STUDY AND COMPARITIVE ANALYSIS OF DIFFERENT MACHINE LEARNING ALGORITHMS

Team Members:

17BCB0018 (Niwedita Kumari)

17BCE0163 (Titichh Mishra)

17BCE2367 (Roshni Sah)

Report submitted for the

Final Project Review of

Course Code: CSE3013 – AI

Slot: F2 + TF2

Professor: Prof. Harshita Patel

1. Introduction:

Bioinformatics is a field of study that uses computation to extract knowledge from a wide range of biological data. It ensures collection, storage, retrieval, manipulation and modelling of data for analysis, prediction or visualization through the use of algorithms. In recent years, the size and number of available biological datasets have increased hugely that has led the bioinformatics researchers to make use of various machine learning algorithms. Today, the machine learning techniques such as deep learning is being deployed for the automatic feature learning using the datasets for many biological research works. With respect to this area of Science, in this project, we are dealing with a fatal disease of white blood cells (WBC) which affects the blood and bone marrow in the human body called Leukemia. We have different Genetic variants and Gene expressions with its gene value associated with the white blood cell which determines the type of Leukemia. Leukaemia is of 2 types: acute and chronic leukaemia.

Leukaemia is a malignant disease characterized by the uncontrolled accumulation of abnormal white blood cells. Leukaemia is a fatal cancer. All blood cells arise in the bone marrow, which occupies the central cavity of the bone, via hematopoiesis. Acute leukaemia causes the fast deterioration of the patient, whereas chronic leukaemia is characterized by slow progression and maybe lymphocytic or myelogenous.

Currently, leukaemia has been classified using two systems: the French-American-British (FAB) classification and the World Health Organization (WHO) proposal [2]. Blast cells found in the peripheral blood smear and characterized Acute Myeloid Leukemia (AMLs), which includes seven types (M-1–M-7). This process is very tedious, time-consuming and not suitable for analyzing a large number of cells. Nevertheless, some mathematical approaches and technologies have been developed to discriminate blood cells, and it is essential for identifying leukaemia [3]These researchers surmised a subjective discovery model of the mapping standard, and from the earlier master learning cannot be considered to better ordaining the parameters, keeping in mind that the end goal is to enhance meeting and decrease the learning time..we also explored other well-known machine learning algorithms such as naive Bayes, support vector machine, k-nearest neighbour, and decision tree.

2. Objective of the project:

To make a comparative analysis of different ML algorithms and obtain the algorithm that can give the highest accuracy for the dataset used.

3. Innovation component in the project:

We made a comparative analysis of different machine learning algorithms to obtain the algorithm which can give the highest accuracy for a dataset which henceforth will help to detect if a patient is suffering from acute leukaemia or chronic leukaemia. In this project, the algorithms implemented are k-Means, Naive Bayes, SVM, logistic regression and XG-Boost.

4. Work done and implementation

Methodology adapted:

We implemented the different algorithms on the dataset using the Jupyter notebook platform on Windows OS. On importing the libraries, we loaded the dataset. We trained the dataset.

We then implemented various algorithms of machine learning for the dataset, ie,k-Means, Naive Bayes, SVM, logistic regression and XG-Boost. On doing so we obtained different accuracy value for each algorithm. We also worked to obtain the confusion matrix for each of the algorithm used. The confusion matrix gives a graphical approach to our prediction nature. Also, it will help to detect which algorithm will prove to be the best in terms of accuracy.

This methodology can be implemented to get the algorithm that will best suit to detect the type of Leukemia the patient is suffering from and hence the cure can be done at the earliest. without going with the traditional method of blood sample collection and analysis. only the gene expression value can ease our task because then we can apply the various machine learning algorithms and get the best algorithm that can deliver the result with the best accuracy.

Dataset used:

https://github.com/titichhm/AI-Project-Dataset

Tools used:

Numpy, Pandas, Matplotlib, Seaborn,mpl_toolkits, Keras,scikit.learn, Tensor

flow, XG-Boost

5. Literature Review Summary Table

Serial no	Authors & Titles	Dataset/s	Contribution	Observation
1.	Muhammad Hammad Waseem et al. & On the Feature Selection Methods and Reject Option Classifiers for Robust Cancer Prediction	Leukemia dataset comprises of 72 patients from which 47 patients with acute lymphoblastic leukemia (ALL) and 25 patients with acute myeloid leukemia (AML).	The classifiers implemented are linear discriminant analysis (LDA), support vector machine (SVM), and k-nearest neighbors (kNN).	using LDA, an accuracy of 67%, is obtained. SVM gives 89% and knn gives 86%
2.	Lokman Faivdullah, Farid Azahar, Zaw Zaw Htike, and Wei Yan Nyein Naing. & Leukemia Detection from Blood Smears	100 blood images corresponding to each of the four leukemia classes (acute lymphoblastic leukemia, acute myeloid leukemia, chronic lymphocytic leukemia, and chronic myeloid leukemia).	Support vector machine classifier (SVM).	An overall accuracy of 79.38% using SVM in classifying leukemia from blood images.
3.	Rana Zeeshan Haider, Ikram Uddin Ujjan and Tahir S. Shamsi. & Cell	A total of 1067 patients with 44 APML (PML-RARA), 181 AML (excluded APML), 89 chronic myeloid leukemia (CML), 51 myelodysplastic	Artificial neural network (ANN) with Principal Component Analysis (PCA).	An accuracy of 95.7% and 97.7% for training and testing data sets, respectively from ANN.

	Population DataeDriven Acute Promyelocytic Leukemia Flagging Through Artificial Neural Network Predictive Modeling	syndrome (MDS), 71 myeloproliferative disorders (MPN) except CML, 10 MDS/MPN, 136 acute lymphocytic leukemia (ALL), 9 Hodgkin's lymphoma (HL), 95 non-Hodgkin's lymphoma (NHL), 32 multiple myeloma, and 349 normal control were prospectively enrolled at the National Institute of Blood Disease and Bone Marrow Transplantation (NIBD & BMT) Karachi, Pakistan.		
4.	Liyan Pan et al. & Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia	A dataset of 661 children aged <16 years, who were newly diagnosed with ALL.	Random Forest (RF) and Decision Tree (DT).	RF exhibited better prediction than DT in 4 of 6 measurements with accuracy as 0.831, specificity as 0.895, PPV as 0.880 and AUC as 0.902.
5.	Jyoti Rawat, Annapurna Singh, H S Bhadauria,Jitendra Virmani. & Computer Aided Diagnostic	In a dataset of 130 ALL infected images, first 65 images is used for training and rest for testing of the proposed system.	Auto support vector machine (SVM) binary classifier with Gray level co–occurrence matrices (GLCM).	Classification accuracies of 86.7% and 72.4% for cytoplasm and nucleus respectively for texture based feature while classification accuracies of 56.1% and 72.4% respectively for shape based features was

	System for Detection of Leukemia using Microscopic Images			obtained but classification accuracy of combined texture-shape feature is 89.8%.
6.	Ahmed S.Negm,Osama A.Hassan,Ahmed H.Kandil. & A decision support system for Acute Leukaemia classification based on digital microscopic images	The total dataset consisted of 757 images.	Panel selection, segmentation using K-means clustering to identify the leukemia cells and features extraction, and image refinement.	The algorithm testing using this dataset demonstrated an overall accuracy of 99.517%, the sensitivity of 99.348%, and specificity of 99.529%.
7.	Jakkrich Laosai,Kosin Chamnongthai. & Classification of acute leukemia using medical-knowledge-based morphology and CD marker	Experiments with 200 and 300 acute-leukemia samples for training and testing respectively.	The cluster of differentiation (CD) marker.	99.67% accuracy is obtained in this.
8.	Rana Zeeshan Haider,Ikram Uddin Ujjan, Tahir S.Shamsi. & Cell Population Data–Driven	Diagnosis of 1067 study subjects with hematological neoplasms.	Artificial neural network (ANN) predictive modeling with principal component analysis (PCA).	ANN model were found acceptable with value of 95.7% and 97.7% for training and testing data sets, respectively.

