Machine Learning and Deep Learning Models in the Cancer Continuum

Tiana Lobo

Abstract

Cancer is one of the leading causes of death worldwide, affecting people from all backgrounds. The need for early detection is essential, as it can prevent the spread of cancer before it takes over the body. However, cancer encompasses hundreds of classifications, making efficient diagnosis difficult. While new technology has been developed, it has trouble identifying abnormal cells, as cancer can resemble healthy cells. Recently, AI has become a ground-breaking tool within oncology, enabling early detection and characterization through advanced image analysis. Two major applications of AI are machine and deep learning models, which analyze data to make clinical predictions. These models support doctors and pathologists in diagnosis, prognosis, and therapy prediction for the patient. Furthermore, as time passes, AI can gather and record patient outcomes, improving the models' results, as well as their detection and prediction techniques. With these newfound advancements, researchers have been forming AI-based models that could potentially save millions of lives. This study will review the innovative technologies of these AI models in oncology and how they are applied within the cancer continuum.

Introduction

Cancer is a puzzling and frightening disease that researchers have struggled for years to deal with (Hauman, 2019). Over the past few decades, scientists have developed different methods, such as screenings, which help determine cancer in earlier stages. Later on, they developed strategies to predict treatment and cancer outcomes. However, accurately, and successfully carrying out these methods is an issue that physicians still struggle with. And even if one does the entire process perfectly, there is still a chance for error and no survival is completely guaranteed (Kourou et al., 2014).

However, in recent years, AI has seen a huge explosion within medicine, specifically its applications in machine learning and deep learning. Oncology is one of the fields that has seen this rise in AI usage as it provides multiple diagnostic problems and copious data for numerous methods. This is because cancer is a complex, diverse, and prevalent disease, requiring a lengthy and demanding process to cure. Fortunately, AI-based models, such as ML and DL, benefit from the environment in oncology, as they utilize data to learn the patterns and relationships of a data set. After processing the information, the model is able to make a precise clinical prediction and assist clinicians directly (Swanson et al., 2023).

Two of the main methods used to help with evaluating cancer are imaging and molecular characterization, both of which can be processed and learned by ML and DL models. Medical imaging is a powerful tool that allows for detection, localization, and monitoring of cancer. Radiology images are used for screenings and diagnosis, along with pathology images. This is vital because they can help guide treatment, and ML/DL models can analyze these data sets to evaluate imaging for various modalities (Hosny et al., 2018). Molecular characterization, such as liquid and solid tumor biopsies, can also identify tumor characteristics and estimate survival outcomes (Kwong et al., 2021). Due to the amount of information provided by these biopsies, ML and DL models can easily infer status and outcomes efficiently. Moreover, ML and DL are not only limited to these specific methods and can also help clinicians in various other ways. Overall, they can aid doctors through every step of the cancer continuum.

However, while AI-based models are extremely helpful, it is important to note that their accuracy is not 100 percent. They are still being developed, and it is not to say they should replace every step. ML and DL models can certainly speed up the process and provide clear results which help aid each step. Yet, it is still important to check the model's predictions with other data until they have been confirmed to be error-free. And while that might not happen in the near future, it is clear that they are already benefiting oncology significantly.

In this review, the new advancements in ML and DL models are explored, along with their different applications. The following sections examine how these AI-based models process data, and how they use it to evaluate current patients and come up with predictions. The use of ML and DL are analyzed through different sections, focusing on detecting, diagnosing, and treating cancer.

Early Detection

One of the leading applications of AI is machine learning, which is now being increasingly used in early detection and screenings for cancer. Machine learning (ML) uses ample amounts of data to learn patterns and structures of a data set. Oncology data sets provide crucial information for ML to be successful, contributing to its rising usage in the field. For instance, early ML used hand-crafted image features that were algorithmically extracted from images (Bera et al., 2021). These features were then entered into a traditional ML model (i.e., support vector machine) which would make a clinical prediction. The AI model goes through various training techniques to connect the patterns of a data set (i.e., transfer learning, data augmentation, regularization, and weak supervision). As a result, ML provides a consistent and detailed image evaluation that can catch cancers missed by time-constrained physicians (Swanson et al., 2023). By identifying abnormalities through patterns, it can detect malignant changes before others, aiding in early detection (Ahmad et al., 2024). Given this, ML is an obvious success within the field due to its sheer efficiency and complete analysis. These findings also support ML's effective role in assisting doctors with flagging cancer. While ML has not reached a 100 percent accuracy, as it is still developing, more data is continuously collected by the day. This leads to the perfect conditions to improve the model's skills and precision. Overall, ML is a highly effective model for early cancer detection within oncology and continues to improve further.

