IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

Using Machine Learning for Chronic Disease Diagnosis and Prediction

Vishal Prasad[1], Utkarsh Raj[2], Umang Dobhal[3]
UG Students, Department of Computer Science and Engineering[1,2,3]
Dronacharya College of Engineering, Gurugram , India

Abstract: *Machine learning had a strong effect on day-to-day life processes, and in the future, it's going to continue influencing this field. A key where machine learning illustrates promise in healthcare, especially in diagnosing and predicting chronic diseases, conditions that last for at least a year. This paper examines various methods through which machine learning is used in healthcare, consider how these models might be improved to perform specialized tasks, like detecting cancer from individual cells, as they are used to analyse medical images. The paper also explores the inner workings of machine learning and enlightens on some of its current limitations. The article addresses concerns such as model bias and the significant expenses associated with implementing machine learning solutions.*

Keywords: Machine Learning, Artificial Intelligence, Cancer Cells.

I. INTRODUCTION

Long-lasting health conditions are a top reason people pass away across the globe. Still, a lot of folks either don't find out they have these issues or learn about them too late. Chronic diseases are among the leading causes of death worldwide, yet many individuals with these conditions go undiagnosed or receive a diagnosis later than they should. Research in Scotland on people who suffered a certain kind of heart attack showed that 60% had signs of blocked heart arteries but hadn't been told about it before [1]. Over 50% of individuals with diabetes are unaware of their condition [2], and close to 50% of lung cancer diagnoses in the U.S. between 2016 and 2020 were made when the disease was already in a late phase [3].

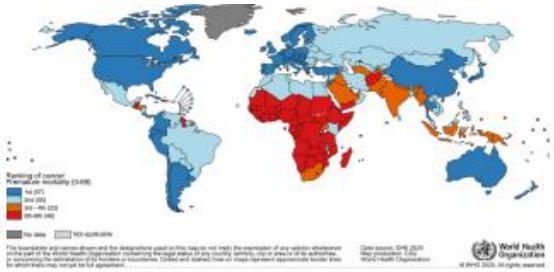

Figure 1-Premature mortality of cancer under 70 years old in 2019; Reprinted from American Cancer Society, Global Cancer Statistics 2020 [1]

Long-term illnesses such as diabetes, cancer, and heart conditions can persist for more than a year, usually requiring continuous medical attention or restricting everyday life. They are among the top reasons for death and long-term health problems in the U.S. These ailments often get worse over time because of unhealthy living habits, like smoking and eating poorly. Long-lasting health troubles sometimes begin with minor hints that are easy to miss, which means they're often not caught until they've gotten pretty serious. If we could spot these problems sooner, we could prevent a lot of heartache. This is where smart computer systems could be super helpful. How can we use these smart systems to spot long-term health issues early on?

II. CURRENT INNOVATION

To understand long-term health problems, doctors are becoming better assisted by high-tech research tools. Gigabytes of pertaining information are searched through to understand if an individual is about to become ill, might have issues

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

like diabetes, heart problems, or should undergo additional examination. It is a clever way of "inspecting" the manner in which people live and what they ate and help identify those who are prone to acquire long-term illnesses [4]. A lot of help for doctors who want to make smart judgments about an individual's health and know what to do afterward to help the person get well or avoid getting sick. Two methods to predict diseases are convolution neural networks and K-nearest neighbors. CNN is a form of deep learning that is effective at pattern recognition and can use medical images to determine a probable diagnosis. They produce results by performing complicated mathematical computations. K-nearest neighbors, on the other hand, is a less intricate system that categorizes data by taking into account how comparable they are to other objects in a dataset. This technique computes the distance to other data points to determine a prediction by finding the closest matches to aid make a final judgment [5]. This contributes to more precise medical image analysis and disease prediction. Super-smart computer programs are boosting the front against breast cancer by the way doctors diagnose using scans. The super-smart programs fine-tune by learning from many breast scans pictures. Typically, artificial intelligence models based on deep learning are employed to verify images that have previously undergone manual assessment and were determined to be negative [6]. Scientists are also finding ways to combine the smart techniques with cancer treatments. For example, strong immune system zapping medicines to check where they're working for the patients [7].

