International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Artificial Intelligence and Machine Learning in Early Diagnosis of Hematological Malignancies

Imran Hussain[1], Ekta A. Andriyas[2], Munzali Hamisu Umar[3], Arun Kumar Saxena[4*]

[1,3,4]Assistant Professor, Integral University, Lucknow
[2]Assistant Professor, Era University, Lucknow

Abstract

In this Era of Machine Learning (ML) and Artificial Intelligence (AI), there is no field left in which these two have not left their impact. Blood cancer or Leukemia is now a days one of the very common hematological disorders. It is very difficult to diagnose leukemia in its very early stages, when it is diagnosed in its later stages it is very difficult to treat it because there are very limited treatments available. So, it is now very important to improve the diagnostic tools and techniques in traditional diagnostics procedures. ML and AI methods have recently garnered a great deal of attention in the field of cancer research by making a noticeable contribution to the growth of predictive medicine and modern oncological practices. Considerable focus has been particularly directed toward hematologic malignancies because of the complexity of detecting early symptoms. Many patients with blood cancer do not get properly diagnosed until their cancer has reached an unnumbered stage with limited treatment prospects. Hence, the state-of-the-art revolves around the latest artificial intelligence applications in hematology management.

Keywords: Artificial Intelligence, Leukemia, Diagnosis, Machine Learning

Objective

This comprehensive review on leukemia provides insights into developments that are taking place in these contemporary segments i.e. ML and AI. This will also provide in-depth status of developments that are taking place in the field of diagnostic hematology with the help of AI. Ur main objective is to explore ML and AI development in diagnosing Leukemia in early Stages.

Methodology

Machine learning and Artificial Intelligence (AI) methods have gained a great deal of attention in cancer research, particularly in hematologic malignancies due to the complexity of detecting early symptoms.
Many patients with blood cancer do not receive a proper diagnosis until advanced stages, leading to the development of the latest AI applications in hematology management, enhancing predictive medicine, and modern oncological practices. We searched Machine learning (ML) and Artificial Intelligence (AI) methods that have recently garnered a great deal of attention in the field of cancer research by making a noticeable contribution to the growth of predictive medicine and modern oncological practices.Considerable focus has been particularly directed toward hematologic malignancies because of

the complexity of detecting early symptoms. Many patients with blood cancer do not get properly diagnosed until their cancer has reached an advanced stage with limited treatment prospects. Hence, the state-of-the-art revolves around the latest artificial intelligence (AI) applications in hematology management.

Result

By using the aforementioned keywords we have search results of around 360 papers among which we have selected 178 papers for the review.

Conclusion

After reviewing all the papers we have concluded that artificial intelligence and machine learning has a very deep impact on the early diagnosis of leukemia. It also helps in the predictive diagnosis of disease.

Introduction

Cancer begins when cells of the body start to develop rapidly. Cells in almost any section of the body can turn out to be cancer and can unfold to different areas of the body. This could also happen in blood and thus is known as blood cancer. So blood cancer or Leukaemia is known as a life-threatening disorder for a human being. It starts in the bone marrow and causes the formation of a massive quantity of unusual cells. By that time, gets entry into the blood cells and causes fatal disease. It occurs in both children and adults that can lead to death if left untreated. There are many types of Leukaemia. According to the World Health Organization (WHO), globally cancer is the 2nd leading disease of death and in 2018, it is responsible for about 9.6 million deaths. It is predicted that in 6 deaths, 1 death is due to cancer (1). Blood cancer also called Leukemia and Lymphoma are among the most common form of malignancies that are of the deadliest type (2). The blood contains three types of blood cells *viz.* White Blood cells also called Leucocytes, Red Blood Cells are also called Erythrocytes, and Platelets which are also known as Thrombocytes. Their primary functions are to fight infections, transport gases, and prevent blood loss by form forming platelet plugs, respectively (3). Ideally, in Leukemia, the maturation and function are abnormal of White Blood Cells (4). As WBC plays a crucial role in eradicating infection from our body, it is crucial for leucocytes to be in a healthy state quantitatively and qualitatively. However, their growth can become disordered when certain hematologic malignancies are present (5). These malignant WBCs get settled in the organs like the spleen, Liver, Kidney, and brain (6). Leukemia is the 10th most common cancer in men and 12th most common in women and constitutes 3% of the global cancer burden. Developing countries bear more than half of the global cancer burden. In India, lympho-hematopoietic malignancies constitute 9.5% of all cancers in men and 5.5% in women. As per available information from population-based surveys, the incidence of leukemia in India varies from 0.8/1, 00,000 in Barshi (Rural area of Maharashtra) to 5/1,00,000 in Delhi. These figures are comparably lower than the rest of the world but under-diagnosis and under-reporting cannot be ruled out. The cell type distribution of leukaemias observed in India is different from that observed in the developed world. Myeloid leukemias predominate in India while lymphoid leukaemias dominate in the Western world mainly because of the higher incidence of chronic lymphatic leukemia. Despite being relatively uncommon, leukemias have been studied more extensively because of the easy accessibility of involved tissue. The incidence of CML was noted highest (45.3%) and that was lowest of CLL (5.7%) in the Capital of India i.e. Delhi during the period of 1970-1979.19 Similar observations were noted in