Building on the success of machine learning, in recent years, AI has also been incorporated into another model which helps in the detection of cancer. This is a subset of ML, referred to as deep learning (DL) which uses neural networks to analyze medical images, such as MRIs, CT scans, and mammograms (Ahmad et al., 2024). The layers of neurons create a representation of the input (embedding), which is used by the final layer of neurons to form an output prediction. As graphics processing units have become increasingly available, DL has taken over traditional ML models for many prediction tasks (Hosny et al., 2018). Also, DL models are much more flexible than ML models, relying less on feature engineering and capable of processing more unstructured data types (Boehm et al., 2023). For instance, DL models can use images from a screening mammogram to predict whether a patient will develop cancer. However, DL models require much more training data compared to ML (Swanson et al., 2023). In this sense, DL is a highly reliable model with more data processing capabilities. However, traditional ML models may be better suited for certain cases, such as data-limited tasks. It all depends on whether an institution has enough information about a certain cancer type, as that determines what model to use for the specific situation. Ultimately, DL is promising and has huge potential for aiding in detecting cancer.

Diagnosis

ML models can also be applied to help diagnosis and characterize specific types of cancer. Diagnosing cancer often involves the following steps. To begin, patients undergo non-invasive imaging, such as radiological scans. These images can reveal tissue that might indicate cancer, resulting in a biopsy being taken and sent to a pathology lab. ML can improve the diagnostic accuracy for these steps by identifying patterns (known and unknown to clinicians) that indicate the presence or absence of cancer (Swanson et al., 2023). In fact, machine learning techniques have been used to detect various cancer types (i.e., lung, brain, breast, skin, blood, hippocampus, pancreas, cardiac, colon cancer, hepatic vessel, spleen, etc.) (Painuli et al., 2022). Furthermore, a reader study designed by Kim et al. (2020) made various radiologists from five institutions evaluate mammograms with and without the aid of an ML model specifically trained on mammograms. Radiologists from multiple institutions had superior performance when working with the ML model rather than without. This suggests how ML can efficiently be used within the cancer diagnosis process. Through multiple methods, ML can detect the specific type of cancer which doctors might miss or are unable to determine. The findings also show that they can make a clinician's diagnosis more accurate and precise. Altogether, ML is a useful tool that makes the diagnosis process much easier for doctors.

Taking a different approach, DL models have a unique way to diagnose and categorize cancer. As mentioned previously, ML methods are typically used for small sized datasets, while DL methods are used for larger ones (Painuli et al., 2022). In pathology samples, DL models are able to use larger data because they operate on small image tiles and then combine the predictions into a diagnostic prediction for the whole slide (Swanson et al., 2023). Another study showcased the DL's process by building a model to diagnose breast cancer from mammograms, evaluating their model on large data sets from the US and the UK (McKinney et al., 2020). Their model had superior performance compared to six radiologists. The findings also demonstrated that in many cases, the model could replace a second reader (a standard requirement in the UK), with their models' prediction and save 88% of the diagnosis time, while maintaining high accuracy. Moreover, DL can support doctors with diagnostic procedures, such as colonoscopies, by analyzing live videos and pointing out suspicious areas of tissue (Swanson et al., 2023). In addition to assisting doctors, DL models can also work efficiently by themselves. Esteva et al. (2020) trained an Inception v3 CNN (DL model) to categorize skin cancer from skin lesion images. Using a set of biopsy-proven clinical images, the model was able to match the performance of 21 dermatologists. Given these outcomes, DL models are able to assist doctors successfully through bigger data sets. Also, these models have been able to diagnose different types of cancer independently, proving themselves capable of operating without oversight. However, only with significant training will they be able to reach this level of success. All in all, DL models are another significant component in cancer diagnosis and are an increasingly established tool in oncology.