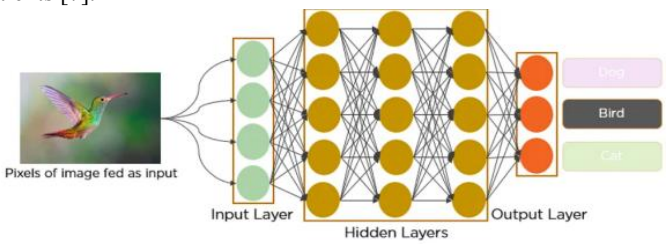

Figure 2-CNN to identify the image of a bird; Reprinted from [2]

III. MACHINE LEARNING

Machine learning, a subset of artificial intelligence, harnesses data to refine its accuracy through learning. This field is primarily divided into two approaches: supervised and unsupervised machine learning. Supervised learning relies on labeled datasets to either categorize data or forecast outcomes. Within this realm, there are two main types of algorithms: classification and regression. Classification algorithms learn from data to predict outcomes, such as the likelihood of chronic diseases. Conversely, regression algorithms forecast continuous, measurable values and are useful for estimating the probability of someone developing a chronic condition. On the other hand, unsupervised machine learning focuses on analyzing and organizing unlabeled data. It's particularly useful for identifying clusters or groups within data, which can help in classifying types of chronic diseases based on similar patterns. Now, imagine you have a bunch of unlabeled pictures of animals. You don't know which ones are cats, dogs, or birds. Unsupervised learning helps find patterns in this data. It groups similar pictures together without any predefined labels. One common technique is clustering, which organizes data into clusters based on similarities [9].

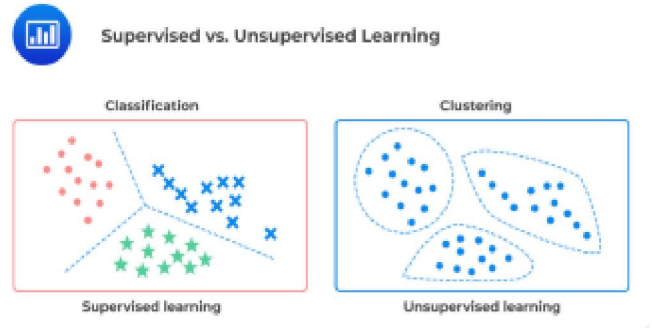

Figure 3-Supervised vs. Unsupervised machine learning diagram; Reprinted from Analyst Prep [3]

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

IV. FUTURE WORK

In cancer diagnosis, machine learning models can help predict whether a single cell might become cancerous later on. Unlike normal cells, cancer cells keep dividing, are invasive, and don't mature into specific cell types quickly [10]. Under a microscope, cancer cells can appear quite different from other cells. They often have varying sizes and irregular shapes in their nuclei [11]. In our bodies, there are cells that appear unusual but don't always turn into cancer [12]. Scientists use special data from cells to create computer programs that can predict if a cell might become cancerous. They look at lots of genes in each cell, especially the ones that are very different between cells. These unique genes help identify the cell's special features [13]. Scientists use special data from cells to create computer programs that can predict if a cell might become cancerous. They look at lots of genes in each cell, especially the ones that are very different between cells. These unique genes help identify the cell's special features [14].