Chandigarh and other metro cities like Mumbai and Calcutta. There was an exception for the incidence of ALL (39.2%) which was highest observed in Kerala state from 1980 to 1983 (7).

Obstacles in Haematological Management

Complete Blood Count (CBC) is routine investigation in Hematology Section of Diagnostic Laboratory, although it's have their limitations. It serves as an initial step in order to identify any type of quantitative or qualitative abnormality (8). CBC also provides the information that either someone is suffering from hematological malignancies or not (9). So, further evaluation is done by making smear examination by using appropriate staining method (8). But there is some drawback to these methods, they are obsolete and requires very highly skilled personal to investigate the blood smear. Sometimes it is also very time-consuming which creates a negative impact on patient efficient and timely treatment (10). One more difficult part of hematology identification is the presence of other blood components around WBCs. Therefore, appropriate classification results are not obtained by the existing identification procedure, which involves manually counting the amount of WBCs that appear aberrant (11). In fact, it has been noted that the difficulties in diagnosing diseases and the intricacy of symptom analysis are the primary causes of diagnostic delays.

Artificial Intelligence in the Management of Hematology

Driven by the impressive successes of artificial intelligence (AI) across a range of domains, the suitability of these algorithms in addressing crucial issues about hematology and oncology has been examined and demonstrated to be effective recently (12). In particular, Machine Learning (ML) and Artificial Intelligence (AI) techniques have been used to help classify different types of cancer, facilitate faster diagnosis, and provide the basis for accurate clinical decisions for better health outcomes. Two of the main challenges in applying AI in medicine are the limitation and limitation of health information.

Traditional Techniques to Leukemia Diagnose

Leukemia may be detected in the early stages of the illness if PB analysis reveals a high concentration of aberrant white blood cells (WBC) or even a very low blood count. Hence, additional morphologic and cytogenetic analysis of BM aspirates is done to confirm the AML diagnosis (13).Multiparameter flow cytometry (MFC) Immuno-phenotyping is a widely accepted method for reliably diagnosing acute myeloid leukaemia (AML) through the detection of intracellular and extracellular markers, such as CD13, CD33, and CD34. However, in order to diagnose AML illness, a traditional cytogenetic investigation is required. Fluorescence in situ hybridization is an alternative method to detect particular abnormalities, gene fusions or chromosome abnormalities associated to myelodysplasia). In addition to these technological methods, many molecular genetic testing have been employed to screen for all genetic abnormalities indicative of AML. These studies are necessary for differential diagnosis and targeted therapeutic interventions. A BM biopsy or BM trephine biopsy is always required for the ultimate diagnosis and surveillance of AML. These are intrusive, unpleasant, and cause leukaemia sufferers agony and suffering (14).As a result, fewer invading techniques are required to guarantee an accurate diagnosis and provide more specific details regarding the course of the illness. Liquid biopsy, or PB biopsy, is one of the less intrusive techniques that has been thoroughly researched and used as a supplemental technique (15).