Treatment Planning

ML also contributes to outcome prediction and treatment planning for a patient. Since ML models analyze hundreds of data sets to establish patterns, they can eventually predict clinical outcomes and cancer status based on molecular features, derived by liquid and solid tumor biopsies (Swanson et al., 2023). For instance, the Gleason grade in prostate cancer is a developed ML model which can predict the subtype or grade of a tumor. This gives physicians information about patient survival rates and the most effective treatments (Bulten et al., 2020). By processing medical images (i.e., radiology and pathology images), ML can guide treatment by suggesting the selection of chemotherapy or immunotherapy, as well as aiding radiotherapy planning (Swanson et al., 2023). This is especially significant as planning radiotherapy is a time-consuming process which benefits from the speed of ML models. McIntosh et al. (2021) displayed their efficiency through a comparison study between human and ML generated radiotherapy treatment plans for prostate cancer. ML was able to generate their treatment plans due to a clever system where they used previous plans most similar to the current patient and developed their own. Among 50 patients, ML-generated plans were selected over human-generated plans 61% of the time and reduced the planning time by 60%. These studies display the numerous benefits that ML models bring to the treatment stage of cancer. They accelerate tedious tasks and make both the doctors' and patients' experience much easier. Their successes also indicate a higher survival rate for the patients, as ML models are able to determine the best treatment planning by analyzing the cancer subtype. However, there are drawbacks, as these ML models can only determine treatment with sufficient previous information about other patients, along with the other steps. Their accuracy and methods also have not been precisely refined, so development is still required to make their predictions exact. In summary, ML is a great helper as it can predict outcomes and the best treatments much more proficiently, yet it can only do so with enough data in store.

More recently, DL models have built on these approaches to bring even more advanced techniques to help treat cancer and determine prognosis stages. To begin with, many DL models can predict patient survival from histology slides. Courtiol et al. (2019) developed one of these models and workflow for prognosis in mesothelioma. On the other hand, Bychkov et al. (2018) predicted survival with DL models for colorectal cancer by using an RNN (DL) to combine the embeddings produced by a CNN (DL) for each tile. Moreover, DL models can also predict the responses to treatments, besides just recommending the best therapy. For instance, Joo et al. (2021) developed a multimodal DL model to predict whether patients will have no cancer left after neoadjuvant chemotherapy for breast cancer. The model made predictions by combining information from pre-treatment MRIs and clinical information. As demonstrated above, there are various DL models, ranging from CNN to RNN, which have successfully helped predict outcomes and reactions, aiding treatment planning in numerous ways. While DL is still being integrated into oncology, it has proven to be extremely effective in helping doctors proceed with the right plan and therapy. Ultimately, DL offers meaningful support that enhances clinicians' ability to support and care for patients.

Conclusion

Altogether, this review examined how ML and DL models are used to assist with detecting, diagnosing, and treating cancer. They are both helpful assistants and can also work independently with a medical professional, overseeing their outcomes. These models show strong promise and are already benefiting oncology in various ways, yet they still have a long way to go before they are fully reliable in clinics.

To recap, ML/DL models are trained on various sets of data which they use to form patterns and structures of a data set. However, DL models are considered more flexible and use a layer of neurons to form an output prediction, yet they require more training. Additionally, ML models use image features, mammograms, biopsies, and more to detect cancer and guide treatment. DL models also process more unstructured data types, support doctors with diagnostic procedures, and can predict patient survival rates. However, as established previously, AI-based models are still developing, and their methods are still not fool-proof. While their accuracy is improving, further development is needed to make their predictions exact. Fortunately, since more data is being collected and becoming more accessible, the models' skills and precision will inevitably improve.

The reason ML/DL models are so vital, especially in oncology, is because cancer affects lives globally, no matter the gender or age. It is such a prevalent problem which requires a demanding process to heal and is very taxing on the body. Plus, many do not even have the access to therapy to fight the disease, as it is costly to pay for multiple treatments, as it is unlikely only one will work effectively. This is why AI is so important for the future of oncology, as it can make the cancer process much more orderly and efficient. Doctors can see more patients at a time and can cure more people. Furthermore, with higher efficiency, the cost of treating cancer will likely become cheaper, and therefore more accessible to all walks of life.

In the future, ML and DL models will continue to develop and gain even more abilities to help assist doctors. It is also possible that they will be able to make reliable clinical predictions with no further oversight, making the process even faster. Moreover, since ML/DL models already catch miniscule details that clinicians overlook, it is possible that in the future they could find new patterns or anomalies that remain undetected today. Further research should focus on applying the models to new and diverse data sets to expand and challenge their pre-existing abilities. Doing so would provide a better idea on where AI-based models can improve their calculations. Using international patient profiles will also help ensure that these models can perform reliably across populations. In conclusion, AI's applications in health care are rapidly expanding, and its potential to save numerous lives and progress accessibility will continue to grow.