	Acute Promyelocytic Leukemia Flagging Through Artificial Neural Network Predictive Modeling			
9.	Ahmed M. Abdeldaim, Ahmed T. Sahlol,Mohamed Elhoseny,Aboul Ella Hassanien. & Computer-Aided Acute Lymphoblastic Leukemia Diagnosis System Based on Image Analysis	The dataset contains 260 cell images: 130 normal and 130 affected by ALL. (Acute Lymphoblastic Leukemia).	KNN,SVM.	KNN achieved the best classification accuracy as 93.2% while for SVM its 87.4%.
10.	Jaroonrut Prinyakupt, Charnchai Pluempitiwiriyawej. & Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers	In dataset 2 ,555 images with 601 white blood cells.In dataset 2, 477 cropped white blood cell images.	Naïve Bayes classifiers were applied for performance comparison.	The overall correction rate in the classification phase is about 94% for naïve Bayes models.
11.	Gonzalez Jesus A, et al. & Leukemia	The training set contains 56 examples and the test set contains 20 examples.	SVM,ANN,RF,k-Means.	95.5% from SVM,ANN is 79.4%,RF is 82.3% and k-Means is 82.6% to distinguish between the acute myeloblastic and

	identification from bone marrow cells images using a machine vision and data mining strategy			lymphoblastic leukemia families.
12.	J. Rodellar et al. & Image processing and machine learning in the morphological analysis of blood cells	A set of 9395 images was analyzed, obtained from Blood smears of 218 patients.	Neural networks, decision trees, and support vector machines (SVM).	The overall accuracy is the mean value SVM classifier is 88.3% while for rest of the methods its lesser than that of SVM.
13.	Oscar Picchi Netto et al. & Applying Decision Trees to Gene Expression Data from DNA Microarrays: A Leukemia Case Study	The training set contains 38 examples and the test set contains 34 examples.	Decision trees using nominal values, SVM.	94.2% accuracy of Decision Trees and for SVM it is 86.4%.
14.	Amjad Rehman et al. & Classification of acute lymphoblastic leukemia using deep learning	Diagnosis of 126 study subjects with Acute Lymphoblastic Leukemia and its subtypes.	Classifiers used as Naive Bayesian, KNN, and SVM.	97.78% accuracy using SVM and of KNN is 82.5%, while of Naive Bayesian it is of 92.3%.
15.	Enrique J.deAndrés-Galiana et al. & Analysis of	Clinical data belongs to a retrospective study of a cohort of 265 Caucasians who were	SVM and KNN.	Precision using KNN is 54.7% and of SVM is 90.1%

	Clinical Prognostic Variables for Chronic Lymphocytic Leukemia Decision-Making Problems	diagnosed with CLL between 1997 and 2007 in Hospital Cabueñes (Asturias, Spain).		
16.	Julián Candia et al. & Uncovering low-dimensional, miR-based signatures of acute myeloid and lymphoblastic leukemias with a machine-learning-driven network approach	Initially, a set of 847 human microRNAs was measured for each sample; however, here and throughout, we focus on a subset of 370 microRNAs obtained after filtering out microRNAs with consistently low or absent expression.	Support Vector Machine (SVM) with a linear kernel.	Prediction performance is 99.5% .
17.	Wanmao Ni et al. & Discrimination of Malignant Neutrophils of Chronic Myelogenous Leukemia From Normal Neutrophils by Support Vector Machine	120 samples of Malignant neutrophils of chronic myelogenous leukemia (CML).	Support vector machine method.	High specificity and sensitivity ≤95.80% and ≤95.30%, respectively.
18.	Identification of Leukemia Subtypes from	Data was collected from two sources: ALL-IDB and ASH Image bank.	Convolutional Neural Network model was used. They performed automatic feature	Experiments showed that the CNN model performance is 88.25% and 81.74% accuracy, in leukemia versus healthy

	Microscopic Images Using Convolutional Neural Network Nizar Ahmed, Altug Yigit, Zerrin Isik, Adil Alpkocak	Data augmentation was also used to increase dataset size and avoid memorization.	extraction from the images. Model was trained with 25 epochs and 32 batch size. Image transformations were used to increase the dataset size.	and multi-class classification of all subtypes, respectively. CNN model has a better performance than other well-known machine learning algorithms.
19.	Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks Sarmad Shafique, Samabia Tehsin	ALL-Image DataBase (IDB) data set was used., 50 more microscopic blood images were collected from Google which were then validated by the expert oncologist.	AlexNet, a pretrained Convolutional Neural Networks, for detection of ALL and classification of its subtypes. Transfer learning was used for deep neural network architecture.	By performing data augmentation, 99.50% accuracy for leukemia detection and 96.06% accuracy for its subtype's classification was achieved. Researchers should deploy different deep learning architectures to get the most efficient architecture for classification.
20.	Detection and Classification of Blood Cancer from Microscopic Cell Images Using SVM KNN and NN Classifier Sachin Paswan, Yogesh Kumar	100 microscopic blood cell images were acquired. Images were preprocessed using image color threshold.	Threshold technique is used for segmentation. Features considered for calculation were Hausdorff dimension, Shape features, Texture features and GLCM. KNN and modified SVM were used to do the classification.	KNN gave accuracy of 61.11%. SVM gave accuracy of 83.33%. Improvements were made to the SVM algorithm, including an initialization step to find 12-neighbor connected component

