite low in number to be significantly used to test, it usually takes a very long period of time to label large volumes of data to use and also clinicians run tests that they need for the patients to get the required results so they can analyze the tests to answer the concerns about patients and asking medical staff to run more tests just for data collection purposes is not a request that we hear of so having a complete and valuable dataset that can have greater effects is a very difficult task.

Baseline comparisons also are absent from a lot of machine learning studies. It is challenging to determine whether the more complex model is preferable to a simpler model due to this lack of transparency. A small gain in performance over a basic model is often not worth the interpretability loss in healthcare applications.

Models are regarded as intellectual property by medical facilities and related software vendors. Reproducibility is problematic in these views, and not just in the medical field. It is now required by some research publications to discuss all necessary procedures down to the level of pseudo-code and to cite any open-source resources that were used. Nevertheless, these opaque problems lead to code duplication amongst organizations and complicate the process of guaranteeing that software defects do not contaminate results.

❖ Conclusion:

We have discussed the potential applications of data science and machine learning for predictive analytics of medical data. We've also seen that the likelihood of the models gaining a clear understanding of the underlying data and producing accurate predictions increases with the amount of data available even though right now we have big data like never before. Nevertheless, there are certain difficulties in applying machine learning to medical data that make it more challenging for scientists but understanding the issues of the data and models can help further achieve the aim of using machine learning in medicine which is greater effect with lower cost, as we have seen in the field progressively.

## ❖ References:


{1} Top 10 Applications of Machine Learning in Pharma and Medicine (2018). Available At: Top 10 Applications of Machine Learning in Healthcare - FWS (flatworldsolutions.com)

{2} How to Build a Disease Detection System Using ML (2020). Available At: How to Build a Disease Detection System Using ML (akkio.com)

{3} Machine-Learning-Based Disease Diagnosis: A Comprehensive Review (2022). Available At: Machine-Learning-Based Disease Diagnosis: A Comprehensive Review - PMC (nih.gov)

{4} Random Decision Forest (2022). Available At: Random Decision Forest - an overview | Science Direct Topics