Machine Learning used in Diagnosing and Predicting Chronic Diseases

Aarushi Banerjee

Abstract

Machine learning has made a huge impact in day to day life and processes, and is showing potential to continue changing the future. One of the fields that machine learning is showing potential in is healthcare, specifically, diagnosing and predicting chronic diseases, which are diseases that last for one or more years. The goal of this paper is to summarize some ways machine learning is being used for this, such as analyzing medical images, and how machine learning models can be developed more in the future on more specific tasks, like predicting cancer from a single cell. Additionally, this paper talks more specifically about machine learning. Finally, this paper also discusses some limitations in current models, such as models developing a bias and cost of implementing machine learning models.

Introduction

Chronic diseases are the world's biggest killers - and a sizable number of people with chronic diseases are either diagnosed late or not diagnosed at all. From a study in Scotland amongst patients who had been diagnosed with a type 2 heart attack, 3 in 5 patients showed signs of coronary artery disease, but were undiagnosed [1]. Over 50% of people with diabetes are undiagnosed [2] and, from 2016 to 2020, 45% of lung cancer patients were diagnosed at a late stage in the US [3]. A chronic disease is a disease that lasts for one year or more and requires medical attention or limit daily activities. Some examples of these are diabetes, cancers, and cardiovascular diseases which are leading causes of death and disability in the United States. Some of these diseases can start to build up and develop over time because of bad lifestyle habits, such as tobacco use and a bad diet. They can also show early signs and symptoms which can often go undiagnosed or diagnosed when the disease has already progressed. If there is a way to detect chronic diseases before it becomes too serious, then we can save lots of lives. This is where machine learning has potential for coming in. How can machine learning be involved in diagnosing chronic diseases from early stages?

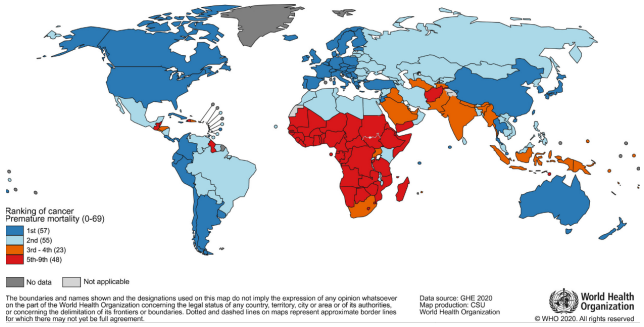

Figure 1-Premature mortality of cancer under 70 years old in 2019; Reprinted from American Cancer Society, Global Cancer Statistics 2020 [1]

Current Innovations

Professionals in the healthcare field have begun the use of machine learning models in the diagnosis of chronic diseases. Machine learning is helping physicians in decision making when they have to decide whether something could be classified as a chronic disease or what the next steps are for the patient. One way is by analyzing patient data, factors like lifestyle habits and diet, to predict their chance for developing a chronic disease which is caused mainly from these factors, like diabetes or cardiovascular diseases[4]. An example of this is using 2 machine learning algorithms, CNN (convolutional neural network) and KNN (K-nearest neighbor), where k is the number of nearest neighbors in a dataset. CNN is a deep learning algorithm used for disease prediction, it can distinguish between its inputs and calculate the output using a mathematical expression. KNN is a supervised machine learning algorithm that can give the final disease prediction through classification. It does this by comparing new data to existing data, by calculating the Euclidean distance from its nearest neighbors, and can find an exact match in a dataset[5]. Additionally, machine learning analyzes medical images to give more accurate results. Specifically in breast cancer imaging, machine learning techniques improve the quality of computer-aided diagnosis, which already provides better automated image detection, by using the image data input as the training set for a machine learning model. CNN can be implemented here too, as it is one of the most developed deep learning algorithms. Artificial intelligence algorithms based on deep learning are most commonly used to double check images that have already been manually analyzed as negative[6]. Adding on, researchers are finding ways to incorporate AI in potential cancer treatments, like immunotherapy, to see whether the patient is responding to immunotherapy and whether it is affecting them[7].