	Rathore			
21.	Leukemia Blood Cell Image Classification Using Convolutional Neural Network T. T. P. Thanh, Caleb Vununu, Sukhrob Atoev, Suk-Hwan Lee, Ki-Ryong Kwon	ALL-IDB1 image database was used.To increase the dataset size, various image transformations were used. Total 1188 images were obtained.	Convolutional neural network of 7 layers was used. First 5 layers perform feature extraction, other 2 layers classify extracted features.	30% of total images were used for testing. Accuracy of 96.60% was obtained. It is concluded that CNN can be very reliable for early detection of blood cancer.
22.	Classification of Leukemia Blood Samples Using Neural Networks Malek ADJOUADI, Melvin Mayala, Mercedes CABRERIZO, Nuannuan Zong, G LIZARRAGA, and MARK ROSSMAN	Samples were collected from various hospitals of both normal and infected cells. Data analysis was further done to extract only the cells that met criterion of interest.	Feature extraction for only 5 features was done before classification. ANN was used for classification of ALL and AML samples.	For ALL classification, 3 cycles of testing and training was conducted. For the highest amount of samples, testing accuracy was 98.46. For AML classification, only one test was conducted. It gave accuracy of 97.27%. By using a reduced amount of parameters, ANNs can be trained to classify AML or ALL.
23.	A Hybrid Deep Learning Architecture for Leukemic B-	Dataset is based on classification of normal versus malignant Cells in B-ALL White Blood Cancer	Hybrid CNN model is proposed, that combines low-level features from intermediate layers. CNN architectures	Results were derived from 967 test images. Accuracy of 96.17% was obtained by the proposed model. Obtained results suggest

	lymphoblast Classification (Sara Hosseinzadeh Kassani,Peyman Hosseinzadeh Kassani,Michal J. Wesolowski, Kevin A. Schneider,Ralph Deters)	microscopic images,provided by SBILab. Dataset contains 76 individual subjects, containing a total cells images of 7272 ALL and 3389 normal cells. Data was preprocessed to increase number of images.	MobileNet and VGG16 were used. Two output neurons associating with normal and malignant cases with softmax non-linearity activation function are used at the classifier layer.	that combining features learned by deep models improves the performance and yield more accurate result.
24.	Recognition of Acute Lymphoblastic Leukemia Cells in Microscopic Images Using K-Means Clustering and Support Vector Machine Classifier Morteza Moradi Amin, Saeed Kermani, Ardeshir Talebi, Mostafa Ghelich Oghli	21 peripheral blood smear and bone marrow slides of 14 patients with ALL and 7 normal persons are used. These were collected at Isfahan Al-Zahra and Omid hospital pathology. Preprocessing and nucleus segmentation was performed.	K-means was used for segmentation. Feature extraction was done after generation and selection. Traditional SVM was used in 1st step. Multiclass SVM classifier was used in 2nd step, because of existence of 6 classes.	K-fold cross validation method with k=10 was applied for evaluation. Binary SVM classifier gave 98%, 95%, and 97%, sensitivity, specificity, and accuracy, respectively. For multi-class SVM classifier, those values 84.3%, 97.3% and 95.6% respectively.
25.	Automated AML Detection from Complete Blood Smear Image Using KNN Classifier	American Society of Hematology online image bank was used. Colour Correlation, image segmentation was performed to improve quality	After feature extraction for some selected features, K nearest neighbours algorithm was used for classification.	Hausdorff Dimension for healthy and cancerous cells were 1.5501 and 1.7828. It was found that KNN is as better as SVM classifier in case of specificity and precision.