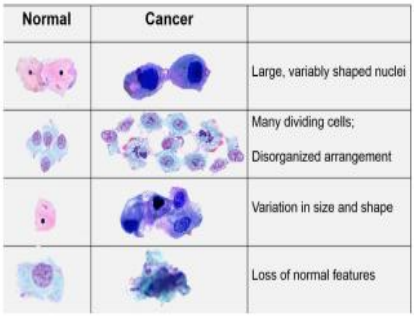

Figure 4-Diagrams of cancerous vs. normal cells; Reprinted from The Biology of Cancer [4]

V. DISCUSSION

The first mention of cancer goes back to 3000 BC to ancient Egypt, when a trauma surgery medical text described it. According to this document, the disease is incurable [15]. Humans are not the only "victims" — cancer can form in plants and animals as well. Despite the fact that we have come a long way in studying cancer throughout the history of its existence and have found effective methods of treatment such as chemotherapy, it is an extremely complex disease, and there are three main reasons for this. Cancer is a very complicated illness. There are more than 100 different types of cancer, and each type varies in how serious it is and how hard it is to treat. The success of cancer treatments can differ greatly based on the specific type of cancer, so a method that's effective for one kind may not work as well for another. The real trick to conquering this illness lies in early detection, and that's where recent advancements in technology come into play. The smart systems that have been developing since the mid-20th century are reshaping our world in many ways [16]. They're not just simplifying everyday tasks or improving various software—they're set to make a significant impact in the medical field, particularly in spotting cancer sooner than ever before. These intelligent systems might just be the breakthrough we need to transform the fight against cancer by catching it early.

VI. LIMITATIONS

Machine learning models used for detecting cancer or chronic diseases have some limitations. These models rely on datasets for training and development. However, if a model is trained on data that lacks diversity in terms of age, ethnicity, socioeconomic backgrounds, or under represents certain groups, its results may not be accurate for everyone [17]. When a model is trained on similar types of data, it might develop biases. If the dataset lacks diversity, the model's performance could suffer, especially when faced with new or different situations. Additionally, research methods may not always translate perfectly to real-world medical scenarios due to variations in data and features presented to the model [18].

Another limitation of using machine learning models in cancer detection is that some medical professionals may be hesitant to use these models. Some doctors might be cautious about using machine learning models for cancer detection because they want to see if these tools can handle a wide range of different cases. They might also be unsure about how AI will fit into their everyday work since it's a new technology for many in the medical field [19].

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

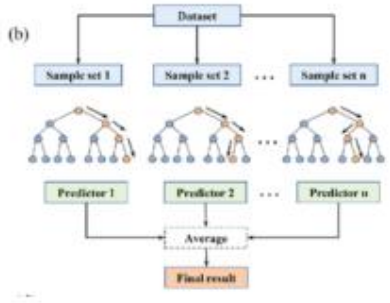

Figure 5-How data is input in a machine learning model; Reprinted from ResearchGate [5]

The expense of setting up and running machine learning systems might be a barrier. Starting such systems in healthcare could range from **$20,000 to $1,000,000**, and they require a dedicated team and additional resources. The everyday expense for these roles can vary from **$600 to $1500 [20]**. As machine learning becomes more common in healthcare, the costs are expected to decrease. However, the initial investment needed to adopt these models is quite substantial, which could restrict their widespread use.

VII. CONCLUSION

The approach to diagnosing long-term illnesses is undergoing a revolution, especially in spotting cancer early on. Innovative techniques are now being applied to sift through patient records and medical images, thanks to sophisticated systems that have greatly enhanced precision. Despite this progress, there are challenges that these systems must still surmount, and experts are diligently fine-tuning and evaluating them. The cornerstone of effectively managing chronic diseases is to catch them early, ideally before they escalate. These cutting-edge techniques have shown great promise, providing sharper and more reliable methods for early diagnosis. The aspiration is that, with ongoing advancements, we'll manage to identify chronic diseases at an initial stage, well before they intensify.

REFERENCES

[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-peop le-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-biomar ker-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, 6 https://www.ibm.com/topics/unsupervised-learning#:~:text=the%20next%20step-,What%20is%20unsupervised%20learning%3F,the%20need%20for%20human%20intervention.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

[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%20m ature.

[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

[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 7

[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/

[21]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

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

[23] " 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-learnin g-unsupervised-machine-learning-deep-learning/

[24] 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

[25] 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