OPEN ACCESS

Volume: 11

Special Issue: 1

Month: July

Year: 2023

E-ISSN: 2582-0397

P-ISSN: 2321-788X

Impact Factor: 3.025

Received: 18.05.2023

Accepted: 23.06.2023

Published: 01.07.2023

Citation:
Kumbhar, Shreedhar Maruti, and S. Manoj. "Explainable Symptom-Based Artificial A Template for Thinking for Leukaemia Detection." *Shanlax International Journal of Arts, Science and Humanities*, vol. 11, no. S1, 2023, pp. 122–28.

DOI:
https://doi.org/10.34293/sijash.v11iS1-July.6326

Explainable Symptom-Based Artificial A template for thinking for Leukaemia Detection

Shreedhar Maruti Kumbhar
Department of Master of Computer Applications
Raja Rajeswari College of Engineering, Bangalore

Manoj S
Department of Master of Computer Applications
Raja Rajeswari College of Engineering, Bangalore

Abstract

Leukaemianevertheless has an untimely end, but it can be extremely costly to cure. Nonetheless, early detection of leukaemia may save lives as well as funds for those affected, particularly children, for whom leukaemia is a common disease type. In this study, we present a directed An AI algorithm that successfully forecasts the possibility of the initial stages leukaemia based solely on side effects. Furthermore, include selection is carried outbased on facts set to demonstrate the power of individual elements and work on the exhibition of characterisation models. We use two AI calculations, the nave bayes classifier and the support vector network. Cross-validation is a strategy for testing ML models that involves training numerous ML models on subsets of the available input data and then evaluating them on the complementary subset. To detect overfitting, use cross-validation. The SVM algorithm based on performance achieved the highest accuracy, precision, recall, and F-measure.

Keywords: Explainable AI, Leukemia, Machine Learning, Symptom-Based Detection.

Introduction

Consequently, the National Cancer Institute, around 60,300 new patients were hospitalised to hospitals in US owing to leukaemia in 2019, with 24,370 of them dying. That has evolved into problem in the past few years. major source of concern. Despite enormous scientific efforts to combat leukaemia and its various forms around the world, the death rate from leukaemia is frightening, with devastating effects, particularly in youngsters. Leukaemia is cell are abnormal cells in the blood. are underdeveloped and exhibit aberrant behaviour such as uncontrollable growth and division.It is very widely used. kind of cancer found in youngsters. Leukaemia is categorised divided into two sections based on the cell type: lymphocyticlymphoblastic and non-lymphocytic (or myeloid). They might be chronic or acute in nature.

Signs and symptoms of leukaemia occur in the body. However, it is frequently identified at a later stage, making therapy more difficult. Early detection of leukaemia can make a significant impact in terms of cost and mortality rate, and moreover patient quality of life. In general, leukaemia detection and screening are completed in hospitals utilising a number of advanced procedures. They use blood samples, full blood counts, bone marrow testing, and so forth. The root or beginning point of leukaemia is bone marrow, where lymphocytes and myeloid cells begin to form.Various studies have indicated that imaging blood cells can also assist detect leukaemia.

The Acute Lymphoblastic Leukaemia Image A record that is extensively discussed Poetry has an extremely strong popular dataset for leukaemia identification. These image-based systems are frequently vulnerable to sophisticated gadgets and imaging techniques. In recent years, genomic approaches and clinical data They were employed. Though combinations of different approaches and multi-modal data are useful, they could be available at a later date. Various machine learning-based methodologies and Algorithmic are currently being used in literature to detect and predict leukaemia.With the growth in available data, it is now possible to describe the problem as a unsupervised machine that is a learning issue can be solved using knowledge-based techniques. Random Forest, support vector machines. Decision Trees, Neural Networks, k-Nearest Neighbour, Fuzzy Systems, Ensemble Methods, and more techniques have been effectively deployed.Often, these machine learning acting do not produce explainable results.

Explainable intelligent technology frequently aids business logic by extracting domain knowledge and norms.

The basic objective of our project is to develop a model that can detect leukaemia or normal blood.

- To use machine learning and cross validation techniques.
- Using machine learning When combined with a gullible Bayes classifier and a support vectors machine, our suggested SVM provides excellent accuracy with predicted outcomes. To improve overall performance.

Literature Review

SubrajeetMohapatra[1] is an Indian businessman. Acute lymphoblastic leukaemia is a type of childhood haematological neoplasia characterised by a high number of lymphoid blasts in the bloodstream. ALL accounts for around 80% of childhood leukaemia and is most common in children aged 3 to 7. The vague character of ALL's indications and symptoms frequently leads to incorrect diagnosis. Diagnostic confusion is often caused by other illnesses imitating similar symptoms. Only a careful microscopic inspection of stained blood smears or bone marrow aspirates allows for an accurate diagnosis of leukaemia. Fluorescence in situ hybridization (FISH), immunophenotyping, cytogenetic analysis, and cytochemistry are also applied to find particular leukaemias. Considering the aforementioned specifics, there is a requirement for leukaemia detection automation.