References

Ahmad, I., & Fahad Alqurashi. (2024). Early cancer detection using deep learning and medical imaging: A survey. *Critical Reviews in Oncology/Hematology*, *204*(104528), 104528–104528. https://doi.org/10.1016/j.critrevonc.2024.104528

Bera, K., Braman, N., Gupta, A., Velcheti, V., & Madabhushi, A. (2021). Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nature Reviews Clinical Oncology*, 1–15. https://doi.org/10.1038/s41571-021-00560-7

Bulten, W., Pinckaers, H., van Boven, H., Vink, R., de Bel, T., van Ginneken, B., van der Laak, J., Hulsbergen-van de Kaa, C., & Litjens, G. (2020). Automated deep-learning system for gleason grading of prostate cancer using biopsies: A diagnostic study. *The Lancet Oncology*, *21*(2), 233–241. https://doi.org/10.1016/s1470-2045(19)30739-9

Bychkov, D., Linder, N., Turkki, R., Nordling, S., Kovanen, P. E., Verrill, C., Walliander, M., Lundin, M., Haglund, C., & Lundin, J. (2018). Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-21758-3

Courtiol, P., Maussion, C., Moarii, M., Pronier, E., Pilcer, S., Sefta, M., Manceron, P., Toldo, S., Zaslavskiy, M., Le Stang, N., Girard, N., Elemento, O., Nicholson, A. G., Blay, J.-Y., Galateau-Sallé, F., Wainrib, G., & Clozel, T. (2019). Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nature Medicine*, *25*(10), 1519–1525. https://doi.org/10.1038/s41591-019-0583-3

Esteva, A., Feng, J., van der Wal, D., Huang, S.-C., Simko, J. P., DeVries, S., Chen, E., Schaeffer, E. M., Morgan, T. M., Sun, Y., Ghorbani, A., Naik, N., Nathawani, D., Socher, R., Michalski, J. M., Roach, M., Pisansky, T. M., Monson, J. M., Naz, F., … Mohamad, O. (2022). Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *Npj Digital Medicine*, *5*(1), 1–8. https://doi.org/10.1038/s41746-022-00613-w

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118. https://doi.org/10.1038/nature21056

Hausman, D. M. (2019). What is cancer? *Perspectives in Biology and Medicine*, *62*(4), 778–784. https://doi.org/10.1353/pbm.2019.0046

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500–510. https://doi.org/10.1038/s41568-018-0016-5

Kim, H.-E., Kim, H. H., Han, B.-K., Kim, K. H., Han, K., Nam, H., Lee, E. H., & Kim, E.-K. (2020). Changes in cancer detection and false-positive recall in mammography using artificial intelligence: A retrospective, multireader study. *The Lancet Digital Health*, *2*(3), e138–e148. https://doi.org/10.1016/S2589-7500(20)30003-0

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*(13), 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

Kwong, G. A., Ghosh, S., Gamboa, L., Patriotis, C., Srivastava, S., & Bhatia, S. N. (2021). Synthetic biomarkers: A twenty-first century path to early cancer detection. *Nature Reviews Cancer*, *21*(10), 655–668. https://doi.org/10.1038/s41568-021-00389-3

McIntosh, C., Conroy, L., Tjong, M. C., Craig, T., Bayley, A., Catton, C., Gospodarowicz, M., Helou, J., Isfahanian, N., Kong, V., Lam, T., Raman, S., Warde, P., Chung, P., Berlin, A., & Purdie, T. G. (2021). Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nature Medicine*, *27*(6), 999–1005. https://doi.org/10.1038/s41591-021-01359-w

McKinney, S. M. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

Painuli, D., Bhardwaj, S., & köse, U. (2022). Recent advancement in cancer diagnosis using machine learning and deep learning techniques: A comprehensive review. *Computers in Biology and Medicine*, *146*, 105580. https://doi.org/10.1016/j.compbiomed.2022.105580

Swanson, K., Wu, E., Zhang, A., Alizadeh, A. A., & Zou, J. (2023). From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell*, *186*(8), S0092–8674(23)00094-6. https://doi.org/10.1016/j.cell.2023.01.035