Machine Learning in Health Sector

Sutapa Ghosh, Sumit Kumar, Kalinga Institute of Industrial Technology, Bhubaneswar, Odisha, India

Abstract:

Machine Learning is a modern and highly sophisticated technological application, and it has become a huge trend in the industry for last couple of decades. Machine Learning gained wide popularity and is present everywhere. It is playing a significant role in many fields such as finance, online shopping, cloud computing, defense, medical science and security etc. The key strength of Machine learning lies in processing, analyzing and finding patterns from a vast data, and in health sector, such medical data sources are abundant and the goal of using Machine Learning is to provide excellent capabilities to predict diseases from the learnt patterns. In this report, we understand the need of Machine Learning in Health sector, how it is applied for developing efficient decision support for healthcare applications. We are also exploring the significance of Interpretable Machine Learning in Health sector, which becomes important when just choosing an algorithm based on output matrix is not enough and an explanation is needed for all decisions made. It has been observed that Machine learning can improve clinical workflows in subtle ways that are distinct from how statistics has shaped medicine. However, currently most machine learning research occurs in siloes, and there are important, unresolved questions about the validation of the models.

I. Introduction

Artificial Intelligence & Machine Learning is a broad discipline that aims to understand and design systems involving statistics, algebra, data processing and knowledge analytics etc., and thus is hard to come up with a novel definition. This is a broad definition that arguably has some cross over with existing statistical techniques. The recent explosion in progress in this field and one family of techniques in particular, deep learning, where computers are programmed to learn associations based on large quantities of raw. Deep learning systems have been studied and applied extensively and set new benchmarks in areas of the economy where high quality digital data are plentiful and there is a strong economic incentive to automate prediction tasks.

This report is structured as follows: First, we briefly touch-down on Machine Learning basics and various flavors of the same. Then we dig into the role of Machine learning specific to the usage in health sector. We understand the need of the study. We also explore potential applications of Machine Learning in health sector and the ways in which this could transform health sector to achieve more by improving efficiency, effectiveness, equity and responsiveness of public health and health care services. Then we discuss two very important association to Machine Learning, namely Interpretable Machine Learning in Health sector and applying IoT and ML for personalized healthcare. Lastly, we also brainstorm on how Machine learning can be further leveraged in health sector to create impact.

Depending on the nature of learning, Machine learning implementations are classified into following major categories:

1. **Supervised learning:** Learns from training data and associated target responses, such as classes or tags, the learning is then used to predict the correct response when posed with new examples. This approach is indeed like human learning under the supervision of a teacher. The student derives general rules from the specific examples, the teacher provided to the student to memorize, and the rules tend to become better and accurate depending on how good the training examples are

2. **Unsupervised learning:** Learns from simple examples without any associated response, leaving to the algorithm to determine the data patterns on its own. These further tend to restructure the data into something else, such as new features that may represent a class or a new series of un-correlated values. This approach resembles the methods humans use to categorize certain objects or events into the same class, such as by observing the degree of similarity between objects.

3. **Reinforcement learning:** Learns from simple examples without any associated response, similar to Unsupervised learning, however, they can be accompanied with positive or negative feedback according to the solution the algorithm proposes. In the human world, it is just like learning by trial and error. Errors help to learn because of the penalty added, such as cost, loss of time, regret, pain etc.

4. **Semi-supervised learning:** Learns from incomplete training signal, i.e. a training set with some of the target outputs missing. Transduction, which is a special case of this principle, where the entire set of problem instances is known at learning time, except that part of the targets are missing.

Also, depending on the nature of output required, Machine learning implementations are classified into following major categories:

1. **Classification:** Inputs are divided into two or more classes, and the model must assign unseen inputs to one or more of these classes. This is typically tackled in a supervised way. Example: Spam filtering, where the classes could be “spam” and “not spam”

2. **Regression:** Typically, also tackled as a supervised problem. Like Classification problems, except outputs are continuous rather than discrete

3. **Clustering:** Unlike in classification, the groups are not known beforehand, making this typically an unsupervised task, where a set of inputs is to be divided into groups/clusters.

In the last few years, machine learning methods were heavily featured in mainstream medical literature. Some latest examples are classified images of retinopathy and breast cancer metastases at a level equal to or better than clinical experts. Using tens of millions of data points from numerous live patient datasets with well-developed models demonstrate diagnostic acumen that surpasses human capabilities and do so at scale.