Mohammad Akter's[2] This study goes over strategies for detecting leukaemia. Red blood cells and immature white cells aredetected utilisingan array of techniques image processing approaches. Anaemia, leukaemia, malaria, vitamin B12 deficiency, and other diseases can be known therein manner. Finding is the aim. and count leukemia-affected cells. Leukaemia can be diagnosed and classified as chronic or acute basedin terms of recognition of immature blast cells. A mixture of methods are used to detect immature cells, including histogram equalisation, linear contrast stretching, and morphological techniques such as area opening, area closure, erosion, and dilation. Watershed transform, K means, histogram equalisation and linear contrast stretching, and shape-based features are 72.2%, 72%, 73.7%, and 97.8% correct, respectively.

Mohammad Akter's[3] This study goes over strategies for detecting leukaemia. Red blood cells and immature white cells are identified using a variety of image processing approaches. Anaemia, leukaemia, malaria, vitamin B12 deficiency, and other diseases can be identified in this manner.

The aim is to identify and count leukemia-affected cells. Leukaemia can be diagnosed and classified as chronic or acute according to the identification of immature blast cells. A variety of methods accustomed to detect immature cells, including histogram equalisation, linear contrast stretching, and morphological techniques such as area opening, area closure, erosion, and dilation. Watershed transform, K means, histogram equalisation and linear contrast stretching, and shape-based features are 72.2%, 72%, 73.7%, and 97.8% correct, respective

Existing Model

Leukaemia is not only devastating, butIn addition, exceedingly expensive to treat. Regardless, early detection of leukaemia can save lives in addition funds for those affected, particularly children, for whom leukaemia is a prevalent type. This study examines provide a realistic administered AI model that accurately forecasts the risk of early-stage leukaemia based solely on adverse reactions. The suggested architecture was made utilising data from two massive medical clinics in Bangladesh. A review of leukaemia and people without leukaemia directed in partnership with a particular clinician assembles 16 dataset features. Our logical managementbasing design upon a decision trees classification algorithm, which produces far better outcomes than other computations and transmits ready-to-use sensible instructions.

We used an apriorial calculation to develop logical leukaemia expectations. Furthermore, spotlight inquiry and element determination are applied to the dataset to demonstrate the force of specific components and their impact on ordering model display. A few of the models are tested within the information set to demonstrate how the suggested approach, which is basic but effective, outperforms the majority of the different models that we have applied. The selected tree model provided During our research achieved 97.45% precision, 0.63 Mathew's Connection Coefficient (MCC), and 0.783 region under Beneficiary Working Trademark (ROC) benton the examt data set. We additionally developed the dataset and source code for the approaches to create this study available to the scholars for further study.

Proposed Methodology

Simple Bayesian Learner

- The Nave The Bayes algorithm is a supervised learning method for resolving issues with classification using the Bayes theorem.
- A widely utilised in text classification having a large training dataset.
- The Nave Bayes A classification model is a quick and efficient segmentation technique for creating machine learning models that can make decisions. quick predictions.
- Being a probabilistic classifier, it makes predictions based on the likelihood of a particular event..
- Filtering junk mail, analysing emotions and article classification are some popular applications of the Nave Bayes Algorithm.

Support Vector Technology

- For Classification and Regression applications, Support Vector Machine (SVM) is a popular Supervised Learning technique. However, it is mostly used in Machine Learning to solve classification problems.
- The SVM algorithm's purpose is to find the optimal line or decision boundary for categorising n-dimensional space so that we may simply place fresh data points in the proper category in the future. A hyperplane is the optimal choice boundary.

- SVM chooses the extreme vectors and points that help build the hyperplane. Support vectors are used in this technique, which is also known as the Support Vector Machine. Take a look at the image below, which displays two separate groupings. that are categorised using a decision boundary.

