International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Challenges of Machine Learning in Healthcare Industry

S. Vasundhara

Department of Humanities and Mathematics, G. Narayanamma Institute of Technology & Science Shaikpet Hyderabad

Abstract: ***Machine Learning Applications are everywhere these are used in many real world scenarios. Especially very important in specific areas such as health care and medical data protection. Different types of Machine Learning algorithms are applied to analyse medical records and disease forecasts. In this study a specific review of several Machine Learning Algorithms Applications and opportunities and challenges for health care sector. This paper fills a research gap for efficient use of Machine Learning Algorithms and applications in the health care industry.***

Keywords: Machine Learning, Data Analytics, HealthCare, Machine Learning Applications. Disease Prediction

1. Introduction

Machine Learning is being used in many industries such as automobile, manufacturing, and retail industries. With the development of machine learning and deep learning algorithms, there are a huge number of useful predictions such as predicting the stock prices**,** house prices and loan default prediction. Furthermore, there is data available in different formats that could be used for machine learning predictions. As the data keeps growing, there is a lot of scope for development in the field of machine learning, and predictions are going to get better and better in the future.

One of the interesting applications of machine learning is in the field of healthcare. We've seen a few movies on the internet where they show robots performing the work of a doctor and making the right kind of diagnosis respectively. Movies such as Ender's Game show how robots are used in medical diagnosis. There are many new applications for machine learning created in the field of medical diagnosis. As a result, there is a lot of scope and improvement in the field.

There are more and more sophisticated algorithms being developed with the aid of machine learning and data science. Some of the cool applications of machine learning in healthcare are to predict the chances of occurrence of cancer and predicting Alzheimer's disease. Taking a look at these applications, we can come to a conclusion that machine learning is still growing and there would be an increase in the demand for it in the future as well. Now, there are more and more sophisticated algorithms being developed that are being used in machine learning to make robust predictions.

There are machine learning models implemented in Radiology where the machines are making predictions which would ensure that we are getting the best results on the test set as well.

Challenges with using Machine Learning in Healthcare
Despite the possibilities of machine learning being used in many industries and especially in healthcare, there are still some challenges that data scientists and machine learning engineers go through when they are trying to use these algorithms for medical diagnosis. Getting to know these challenges is beneficial so that one could gather the resources and tools to tackle them and generate better outcomes with artificial intelligence.

There could be challenges when trying to implement machine learning models for healthcare data. One of the challenges when dealing with healthcare data is that the data might be causal for machine learning models. What is meant by causality is that when there is data and if one feature causes the occurrence of the other, the relationship could be said to be having a high causality. In machine learning for most of the algorithms, we assume that the features are independent of each other without one feature causing the other to occur and vice - versa. Therefore, this assumption would be weakened when there is high causality between features respectively.

Shortage of Data Scientists and Machine Learning Engineers
So far, we've talked about data and machine learning algorithms that are the limitations of using data science in healthcare. However, if there is a shortage of people who would be using these algorithms, then the implementation of artificial intelligence in the medical industry would be a major challenge as well. Since the number of machine learning and data science courses in institutions has increased, there is a higher chance of talented professionals entering the field and making it a success. Demand for machine learning engineers is high and only expected to grow as the complexity of and access to machine increases.

Bias would be present in machine learning:
When performing the machine learning tasks, there would be a presence of bias which could lead to the machine learning models not performing well on the unseen or the test set respectively. The bias that is present in machine learning models would be due to the type of data that is fed to machine learning models. Consider, for example, if the data that is given to the models contain a lot of information about a particular class and less information about the minority class without taking into account different scenarios, the machine learning models would make predictions on the test set that is reflective of the results on the training set that contains bias. As a result, the models achieve very high accuracy of the dominant class (majority

class) while performing poorly on the minority class.

Lack of Quality Data

Machine learning and artificial intelligence are some of the most rapidly developing industries in the entire world right now. The global machine learning market is projected to have mind - boggling growth of many times its size over the next few years. Technologies and tools are evolving quickly and becoming more popular by the day. Machine learning finds its use in analytics, computer vision systems, recommendation engines, and many other fields. Because of this, there is a constantly growing need for models to improve: to learn to do things faster, better, and more efficiently. Machine learning models use algorithms to recognize patterns in data and learn from it. There are many metrics to their performance, as there are many tools and resources to improve and optimize it. Unfortunately, no tools and expertise can improve the data model if the input data is incorrect.

The data used in machine learning is the most significant limiting factor for what an ML model can do or how much it can grow. No matter how much you may try to improve the model or how much work and time you and your team put into the process, it will not improve past the level of quality of its data. Data is a hard limit when it comes to machine learning.

Though there is a huge possibility of machine learning algorithms to be put to use, there is still a lot of data that is required in the field of medicine to make the most use of it. The data that is present in the form of medical images is quite low in number to be significantly used to test. Furthermore, data that is present is not labelled so that it could be used for machine learning purposes. It really takes a long time to label large volumes of data for machine learning.

Data annotation must be done accurately

Data is present everywhere in the form of medical images and other useful information. Though there is this huge presence of data, annotated examples or the output label for predictions are not present. Since some of the best machine learning algorithms would be working well under supervision when there is an output class label, we need to be providing the data that is annotated. This would ensure that the annotated examples would help the machine learning models to be fitting and ensuring that there are predictions that are accurate respectively. In medical data, there is a requirement to annotate the data which is a time - consuming process. Therefore, this is one of the challenges for using machine learning in medicine respectively.