Artificial Intelligence and Machine Learning in Diagnosis of Leukemia

Models Based on Machine Learning, about 20% of children who receive full treatment for childhood ALL, a malignant disease that is the primary cause of cancer-related deaths in children, experience a recurrence (16). Hence, it is essential to anticipate relapse in order to address the various risk groups appropriately. An ALL relapse prediction model based on ML algorithms was presented by Pan et al. to help classify patients with ALL into suitable risk categories, hence facilitating improved care and follow-up planning. To separate relapses from non-relapses in the three clinically defined risk categories—standard, intermediate, and high risk—four classification algorithms—random forest (RF), decision tree, support vector machine (SVM), and linear regression—were trained using 103 clinical variables during the model selection process (16).Hauser et al. (17) investigated the prospect of predicting CML prior to diagnosis using simply CBC test data and ML algorithms such as XGBoost and LASSO algorithms on 1623 patients with a definitive CML status, whereas Pan et al. (16) developed a model to predict illness relapse. The study included laboratory CBC test results, patient demographic data (age and gender), and patient encounter data (number of visits to outpatient clinics, etc.) as factors. The most promising predictors' predictive ability was evaluated using a forward feature selection procedure. The data set was then split into seven subsets, each corresponding to a different time point between the patient's diagnosis and the remaining six sets times in front of the diagnostic examination. It's interesting to note that depending on the time between data collection, variable selection produced distinct features for the models.In each of the 100 training sets, a 10-fold cross validation was used to assess the performance of the selected classifiers (16). The recommended method, meanwhile, is thought to be insufficient for internal validation, which calls for at least 50 repetitions (18). In contrast, the selected data set (17) was split into two separate groups: the test group and the train/validation group. Given the enormous sample size, the latter split-sample validation approach is fair and justified to apply in this scenario; nonetheless, there are still a number of aspects that require attention throughout the application, and potential negatives should be considered. For example, significant patient imbalances with regard to the predictor and output distributions may have developed since the sample split was carried out completely at random. Additionally, 20% was used to evaluate the model, which could have resulted in a biased assessment of the model's efficacy (18).Moreover, it is often recognized that leukemia patients frequently experience health problems as a result of recurrent infections, which, if not diagnosed early, can be fatal. In order to do this, Agius et al. (19) created the CLL therapy-Infection Model (CLL-TIM) to study the risk of infection resulting from a cytotoxic therapy or a compromised immune system right after a diagnosis of CLL. The target outcome was the 2-year infection risk or the course of CLL treatment, with the prediction point for each patient being established at three months following diagnosis. The ultimate number of participants in the trial group was 3729, after 74 patients who passed away and 373 patients who started treatment before the predicted point were excluded. By dividing the data set into 65% training and 17.5% test and internal validation sets, Agius et al. (19) used stratified sampling, in contrast to Hauser et al. (17), to preserve class distributions and make up for the 52% International Prognostic Index for CLL (CLL-IPI) missing variables. The CLL-TIM ensemble algorithm was made up of 28 ML models that could identify patients at a high risk of infection to increase their chances of survival. The 7288 features that were obtained from a collection of variables from various sources included baseline variables at the time of diagnosis, such as age, gender, etc.; routine laboratory tests; microbiology findings; pathology reports; and diagnosis codes for all patients.

Conclusion

After reviewing all the papers we have concluded that artificial intelligence and machine learning has a very deep impact on the early diagnosis of leukemia. It also helps in the predictive diagnosis of disease. Many research suggested that early prediction of Disease may decrease the mortality rate. In leukemia, relapses are common, although there is no such precise tool and techniques available to early diagnose the relapses. Here AI and machine learning might be a tool that can early diagnose the relapse pattern and predict it inthe near future.