	Nayana B. Sen, Mercy Mathew	and get important regions of images.		
26.	Automated Detection of Acute Leukemia using K-mean Clustering Algorithm Sachin Kumar, Sumita Mishra, Pallavi Asthana, Pragya	Samples for the proposed work were obtained from Dr RML Awadh hospital, Lucknow. Preprocessing was performed suppress undesired distortions and enhance the quality.	After feature extraction for some selected features, proposed algorithm wass tested with kNN and Naïve Bayes Classifier on the dataset of 60 pretested samples.	The accuracy achieved was 92.8%. Both kNN and Naïve Bayes achieved nearly same senstivity, but specificity of Naïve Bayes classifier was very less compared to kNN.
27.	Prediction and Diagnosis of Leukemia Using Classification Algorithms Khaled A.S. Abu Daqqa, Ashraf Y.A. Maghari, Wael F. M. Al Sarraj	Dataset was collected from CBC tests repository of European Gaza Hospital.It contains 4000 instances including 2000 instances with leukemia disease. It contains 18 attributes.	Data preprocessing was done to improve accuracy. 3 classifiers, namely SVM, DT, and KNN were used. These were applied using RapidMiner to get the precision.	DT had the accuracy of 77.30%, which was highest among the 3 algorithms. It also obtained properties regarding outer attributes.
28.	A fuzzy neural approach for leukemia cancer classification Dr BBM Krishna Kanth	Dataset is a collection of expression measurements reported by Golub. Profiles have been constructed from 72 people who have either ALL or AML.	Classification dimensionality reduction methods such as Signal-to-Noise Ratio, Class-Separability, etc., were used to find best genes for classification. Fuzzy HyperSphere Neural Network Classifier was used for classification.	High accuracy was obtained by using only two genes. Average training time and testing time of FHSNN classifier with ALL/AML dataset was much faster than that of traditional methods like KNN and SVM. FHSNN yielded 100% accuracy, while traditional methods were 97.1% accurate.
29.	Classification of white	BCCD dataset, consisting of	Deep convolutional neural network to	With increasing number of epochs, the model

	blood cells for blood cancer diagnosis using deep learning neural networks Dhvani Kansara, Shaunak Sompura, Saifil Momin, Mitchell D'Silva	12500 images was used. 3000 images of each type(Eosinophil, Lymphocyte, Monocyte, and Neutrophil) were augmented. 410 images were original, pre-augmented.	classify images into cell types. Multiple layers of convolution and pooling are added in the beginning to extract all possible features. Image transformations were used to create a set of WBC images in different orientations.	obtained improved accuracies, because of back-propogation. Validation set accuracy remained near constant after 30 epochs. Value of 83% is obtained for precision. Recall and F1 score was 78%.
30.	Classifying White Blood Cells in Blood Smear Images using a Convolution al Neural Network Gulshan Sharma, Rakesh Kumar	Public dataset of 364 colored microscopic images of dyed WBC. 80% data was used as training set. 10% data was used for validation. Rest 10% was used as testset.	Convolutional neural networks of 5 2D convolutional layers are used. The maximum number of epochs set was 10.	High accuracy rates were observed. For binary classification, accuracy was 99.76%. For multi classification, accuracy was 98.14.
31.	Classifying White Blood Cells in Blood Smear Images using a Convolution al Neural Network Gulshan Sharma,	Public dataset of 364 colored microscopic images of dyed WBC. 80% data was used as training set. 10% data was used for validation. Rest 10% was used as testset.	Convolutional neural networks of 5 2D convolutional layers are used. The maximum number of epochs set was 10.	High accuracy rates were observed. For binary classification, accuracy was 99.76%. For multi classification, accuracy was 98.14.

	Rakesh Kumar			
32.	Convolutional Neural Networks for Recognition of Lymphoblast Cell Images Tatdow Pansombut, Siripen Wikaisuksakul, Kittiya Khongkraphan, Aniruth Phon-on	Two datasets were used. 1st dataset had 93 normal white blood cells, obtained were from Labati et al. Second collection is composed of ALL subtypes: pre-T and pre-B cells from ASH image bank.	A convolutional neural network, named ConVNet is used for classification. It performed automatic feature extraction. For feature extraction for SVM-GA, MLP, and Random Forest, 46 features were selected. GA based feature selection and parameters optimization was used.	The accuracies obtained for ConVNet, SVM-GA, MLP, and random forest were 81.74%, 81.65%, 76.12 % and 78.43% on average respectively. CNN was superior to MLP and random forest in all three classes.
33.	An Automated Leucocyte Classification For Leukemia Detection Gayathri.S, Jyothi R L	All dataset images are microscopic images from the laboratory through Canon Power Shot G5 camera. Segmentation and cleaning was performed after acquistion.	An ANN was implemented first. It's performance was compared with SVM's performance. Proposed work is also implemented using CNN.	Feature extraction method based recognition system produce an efficiency of 89.47% with SVM and 92.10% with ANN. CNN based feature extraction method produce an efficiency of 93%.
34.	Machine learning applications in the diagnosis of leukemia: Current trends and future directions Haneen T. Salah Ibrahim N. Muhsen Mohamed E. Salama Tarek	The search strategies used Boolean logic with MeSH terminology including terms of leukemia and its subtypes (eg, "Leukemia" and "Leukemia, Myeloid/") and terms pertaining to AHI techniques. The studies were classified according to the type of leukemia	Machine learning (ML) and deep learning algorithms like SVM, KNN, RF,LR,RC,CNN are used. The most common segmentation algorithm methodology was pattern recognition–based (eg, fuzzy c-mean and k-means), followed by threshold-based methodologies (eg, watershed).	Accuracy of algorithms used ranged from 74% to 99.5% for ALL. An accuracy of ALL detection of 74% using SVM algorithm. Accuracy of algorithms used has ranged from 82% to 97% in AML. Achieved 99.6% accuracy in flow cytometric diagnosis of CLL after using multiple algorithms, of which Bayesian clustering (BC) was the most accurate.