II. Need of Machine Learning in Health sector

To answer this, let's go to the roots of the problem that current health sector faces, and how machine learning could help overcome those. Globally, health sector faces multiple challenges, such as:

1. Increasing burden of illness, multimorbidity and disability driven by ageing and epidemiological transition
2. Greater demand for health services, higher societal expectations and increasing health expenditures
3. Inefficiency, with poor productivity, that exists against a background of fiscal conservatism, with misplaced economic austerity policies that are constraining investment in health systems, and
4. The bulk of medical knowledge is unstructured information within the variety of plethora of completely different notes, images, audio and video recording, reports, and discharge summaries. Quantifying and analyzing a conversation between the supplier and the patient are terribly exhausting and the conversation is incredibly personalized and might take many alternative directions

To overcome these challenges and to achieve universal well-being, fundamental transformation of health systems is important. Machine learning, with the newest cutting edge technologies in digital transformation, holds the promise of achieving more with less, and could be the catalyst for such a transformation, but the nature and extent of this promise has to be systematically assessed.

However, the impact of digital technology on health systems has been equivocal. Bringing machine learning models from equations derived on a blackboard to care at the bedside requires intense transdisciplinary collaboration, alignment of goals, and capabilities that are hard to find in health sector today.

III. Application of Machine Learning in Health Sector

Machine learning shows a vital performance in health sector, which includes computer-aided diagnosis, image registration, image annotation, image-guided medical aid, and image database retrieval, multimodal image fusion, medical image segmentation, where deficiency might be incurable. Machine learning is applied in health sector to decrease the increasing price of healthcare and to create an improved patient-doctor clinical communication. These solutions are used for an inordinateness of health-relevant uses; some include serving to clinicians identify a lot of customized prescriptions and therapy for patients and additionally serving to patients identify once and if they must record follow up appointments.

The Machine learning techniques are useful in identifying complicated patterns within prosperous and huge data. This is especially well-suited to clinical applications, particularly those people who rely on advanced genomics and proteomics measurements. It is often used in numerous illness diagnosing and detection. In medical applications, machine learning algorithms, by suggestions of implementing useful health-care system, will manufacture higher decisions regarding treatment plans for patients.

Health sector is utilizing the methods to forecast wait times for patients in emergency department waiting for places. Factors like patient information, illness symptoms and information, emergency department charts, and even the layout of the hospital room etc. is used to develop the model, which derives wait times. So, machine learning application could help patients by decreasing price, rising accuracy, and reducing wait times.

Two core information processing tasks are involved while provisioning health care:

1. **Screening and diagnosis:** Classification of cases based on history, examination and investigations, and

2. **Treatment and monitoring**: Planning, implementation and monitoring of a multistep process to deliver a future outcome.

Both the above information processing tasks involve hypothesis generation, hypothesis testing and action. Machine learning can reveal previously hidden trends in data, analyzing which has the potential to improve hypothesis generation and hypothesis testing tasks within a health system and thus has the potential for substantial impact both at the individual patient and system level.

Machine learning can find pattern in data by exploiting existing statistical techniques, utilizing methods that are not based on a priori assumptions about the distribution of the data. The pattern can then in turn be used to formulate hypotheses and hypothesis tests. Thus, machine learning models can incorporate many more variables and are generalizable across a much broader array of data types and can produce results in more complex situations.

IV. Machine Learning Techniques used in Health sector

Below are some of the most popular Machine Learning models used to solve problems in health sector:

1. **Naive Bayes classification:** Bayesian classifiers can identify the class membership probabilities based on given class label. It performs one scanning of data and hence classification is easy

2. **Decision Tree:** Mostly used for classification containing internal node and one leaf node with a class label. The construction is very simple, which won't require any parameters, thus making it quite popular

3. **Support Vector Machine:** A collection of training samples is given, and each sample is divided into different categories. Mainly used for classification and regression problems

4. **K-nearest neighbor:** Frequently used approach for classification of samples. In this technique, distance measure from N number of training samples are calculated and then categorized

5. **Regression Tree:** In regression trees, the target variables are represented as categorical and continuous. These variables are then used to predict values in the tree

Below are important medical areas, where machine learning models are used:

1. **Heart Diseases:** The purpose is to increase the accuracy of diagnosing heart diseases. Naïve Bayes, Support Vector Machines, and Bayes Net have been used so far to solve this problem, and obtained accuracies of 74%, 94.60% and 84% respectively

2. **Diabetic Diseases:** The purpose is to increase the accuracy of scrutiny of diabetic diseases. Naïve Bayes, Decision Trees, and Regression Trees have been used so far to solve this problem, and obtained accuracies of 79.56%, 76.95% and 77.48% respectively

3. **Breast Cancer:** It is one of the top leading reasons of death of women in the United States and in Asia countries. The purpose is to identify breast cancer risks in women. j48, Naïve Bayes, Support Vector Machines and K-nearest neighbor etc. have been used so far to solve this problem, and obtained accuracies of 94.2%, 82.60%, 96.40%, 95.15% respectively

4. **Thyroid Cancer:** The purpose is to predict thyroid diseases among individuals. SVM and fuzzy map with data mining algorithms are readily used for diagnosis of various diseases

5. **Brain Imaging:** These are instrumental for mining vast amounts of neural data of ever-increasing measurement precision and detecting minuscule signals from an overwhelming noise floor

6. **Cardiac MRI Analysis:** Cardiac MRI scans are frequently used to measure heart structure and function that guide patient care and treatment recommendations. Improving the performance of these measures could improve patient management and outcomes.

V. Interpretable Machine Learning in Health Care

While learning about the need and application of machine learning algorithms in health sector, it is important to understand Interpretable Machine Learning, because the usage is quite risky as it deals with human lives. When a machine learning model makes certain predictions, often questions are asked to provide explanation on why certain model is preferred over other. This explanation is called Interpretable Machine Learning.
In many domains, where predictions of machine learning systems need user trust, just providing traditional machine learning metrics like accuracy score, precision, and recall etc. may not be enough. The expansion of these techniques into fields like healthcare after decades of usage have led to an increased emphasis for explanations of machine learning systems. Decision makers in healthcare note interpretability of model predictions as a priority for implementation and utilization. Decisions based on machine learning predictions could inform diagnoses, clinical care pathways, and patient risk stratification, among many others.

VI. Applying Internet of Things (IoT) in Personalized healthcare

Currently, Doctors most of the times, prescribe medicines by trial and error and a one-size-fits-all approach. Although, it is known that the response to a drug might be different for each person. While some patients may respond to a drug positively, whereas some patients may have either minimal effect or severe side effects from the same medication. According to some studies, patients only fill and take prescriptions as instructed only half of the time, reasons contributing to the fact that people are getting more aware towards the side effects of some medications. Personalized Healthcare can improve the quality of care and decrease cost at the same time, and also help us predict the right therapy with the fewest side effects for individual patients.

It utilizes the internet to allow the exchange of information between humans and devices. Concepts like smart home, smart transportations, etc. are all by-products of IoT. Internet of Things (IoT) has gathered huge attention during the past few years for its potential to make major advancements in the healthcare system. With the rapid growth in cloud computing it is being used in almost every field including health sector and has been highly benefitted. This is because IoT devices in health sector uses wearable sensors that gather a huge volume of meaningful data from people worldwide and this huge volume often makes it challenging to work with this volume of data.

Patient data collected from such devices is one of the focuses of this new advancement. Other sources of data are Electronic health records (EHR), wearables and mobile devices, web-based information and social media. Personalized healthcare applies Machine Learning techniques heavily to the collected dataset to improve disease progression technique, disease prediction, and

patient self-management. These techniques are widely used in this regard to develop analytic models. These models are then integrated into different healthcare service applications and clinical decision support systems. These models mainly analyze the collected data to identify behavioral patterns and clinical conditions of the patient. For example, a user's wearable watch may be used to collect personal exercise routine, heart rate metrics, sleep pattern and then these models identify the patient's improvements, habits and anomaly in daily routine, changes in sleeping and mobility, eating, drinking and digestive pattern. Based on those patterns, the systems recommend lifestyle advice, special treatment and care plans for the patient. The doctors and clinical authorities also engage in the care plan process to validate lifestyle advice.