References-

1. O. Gal, N. Auslander, Y. Fan, and D. Meerzaman, "Predicting complete remission of acute myeloid leukemia: Machine learning applied to gene expression," Cancer informatics, vol. 18, p. 1176935119835544, 2019.
2. Das M, Kumar Saha P. A Comprehensive Study of Machine Learning algorithms for Predicting Leukemia Based on Biomedical Data. International Conference on Innovation in Engineering and Technology; December 23-24, 2019; Dhaka, Bangladesh. 2020. Dec 21, pp. 1–5.
3. Sahlol AT, Kollmannsberger P, Ewees AA. Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features. Sci Rep. 2020 Feb 13;10(1):2536. doi: 10.1038/s41598-020-59215-9. doi: 10.1038/s41598-020-59215-9.10.1038/s41598-020-59215-9.
4. Acharya V, Ravi V, Pham TD, Chakraborty C. Peripheral Blood Smear Analysis Using Automated Computer-Aided Diagnosis System to Identify Acute Myeloid Leukemia. IEEE Trans. Eng. Manage. 2021:1–14. doi: 10.1109/tem.2021.3103549.
5. Ratley A, Minj J, Patre P. Leukemia Disease Detection and Classification Using Machine Learning Approaches: A Review. 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T); January 3-5, 2020; Raipur, Chhattisgarh, India. 2020. pp. 161–165.
6. Renuka T, Surekha B. Acute-lymphoblastic leukemia detection through deep transfer learning approach of Neural Network. First Doctoral Symposium on Natural Computing Research; August 8, 2020; Pune, Maharashtra, India. 2021. Mar 19, pp. 163–170.
7. Rajnikant Ahirwar, Rajendra Kumar Nigam, Deepankar Parmar. A study of leukemias Profile in central India: Tropical Journal of Pathology and Microbiology: April - June, 2018/ Vol 4/ Issue 2 : DOI: https://doi.org/10.17511/jopm.2018.i02.12
8. Ratley A, Minj J, Patre P. Leukemia Disease Detection and Classification Using Machine Learning Approaches: A Review. 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T); January 3-5, 2020; Raipur, Chhattisgarh, India. 2020. pp. 161–165.
9. Lee S, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, Estey EH, Miller CP, Chien S, Dai J, Saxena A, Blau CA, Becker PS. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat Commun. 2018 Jan 03;9(1):42. doi: 10.1038/s41467-017-02465-5. doi: 10.1038/s41467-017-02465-5.10.1038/s41467-017-02465-5
10. Avalos M, Touchais H, Henríquez-Henríquez M. Optimising criteria for Manual Smear Review following Automated Blood Count Analysis: A machine learning approach. Advances in Intelligent Systems and Computing. 2021 Apr 10;91:380–91. doi: 10.1007/978-3-030-73603-3_35.
11. Darane B, Rajput P, Sondagar Y, Koshy R. Recognizing Presence of Hematological Disease using Deep Learning. 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics

and Cloud) (I-SMAC); December 12-14, 2019; Palladam, Tamil Nadu, India. 2020. Mar 12, pp. 310–315.

12. Schenone D, Dominietto A, Campi C, Frassoni F, Cea M, Aquino S, Angelucci E, Rossi F, Torri L, Bignotti B, Tagliafico AS, Piana M. Radiomics and Artificial Intelligence for Outcome Prediction in Multiple Myeloma Patients Undergoing Autologous Transplantation: A Feasibility Study with CT Data. Diagnostics (Basel) 2021 Sep 24;11(10):1759. doi: 10.3390/diagnostics11101759.
13. Orazi A. Histopathology in the Diagnosis and Classification of Acute Myeloid Leukemia, Myelodysplastic Syndromes, and Myelodysplastic/Myeloproliferative Diseases. Pathobiology. 2007;74:97–114. doi: 10.1159/000101709.
14. Khoo B.L., Shang M., Ng C.H., Lim C.T., Chng W.J., Han J. Liquid biopsy for minimal residual disease detection in leukemia using a portable blast cell biochip. Npj Precis. Oncol. 2019;3:30. doi: 10.1038/s41698-019-0102-5.
15. Abdulmawjood B., Roma-Rodrigues C., Fernandes A.R., Baptista P.V. Liquid biopsies in myeloid malignancies. Cancer Drug Resist. 2019;2:1044–1061. doi: 10.20517/cdr.2019.88.
16. Pan L, Liu G, Lin F, Zhong S, Xia H, Sun X, Liang H. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. Sci Rep. 2017 Aug 07;7(1):7402. doi: 10.1038/s41598-017-07408-0. doi: 10.1038/s41598-017-07408-0.10.1038/s41598-017-07408-0
17. Hauser R, Esserman D, Beste L, Ong S, Colomb D, Bhargava A, Wadia Roxanne, Rose Michal G. A Machine Learning Model to Successfully Predict Future Diagnosis of Chronic Myelogenous Leukemia With Retrospective Electronic Health Records Data. Am J Clin Pathol. 2021 Nov 08;156(6):1142–1148. doi: 10.1093/ajcp/aqab086.6310949.
18. Steyerberg EW. Statistics for Biology and Health. Berlin, Germany: Springer; 2020. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating.
19. Agius R, Brieghel C, Andersen MA, Pearson AT, Ledergerber B, Cozzi-Lepri A, Louzoun Y, Andersen CL, Bergstedt J, von Stemann JH, Jørgensen Mette, Tang ME, Fontes M, Bahlo J, Herling CD, Hallek M, Lundgren J, MacPherson CR, Larsen J, Niemann CU. Machine learning can identify newly diagnosed patients with CLL at high risk of infection. Nat Commun. 2020 Jan 17;11(1):363. doi: 10.1038/s41467-019-14225-8. doi: 10.1038/s41467-019-14225-8.10.1038/s41467-019-14225-8.