Need for Hyper Parameter Tuning

There are a lot of complex ML models being developed with some of them being random forests, decision trees and neural networks (deep learning). Some of the setbacks with these algorithms is to be able to tune (change) the hyper parameters so that they result in a very good performance on the test data (unseen data). In order for them to work well, these hyper parameters need to be changed and monitored constantly so that there is a boost in their performance. However, this could be a tedious task especially when there are so many parameters to tune and monitor in order to get the most optimum results. Therefore, setting the right hyper parameters and altering them to get the most optimum results can also be one of the challenges in applying machine learning to healthcare.

Hyper parameters contain the data that govern the training process itself. Training application handles three categories of data as it trains your model:

- I*nput data (*also called training data) is a collection of individual records (instances) containing the features important to your machine learning problem. This data is used during training to configure your model to accurately make predictions about new instances of similar data. However, the values in your input data never directly become part of your model.
- Model's *parameters* are the variables that your chosen machine learning technique uses to adjust to your data. For example, a deep neural network (DNN) is composed of processing nodes (neurons), each with an operation performed on data as it travels through the network. When your DNN is trained, each node has a weight value that tells your model how much impact it has on the final prediction. Those weights are an example of your model's parameters. In many ways, your model's parameters *are* the model—they are what distinguishes your particular model from other models of the same type working on similar data.
- *Hyper parameters* are the variables that govern the training process itself. For example, part of setting up a deep neural network is deciding how many hidden layers of nodes to use between the input layer and the output layer, and how many nodes each layer should use. These variables are not directly related to the training data. They are configuration variables. Note that parameters change during a training job, while hyper parameters are usually constant during a job.

Model *parameters* are optimized (you could say "tuned") by the training process: you run data through the operations of the model, compare the resulting prediction with the actual value for each data instance, evaluate the accuracy, and adjust until you find the best values. *Hyper parameters* are tuned by running your whole training job, looking at the aggregate accuracy, and adjusting. In both cases you are modifying the composition of your model in an effort to find the best combination to handle your problem.

Without an automated technology like AI Platform Training hyper parameter tuning, you need to make manual adjustments to the hyper parameters over the course of many training runs to arrive at the optimal values. hyper parameter tuning makes the process of determining the best hyper parameter settings easier and less tedious.

How hyper parameter tuning works

Hyper parameter tuning works by running multiple *trials* in a single training job. Each trial is a complete execution of your training application with values for your chosen hyper

parameters, set within limits you specify. The AI Platform Training training service keeps track of the results of each trial and makes adjustments for subsequent trials. When the job is finished, you can get a summary of all the trials along with the most effective configuration of values according to the criteria you specify.

Hyper parameter tuning requires explicit communication between the AI Platform Training service and your training application. Your training application defines all the information that your model needs. You must define the hyper parameters (variables) that you want to adjust, and a target value for each hyper parameter.

To learn how AI Platform Training uses Bayesian optimization for hyperparameter tuning. In addition to Bayesian optimization, AI Platform Training optimizes across hyperparameter tuning jobs. If you are doing hyper parameter tuning against similar models, changing only the objective function or adding a new input column, AI Platform Training is able to improve over time and make the hyper parameter tuning more efficient.

2. Conclusion

All in all, how machine learning and data science could be used for making predictions in medical data. And also seen that the more data we have, the more is the possibility for the models to get a very good picture about the underlying data and making the right predictions. However, we've seen some of the challenges in the field of machine learning when it comes to applying them to medical data. Hope it was able to give a good idea about the usage of machine learning models in healthcare.

References

[1] Bishop, Pattern Recognition and Machine Learning, 2006. Sutton and Barto, Reinforcement Learning: An Introduction, MIT Press, 1998
[2] S. Vasundhara Machine Learning Applications in Health Care systems Turkish Online Journal of Qualitative Inquiry (TOJQI)
[3] R. Atallah and A. Al - Mousa, "Heart disease detection using machine learning majority voting ensemble method, " in *Proceedings of the 2nd International Conference on new Trends in Computing Sciences*, Amman, Jordan, October 2019. View at: Google Scholar
[4] P. S. Kohli and S. Arora, "Application of machine learning in disease prediction, " in *Proceedings of the 4th International Conference on Computing Communication and Automation*, Greater Noida, India, December 2018. View at: Google Scholar
[5] A. Bharat, N. Pooja, and R. A. Reddy, "Using machine learning algorithms for breast cancer risk prediction and diagnosis, " in *Proceedings of the 3rd International Conference on Circuits, Control, Communication and Computing*, Bangalore, India, July 2018. View at: Google Scholar
[6] R. C. Deo, "Machine learning in medicine, " *Circulation*, vol.132, no.20, pp.1920–1930, 2015. View at: Publisher Site | Google Scholar
[7] S. Sánchez Martínez, O. Camara, G. Piella et al., "Machine learning for clinical decision - making: challenges and opportunities, " 2019. View at: Google Scholar
[8] S. Debnath, D. P. Barnaby, K. Coppa et al., "Machine learning to assist clinical decision - making during the COVID - 19 pandemic, " *BioMed Central*, vol.6, pp.2332–8886, 2020. View at: Publisher Site | Google Scholar.
[9] A. Calamuneri, L. Donato, C. Scimone, A. Costa, R. D'Angelo, and A. Sidoti, "On machine learning in biomedicine, " *Life Safety and Security*, vol.5, no.12, pp.96–99, 2017. View at: Google Scholar
[10] S. Vasundhara Impact of Polycystic Ovary Syndrome - Approach with Machine Learning Algorithms. ElseveirSSRN.