However, the field being so vast and dynamic, there are many uncertainties and a grey area when it comes to applying machine learning in the same context. Medical, behavior and lifestyle data in nature are very sensitive. There might be different types of biased involved in the process of data collection and interpretation. All these could finally lead to an incorrect decision from the system without the user's knowledge.

Many tech companies invested and are actively investing in developing a machine learning based personal healthcare system. So is the case in academia, where the researchers also have been developing algorithms to train the machine to diagnose complex symptoms from 2d and 3d images.

One of the successful case studies of Personalized Healthcare with IoT and machine learning is the diabetic management. The food habit and insulin response of each person are different, which is key to identify diabetes patterns and trends. Bluetooth enabled blood sugar monitoring provides dietary advice to the patient when and how much food should be taken and the dietary advice on the impact of unsupervised eating. Personal mobile devices can be used to pop-up such advices.

VII. Potential future research areas

Machine learning has now become a "General Purpose Technology", which can be improved over time and has the potential to spawn complementary innovations. To see how this general trend might apply in the health sector, it is useful to examine the clinical area that is currently best represented in machine learning literature, which is diagnostic radiology. Machine learning will create processes performed by a hybrid of *human and computer*. These instances offer the potential to achieve optimal combination of leveraging human ability to generate hypotheses, collaborate and oversee systems to harness Machine learning's ability to analyze large volumes of data to find associations with predictive power or optimize against a success criterion. Creation of high resolution clinical data sets and the necessary mechanisms for sharing of data and collaborative investigation to establish both efficacy and safety can be seen as a huge investment opportunity to seed growth in Machine Learning. There are emerging uses of machine learning in health sector. For example, Epic Systems Corporation's cognitive computing platform supports machine learning models; the Food and Drug Administration has approved software to assist with medical imaging segmentation etc. The adoption of such technologies and models serves as a key foundation for machine learning to diffuse across institutions into clinical care operations.

VIII. Conclusion

Machine learning techniques are crucial in different business fields. Health sector is one of the leading business areas, facing more problems and it is becoming more expensive. Hence applying Machine Learning in this sector is of utmost priority in today's technological advancement challenges. In this report we talked about basics of Machine Learning such as how it can be defined and categorized based on nature of learning and output. Next, we understood the challenges faced by health sector that Machine Learning aims to resolve. Further we touched down on the application of Machine Learning in this field, and the popular techniques such as Support Vector Machine, Decision Tress and Naïve Bayes. We also observed from the accuracy reports of various techniques based on previous research that choice of a suitable algorithm to apply is a delicate process, and it depends on the problem and example scenario. Hence, we further understood the concept of Interpretable Machine Learning, where the explanation becomes important for decision makers on which technique to prefer over others. IoT devices with sensors and wearables accumulate significantly high amount of health data, in this report we also briefly discussed about how this big data can be utilized to find personalized health patterns and could be highly beneficial to health sector. In future, we can try to improve the accuracy of predictions by using different machine learning algorithms. To benefit from the full potential of machine learning in health care, we must acknowledge breakthroughs in technology and adoption, address barriers to progress, and critically reflect on the strategic priorities necessary to bring health care into a new digital age.

References

[1] T. Mitchell, "Machine Learning", McGraw Hill. P. 2, 1997.

[2] Apoorva Mishra and Anupam Shukla, "From Machine Learning to Deep Learning Trends and Challenges", CSI Communications, December 2018.

[3] LP. Kaelbling and Aw. Moore, "Reinforcement Learning A Survey", Journal of Artificial Intelligence Research, vol-4, pp.237-285, 1996.

[4] T.G. Thompson and D.J. Brailer, "The decade of health information technology: Delivering consumer centric and information-rich health care: Framework for strategic action", 2007.

[5] K. Rajalakshmi, S. Chandra Mohan, S. Dhinesh Babu, "Decision Support System in Healthcare Industry", International Journal of Computer Applications, vol.9, 2013.

[6] Rohan Bhardwaj, Ankitha R. Nambiar, "A Study of Machine Learning in Healthcare", IEEE 41st Annual Computer Software and Applications Conference, 2017.

[7] Mrs. J. Sukanya, "Applications of Big Data Analytics and Machine Learning Techniques in Health Care Sectors", International Journal of Engineering and Computer Science, vol.6, pp.21963-21967, 2017.