	Owaidah Shahrukh K. Hashmi	into: ALL (13), AML (8), CLL (3), and CML (1). Two studies proposed diagnostic models for both AML and ALL. A widely used digital library was ALL-Image DataBase (IDB), which was used in 5 studies (42%).42 ALL-IDB has two data sets, data set (1) cells are not segmented thus allowing for both segmentation and classification exercises, whereas data set (2) cells are segmented.		
35.	ACUTE LEUKEMIA CLASSIFICATION USING CONVOLUTION NEURAL NETWORK IN CLINICAL DECISION SUPPORT SYSTEM Thanh.TTP, Giao N. Pham, Jin-Hyeok Park, Kwang-Seok Moon, Suk-Hwan Lee, and Ki-Ryong Kwon.	Training set- Normal cell 40 Abnormal cell 40 Total 80 Test Set- Normal cell 19 Abnormal cell 9 Total 28	Novel approach to perform acute leukemia classification is based on Convolution Neural Network (CNN). CNN network contains 4 layers. The first 3 layers for detecting features and the other two layers (Fully connected and Softmax) are for classifying the features.	provides an excellent performance in classification process that reaches 96.43% of accuracy to discriminate normal and abnormal cell images from given database.

36.	Image processing and machine learning in the morphologic al analysis of blood cells J. Rodellar S. Alférez A. Acevedo A. Molina A. Merino	A set of 9395 images was analyzed. a set of 220 new independent images.	Image processing and segmentation techniques, Two main approaches for texture analysis are the granulometry and the gray- level co- occurrence matrix (GLCM) Neural networks, decision trees, and support vector machines (SVM) are used.	SVM classification of 5 types of normal leukocytes and only 1 group of abnormal lymphoid cells (CLL) was reported in. Although average accuracy was around 94%, CLL accuracy was 88%.
37.	Acute Lymphoblast ic Leukemia Detection and Classificatio n of Its Subtypes Using Pretrained Deep Convolution al Neural Networks Sarmad Shafique, MS1 and Samabia Tehsin, PhD	compared the data sets with different color models to check the performance over different color images. This data set was divided into 2 versions. Acute lymphoblastic leukemia-IDB 1 consisted of 108 images where 59 images were from healthy patients and 49 images were from patients affected with leukemia. Acute lymphoblastic leukemia-IDB 2 data set consisted of 260 images having single cell where 130 images were from patients affected by leukemia and 130 were normal images.	To reduce overtraining, data augmentation technique was used. SVM(support vector machine), CNN(convolutional neural network); DCCN(deep convolutional neural network)	For acute lymphoblastic leukemia detection, achieved a sensitivity of 100%, specificity of 98.11%, and accuracy of 99.50%; and for acute lymphoblastic leukemia subtype classification the sensitivity was 96.74%, specificity was 99.03%, and accuracy was 96.06%.
38.	Recognition	41 clinical and	Artificial Neural	Performance of learning