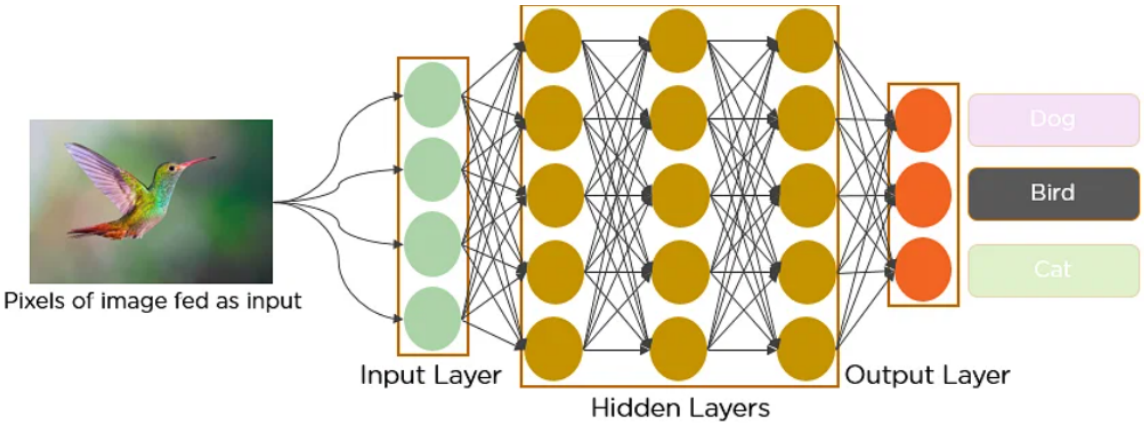

Figure 2-CNN to identify the image of a bird; Reprinted from [2]

Machine Learning

Machine learning, a branch of AI that uses data to learn and improve its accuracy, is the one being implemented in the examples above. There are two basic approaches to machine learning, supervised and unsupervised machine learning. Supervised machine learning uses

labeled data sets to classify data or predict outcomes. There are two types of supervised machine learning algorithms, classification and regression. Classification algorithms can learn from the data to predict an outcome and can be used in the prediction of chronic diseases [8]. Regression algorithms are used to predict continuous values, which are values that can be measured. They can be used for predicting the chance of a person developing chronic disease. Unsupervised machine learning is used to analyze and cluster unlabeled data sets, so it can be used to classify chronic diseases. Clustering algorithms are a part of unsupervised machine learning, it is used to group unlabeled data based on similar patterns[9].

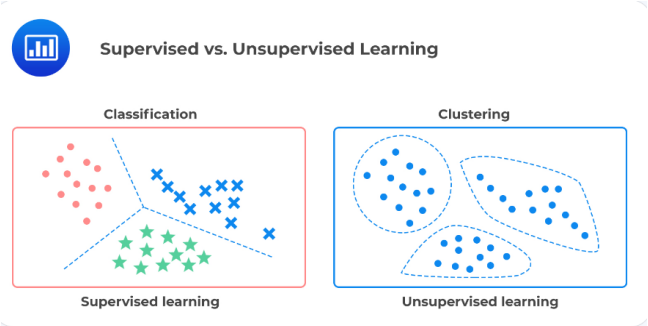

Figure 3-Supervised vs. Unsupervised machine learning diagram; Reprinted from Analyst Prep[3]

Future Work

One way that machine learning models can be used and developed more, specifically in cancer diagnosis, is by being able to predict whether a single cell will be cancerous in the future. Cancer cells are different from normal cells because they keep dividing, are invasive and since they divide quickly, they don't mature into a distinctive type of cell[10]. Because of these characteristics, cancer cells can look distinctly different from other cells underneath a microscope, often having different sizes and abnormal shapes/nucleus[11]. The body can also have precancerous cells, which looks abnormal but doesn't necessarily always develop into cancer[12]. Machine learning models to predict cancer from a single cell are developed on single-cell RNA sequencing datasets, which have a lot of cells in a sample with lots of genes expressed in each cell. Some of these genes show big differences between the cells, they are called high-variable genes and are the ones that show the unique characteristics of each cell in the dataset[13]. Machine learning models can analyze these datasets to find similarities and understand differences between the data along with using known genes that are important to transfer knowledge between datasets, making it easier to compare them[14].

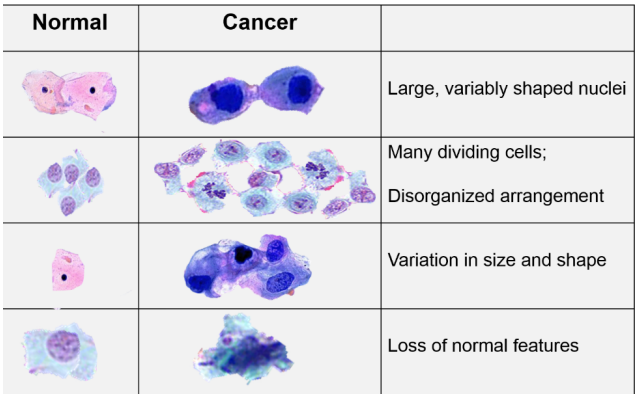

Figure 4-Diagrams of cancerous vs. normal cells; Reprinted from The Biology of Cancer [4]