[8] P. Sharma and A. P. R. Bhartiya, "Implementation of Decision Tree Algorithm to Analysis the Performance", vol. 10, pp. 861–864, 2012.

[9] Jalpaiguri et al., "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, vol.10, pp. 2137-2159, 2017.

[10] Parthiban, G. and Srivatsa, S.K., "Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients", International Journal of Applied Information Systems, vol.3, pp.25-30, 2012.

[11] Iyer A., Jeyalatha S. and Sumbaly R, "Diagnosis of Diabetes Using Classification Mining Techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP), vol.5, pp. 1-14,2015

[12] Chaurasia, V. and S. Pal," Data Mining Approach to Detect Heart Disease", International Journal of Advanced Computer Science and Information Technology, vol.2,pp.56-66,2018.

[13] Vembandasamy et al., “Heart Diseases Detection Using Naive Bayes Algorithm”, vol.2, pp. 441-444, 2015.
[14] Kumari, V.A. and R. Chitra, “Classification of Diabetes Disease Using Support Vector Machine”, International Journal of Engineering Research and Applications, vol.3, pp. 1797-1801, 2013.
[15] Sarwar, A. and Sharma, V., “Intelligent Naïve Bayes Approach to Diagnose Diabetes Type-2. Special Issue”, International Journal of Computer Applications and Challenges in Networking, Intelligence and Computing Technologies, vol. 3, pp.14-16, 2012.
[16] Shrivastava et al., “An Overview on Data Mining Approach on Breast Cancer data”, *International Journal of Advanced Computer Research, vol.*3, pp. 256-262, 2013.
[17] E. Venkatesan et al., “Performance analysis of decision tree algorithms for breast cancer classification”, Indian Journal of Science and Technology, vol.8, pp.1-8, 2015.
[18] K. Sivakami, “Mining Big Data: Breast Cancer Prediction using DTSVM Hybrid Model”, International Journal of Scientific Engineering and Applied Science vol.1, pp.418-429, 2015.
[19] S. Shajahaan et al., “Application of Data Mining techniques to model breast cancer data” International Journal of Emerging Technology and Advanced Engineering, vol.3, pp.362-369, 2013.
[20] M. R. Nazari Kousarrizi et al., “An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification”, International Journal of Electrical& Computer Sciences, vol.1,pp.13-19,2012. IEEE Conference Record # 42487; IEEE Xplore ISBN:978-1-5386-0965-1
[21] Muhammad Aurangzeb Ahmad, Carly Eckert, Ankur Teredesai et. al., KenSci Inc. “Interpretable Machine Learning in Healthcare”, 2018 IEEE International Conference on Healthcare Informatics.

[22] Puja Ghosal, Debanjan Das, Indrajit Das et al., “Extensive Survey on Cloud-based IoT-Healthcare and Security using Machine Learning”, 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)
[23] K. Shailaja, B. Seetharamulu, M.A. Jabbar et al., “Machine Learning in Healthcare: A Review”, Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018)
[24] J. Qi, P. Yang, G. Min, O. Amft, F. Dong, and L. Xu, "Advanced Internet of Things for personalised healthcare systems: A survey," *Pervasive and Mobile Computing*, vol. 41, pp. 132-149, 2017.
[25] M. Ermes, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE transactions on information technology in biomedicine*, vol. 12, no. 1, pp. 20-26, 2008.
[26] A. Ara and A. Ara, "Case study: Integrating IoT, streaming analytics and machine learning to improve intelligent diabetes management system," in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, 2017, pp. 3179-3182.
[27] R. Vargheese and Y. Viniotis, "Influencing data availability in IoT enabled cloud based e-health in a 30 day readmission context," in *10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2014, pp. 475-480.
[28] Mark Sendak, Michael Gao, Marshall Nichols, Anthony Lin, Suresh Balu et al., “Machine Learning in Health Care: A Critical Appraisal of Challenges and Opportunities”, eGems
[29] Trishan Panch, Peter Szolovits, Rifat Atun et al., “Artificial intelligence, machine learning and health systems, J. Glob Health, 2018 Dec; 8(2): 020303
[30] Farhad Ahamed, Farnaz Farid et al., “Applying Internet of Things and Machine-Learning for Personalized Healthcare: Issues and Challenges”, 2018 International Conference on Machine Learning and Data Engineering (iCMLDE)