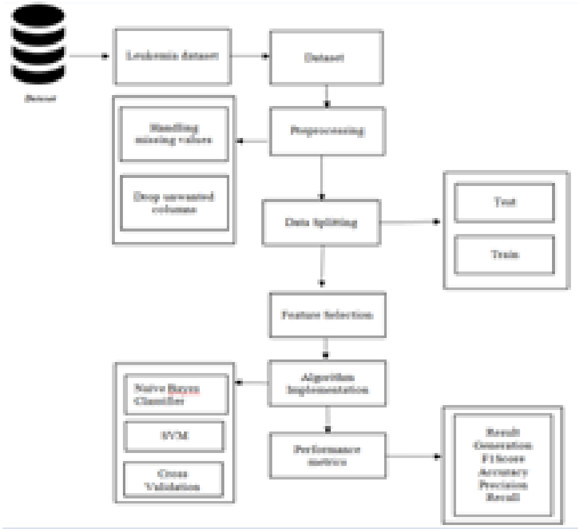

Figure 1 Proposed Architecture

Implementation

Module Description

- Data selection
- Data preprocessing
- Data splitting
- Feature Selection

Data Collection

- The data being used was acquired using the web for a website called dataset repository.
- This work includes a test dataset and a train dataset, with the trial information having a 5000 dataset and the train dataset having an 8000 dataset.
  This technique read from our obtained dataset using pandas.

Data Preprocessing

- Data preparation is the approach of deleting undesirable data a dataset, please. • Pre-processing Techniques for data transformation are employed. to turn the dataset into a machine learning-friendly structure.
- Washing is also part of this procedure. the dataset by deleting extraneous or corrupted data that could impair the dataset's correctness,enhancing it efficient.
- Elimination of missing data

Data Splitting

- Data are required throughout the machine learning process in order for learning to occur.
- Along with to generating training facts, testing data are necessary to assess the algorithm's efficiency; however, training and testing datasets are provided independently below.

- We must divide practise and assessment into x_train, y_train, x_test, y_test in our method.
- Data The technique of dividing the data accessible into twohalves is known as splits. typically for cross-validation purposes.
- A single set of facts is employed to build a predictive model, while the other is utilised to evaluate the system's efficiency.

Feature Selection

- Inclusion choices, also known as credit determination, is the most often used method for recognising the most important features from an information set and then utilising AI tactics for use on the algorithm's presentation.
- A large number of irrelevant features exponentially increases training time and raises the risk of overfitting.
- For categorical features in a dataset, the Chi-square test is utilised. We compute the Chi-square between each feature and the target and select the features with the highest Chi-square scores.

Result

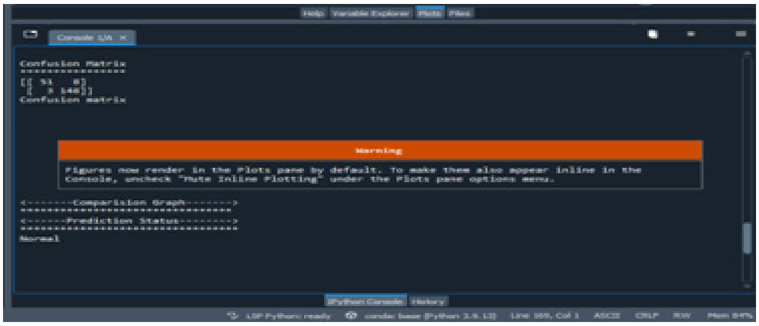

Figure 2 Prediction of Status

Graphs

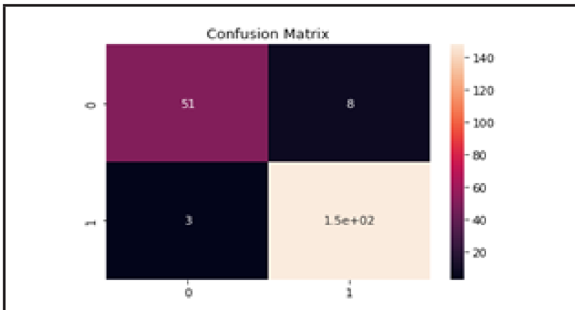

Figure 3 Confusion Matrix

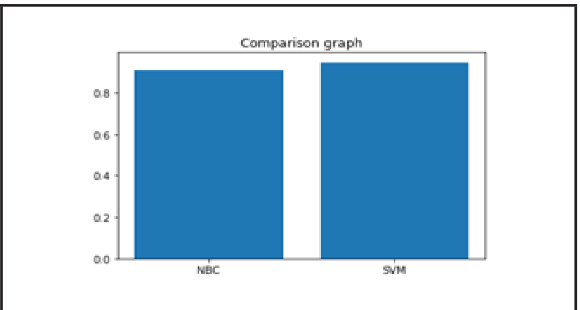

Figure 4 Comparison Graph

Fig 2, 3 and 4 shows the status of the prediction using dataset, the graphs shows the leukaemia dataset prediction. This predicts the level of leukaemia disease and gives the result.

Conclusions

We established an understandable machine learning model was used in this work to diagnose leukaemia.We employed a Leukaemia dataset. However, after selecting the data, we used multiple ML models and presented a comparison analysis. In addition, we used explainable analysis to develop a classification and cross validation approach. Our suggested SVM method provides high accuracy and precise prediction results.