Discussion

The first recorded description of cancer goes all the way back to about 3000 BC in Egypt and was found in an Egyptian textbook about trauma surgery. In this textbook it is written that this disease has no treatment[15]. Cancer is such a disease that not only occurs in humans, but in plants and animals too. Our understanding of cancer has progressed a lot over the years of human history, which has allowed researchers to develop many treatments, such as chemotherapy. Cancer is an extremely complex disease, with over 100 different types of cancers, each having a different range of seriousness and difficulty to treat. So, a treatment that is effective for one type of cancer may not be as effective for another. The best 'cure' for this disease is early detection and this is where machine learning is showing great potential. The development of machine learning started in 1943 and has gone on to completely change the way things work[16]. From making everyday tasks easier, to improving the quality of programs, machine learning will impact the healthcare field greatly as well. Machine learning can be the piece that will revolutionize cancer detection from early stages.

Limitations

There are still some limitations in machine learning models for cancer/chronic disease detection. Machine learning models require datasets, and are trained/developed on those datasets. So, one limitation is the type of data a model is trained on. If a model has been developed on a dataset that does not show diversity in age, ethnicity, socioeconomic backgrounds or underrepresents certain populations then that model will not provide accurate results for everyone[17]. This could result in bias being built into the model, because it was trained on homogenous datasets. Additionally, if a dataset has more diverse diseases but the classification methods of the machine learning model have only been trained with a few specific diseases, then the results will also not be as accurate. Methods that are used in research may not work as well in real medical settings because the model could have limited data or different features than the data presented to the model in a real-time medical setting[18].

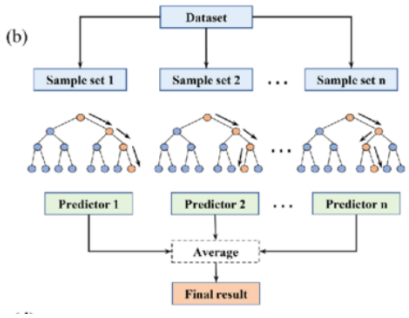

Figure 5-How data is input in a machine learning model; Reprinted from ResearchGate[5]

Another limitation of using machine learning models in cancer detection is that some medical professionals may be hesitant to use these models. This could be because they would want to wait until its performance has been proven to be able to work with more diverse data. They also might be hesitant because they are not familiar with how AI models would be integrated in their work[19]. The cost of machine learning models and maintenance can also limit its use. Implementing AI/machine learning models in the healthcare field can cost between $20,000-$1,000,000, but they will also need a specific team set up and many other resources as well. The daily cost of filling these positions can be from $600-$1500[20]. As these models get used more in the healthcare field, the cost will also go down. But, the cost for starting to use machine learning models is high, which can limit its use.

Conclusion

Machine learning has the potential to impact chronic disease diagnosis greatly, especially early cancer diagnosis. It has already been implemented in several use cases like analyzing patient data and medical images. Machine learning has developed a lot over the years and with advanced computational models like CNNs and KNNs, they are even more accurate. However, some limitations exist for current models, which can be overcome as researchers continue to test and improve these models. Overall, the best cure for chronic diseases is catching the disease right from the start, and being able to predict the disease before it happens. The development of machine learning has shown more precise and accurate ways of doing this, and hopefully in the future, chronic disease will be caught before it becomes serious.

References

Introduction:

[1] Williamson Laura. “Undiagnosed heart disease may be common in people with heart attacks not caused by clots”. Heart Attack and Stroke Symptoms. 28 March, 2022, https://www.heart.org/en/news/2022/03/28/undiagnosed-heart-disease-may-be-common-in-people-with-heart-attacks-not-caused-by-clots

[2] International Diabetes Federation. “Facts and Figures”. 2021 https://idf.org/about-diabetes/diabetes-facts-figures/

[3] Centers for Disease Control and Prevention. “U.S. Cancer Statistics Lung Cancer Stat Bite”. 23 June, 2023, https://www.cdc.gov/cancer/uscs/about/stat-bites/stat-bite-lung.htm

[4]Rashid, Junaid, et al. "An augmented artificial intelligence approach for chronic diseases prediction." Frontiers in Public Health 10 (2022): 860396. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9008324/

[5] Alanazi, Rayan. "Identification and prediction of chronic diseases using machine learning approach." Journal of Healthcare Engineering 2022 (2022). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8896926/

[6]Zheng, Dan, Xiujing He, and Jing Jing. "Overview of artificial intelligence in breast cancer medical imaging." Journal of Clinical Medicine 12.2 (2023): 419. https://www.mdpi.com/2077-0383/12/2/419

[7] Emory Winship Cancer Institute. “New AI-based biomarker can help predict immunotherapy response for patients with lung cancer”. 4 January, 2023. https://winshipcancer.emory.edu/about-us/newsroom/press-releases/2023/new-ai-based-biomarker-can-help-predict-immunotherapy-response-for-patients-with-lung-cancer.html

[8]Alanazi, Rayan. "Identification and prediction of chronic diseases using machine learning approach." Journal of Healthcare Engineering 2022 (2022). https://www.hindawi.com/journals/jhe/2022/2826127/

[9] Delua, Julianna “Supervised vs. Unsupervised Learning: What’s the Difference?” IBM. 12 March, 2021,

https://www.ibm.com/topics/unsupervised-learning#:~:text=the%20next%20step-,What%20is%20unsupervised%20learning%3F,the%20need%20for%20human%20intervention.

[10] Tan, Bradford. “Cancer cells vs. Normal Cells”. City of Hope. 15 November, 2021, https://www.cancercenter.com/cancer-cells-vs-normal-cells#:~:text=Normal%20cells%20follow%20a%20typical,%2C%20liver%2C%20lungs%20and%20pancreas.

[11] “Cancer Cells”. Cancer Research UK. 6 October, 2023, https://www.cancerresearchuk.org/about-cancer/what-is-cancer/how-cancer-starts/cancer-cells#:~:text=Unlike%20healthy%20cells%2C%20cancer%20cells,have%20a%20chance%20to%20mature.

[12] Lynne Eldridge. “What It Means if You Have Precancerous Cells”. Verywell Health. 7 July, 2023, https://www.verywellhealth.com/what-are-precancerous-cells-2248796

[13]Jovic, Dragomirka, et al. "Single‐cell RNA sequencing technologies and applications: A brief overview." Clinical and Translational Medicine 12.3 (2022): e694. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8964935/#:~:text=The%20single%E2%80%90cell%20RNA%20sequencing%20data%20set%20is%20high%E2%80%90dimensional,genes%20expressed%20in%20each%20cell.

[14] Dohmen, Jan, et al. "Identifying tumor cells at the single-cell level using machine learning." Genome Biology 23.1 (2022): 123. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02683-1#Sec1

Discussion:

[15] “Understanding What Cancer Is: Ancient Times to Present”. American Cancer Society, 4 January, 2018, https://www.cancer.org/cancer/understanding-cancer/history-of-cancer/what-is-cancer.html

[16] Firican, George. “The history of Machine Learning”. LightsonData, 2015, https://www.lightsondata.com/the-history-of-machine-learning/#t-1641519420904

[17] Hickman, Sarah E., Gabrielle C. Baxter, and Fiona J. Gilbert. "Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations." British journal of cancer 125.1 (2021): 15-22.
https://www.nature.com/articles/s41416-021-01333-w#Sec10

[18] Rashid, Junaid, et al. "An augmented artificial intelligence approach for chronic diseases prediction." Frontiers in Public Health 10 (2022): 860396.
https://www.frontiersin.org/articles/10.3389/fpubh.2022.860396/full

[19] “Machine Learning’s Potential to Improve Medical Diagnosis”. US Government Accountability Office,10 November, 2022,
https://www.gao.gov/blog/machine-learnings-potential-improve-medical-diagnosis

[20] Ali, Afnan. “The Financial Cost of AI in Healthcare – A Comprehensive Guide for 2023”. Folio3 Digital Health, 25 September, 2023,
https://digitalhealth.folio3.com/blog/cost-of-ai-in-healthcare/amp/

Images:

[1]Sung, Hyuna, et al. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: a cancer journal for clinicians 71.3 (2021): 209-249.
https://acsjournals.onlinelibrary.wiley.com/doi/10.3322/caac.21660

[2] Biswal, Avijeet. “Convolutional Neural Network Tutorial”. Simplilearn, 26 October, 2023,
https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network

[3] “ Supervised Machine Learning, Unsupervised Machine Learning, and Deep Learning”. Analyst prep, 6 March 2021,
https://analystprep.com/study-notes/cfa-level-2/quantitative-method/supervised-machine-learning-unsupervised-machine-learning-deep-learning/

[4] LaMorte W, Wayne. “ Characteristics of Cancer Cells”. The Biology of Cancer, Boston University School of Public Health, 4 May, 2016,
https://sphweb.bumc.bu.edu/otlt/mph-modules/ph/ph709_cancer/ph709_cancer7.html

[5] Lei, Guoqing, et al. "A comparison of physical-based and machine learning modeling for soil salt dynamics in crop fields." Agricultural Water Management 277 (2023): 108115.
https://www.sciencedirect.com/science/article/pii/S037837742200662X