In the future, further information based on datasets will be discovered,like the kind of leukaemia detected. To employ a different approach to extract characteristics. In addition, we intend to employ more dataset to validate the efficiency and efficacy of our strategies.

References

1. (2019).MedicalNewsToday.Accessed:Nov.2,2020.[Online].Available:https://www.medicalnewstoday.com/articles/142595
2. (2019).MedicalNewsToday.Accessed:Nov.2,2020.[Online].Available:https://www.medicalnewstoday.com/articles/282929
3. U. Zelig, S. Mordechai, G. Shubinsky, R. K. Sahu, M. Huleihel, E. Leibovitz, I. Nathan, and J. Kapelushnik, ''Pre-screening and followup of childhood acute leukemia using biochemical infrared analysis of peripheral blood mononuclear cells,'' Biochim. et Biophys. Acta (BBA) Gen. Subjects, vol. 1810, no. 9, pp. 827–835, Sep. 2011.
4. K. A. S. A. Daqqa, A. Y. A. Maghari, and W. F. M. A. Sarraj, ''Prediction and diagnosis of leukemia using classification algorithms,'' in Proc. 8th Int. Conf. Inf. Technol. (ICIT), May 2017, pp. 638–643.
5. N.Mahmood, S.Shahid, T.Bakhshi, S.Riaz, H.Ghufran, and M.Yaqoob, ''Identification of significant risks in pediatric Machine learning (ML) for the diagnosis of acute lymphoblastic leukaemia (ALL)approach,'' Med. Biol. Eng. Comput., vol. 58, no. 11, pp. 2631–2640, Nov. 2020.
6. E. Fathi, M. J. Rezaee, R. Tavakkoli-Moghaddam, A. Alizadeh, and A. Montazer, ''Design of anintegrated model for diagnosis and classification of pediatricacuteleukemia using machine learning,''Proc.Inst.Mech. Eng., H, J. Eng. Med., vol. 234, no. 10, pp. 1051–1069, Oct. 2020.
7. T. Markiewicz, S. Osowski, B. Marianska, and L. Moszczynski, ''Automatic recognition of the blood cells of myelogenous leukemia using SVM,'' in Proc. IEEE Int. Joint Conf. Neural Netw., vol. 4, Jul. 2005, pp. 2496–2501.
8. S.-H. Hsieh, Z. Wang, P.-H. Cheng, I.-S. Lee, S.-L. Hsieh, and F. Lai, ''Leukemia cancer classification based on support vector machine,'' in Proc. 8th IEEE Int. Conf. Ind. Informat., Jul. 2010, pp. 819–824.
9. M.Reiter,M.Diem,A.Schumich,M.Maurer-Granofszky,L.Karawajew, J. G. Rossi, R. Ratei, S. Groeneveld-Krentz, E. O. Sajaroff, S. Suhendra, M. Kampel, and M. N. Dworzak, ''Automated flow cytometric MRD assessmentin childhooda cute B-lymphoblasticleu kemia using supervised machine learning,'' Cytometry A, vol. 95, no. 9, pp. 966–975, Sep. 2019, doi: 10.1002/cyto.a.23852.
10. E. B. Leinoe, M. H. Hoffmann, E. Kjaersgaard, J. D. Nielsen, O. J. Bergmann, T. W. Klausen, and H. E. Johnsen, ''Prediction of haemorrhage in theacute myeloid stage in its infancy leukaemia by flow cytometric analysis of platelet function,'' Brit. J. Haematol., vol. 128, no. 4, pp. 526–532, Feb. 2005.
11. S. Shafique and S. Tehsin, ''Acute lymphoblastic leukemia detection and classification of employing pre-trained deep convolutional neural networks, its subtypes networks,'' Technol. Cancer Res. Treatment, vol. 17, Sep. 2018, Art. no. 1533033818802789.
12. ''Infection level identification for leukaemia detection using optimised support vector neural network,'' B. K. Das and H. S. Dutta, Imag. Sci. J., 67(8), 417-433, November 2019. [13] M. Fatma and J. Sharma, ''Identification and classification neural network,using for acute leukaemia '' inProc. Int. Conf. Med. Imag., m-Health Emerg. Commun. Syst. (MedCom), Nov. 2014, pp. 142–145.

13. J.. Rawatt, A. singh, B. Hs, J. Virmani, and J. S. Devgun, ''Computer assisted classification framework for prediction of acutelymphoblastic and acute myeloblastic leukemia,'' Biocybern. Biomed. Eng., vol. 37, no. 4, pp. 637–654, 2017.
14. K. K. Jha, P. Das, and H. SDutta, ''FAB classification based leukemia identification and prediction using machine learning,'' in Proc. Int. Conf. Syst., Comput., Autom. Netw. (ICSCAN), Jul. 2020, pp. 1–6.