	and prediction of leukemia with Artificial Neural Network (ANN) Saeid Afshar, Fahimeh Abdolrahmani, Fereshte Vakili Tanha, Mahin Zohdi Seif, Kobra Taheri	laboratory parameters of 131 patients (63 of them were cancerous and others non-cancerous) who had pathological results were selected from patients' documents	Network (ANN), LM algorithm	was 0.094. The Relationship between the output of trained network for test data and real results of test data was high and the area under ROC curve was 0.967. Therefore can use artificial neural network for rapid and reliable leukemia recognition.
39.	White Blood Cells Segmentation and Classification to Detect Acute Leukemia Ms. Minal D. Joshi1, Prof. Atul H. Karode, Prof. S.R.Suralkar	There are two types of datasets are available. The ALL-IDB1 can be used both for testing segmentation capability of algorithms, as well as the classification systems and image preprocessing methods and ALL-IDB2 has segmented WBCs to test the classification of blast cells.	kNN classifier has been utilized to classify blast cells from normal lymphocyte cells. K-Mean clustering, fuzzy C-Mean clustering are utilised.	Leukemia detection with proposed features were classified using kNN classifier giving overall accuracy of 93%
40.	Intelligent leukaemia diagnosis with bare-bones PSO	The proposed algorithms is evaluted using a cross-domain sonar data set from the UCI Machine Learning Repository. 140 and 68 instances for training and	Bare-bones Particle Swarm Optimization (BBPSO) algorithms are proposed to identify the most significant discriminative characteristics of healthy and blast cells to enable efficient ALL classification. Both 1-Nearest Neighbour (1NN) and Support Vector	proposed algorithms achieve superior geometric mean performances of 94.94% and 96.25%,

	based feature optimization Worawut Srisukkham, Li Zhang, Siew Chin Neoh, Stephen Todryk, Chee Peng Lim	test, respectively.	Machine (SVM) with Gaussian Radial Basis Function (RBF) kernel are used to classify lymphocytes and lymphoblasts using the identified optimal feature subsets. SDM-based clustering algorithm.	
41.	Detection of leukemia in human blood sample based on microscopic images. Shailesh J. Mishra, Mrs. A.P. Deshmukh	describes a preliminary study of developing a detection of leukemia types using microscopic blood sample images.	used detection of leukemia cells in normal blood cells using MATLAB	using the MATLAB programming the different operation perform on the images like as enhancement, restoration, segmentation and color image processing using this method obtained the edge of cancerous blood cells. This cancerous blood cells in abnormal in shape and size.
42.	Automatic Recognition of Acute Myelogenou	a total number of 1500 data, 750 data for ALL, and 750 data for AML were used. Out of	Images are classified to cancerous and noncancerous images by binary	The results have shown that k-NN produced good performance in classifying both AML and ALL with high percentage of

	s Leukemia in Blood Microscopic Images Using K-means Clustering and Support Vector Machine Fatemeh Kazemi, Tooraj Abbasian Najafabadi, and Babak Nadjar Araabi	1500, 1200 was considered as the train data while the rest of the data was considered as the test data.	support vector machine (SVM) classifier, k-means clustering, fuzzy C-means clustering applied to separate the foreground and background.	accuracy up to 86%.
43.	Automated Detection of Acute Leukemia using K-mean Clustering Algorithm Sachin Kumar, Sumita Mishra, Pallavi Asthana, Pragya	the dataset of 60 samples.	using K-mean Clustering Algorithm, Image processing	The method implemented uses basic enhancement, morphology, filtering and segmenting technique to extract region of interest using k – means clustering algorithm. The proposed algorithm achieved an accuracy of 92.8% and is tested with Nearest Neighbor (kNN) and Naïve Bayes Classifier
44.	Application of Support Vector Machine and	The available data set was split into two parts: One part that contains two third of the	genetic algorithm (GA) and a support vector machine (SVM) to the recognition of	Applying the GA, were able to increase the accuracy of the blood cell recognition by more than 25% (in relative terms)

	Genetic Algorithm for Improved Blood Cell Recognition Stanislaw Osowski, Robert Siroi ´c, Tomasz Markiewicz, and Krzysztof Siwek	data set has been used in learning, and the remaining part (one third) has been used for testing only. The data from the first set (two third of the data) have also been split into two halves. The first half was used for pure learning of the SVM classifier, and the second half was used for calculating the fitness function (the validation of the SVM model). The remaining one third of data has only been used for the testing of the trained classifiers.	blood cells based on the image of the bone marrow.	with respect to the best method of feature selection (linear SVM ranking).
45.	Classificatio n of Acute Leukaemia Cells using Multilayer Perceptron and Simplified Fuzzy ARTMAP Neural Networks Aimi Abdul Nasir1, Mohd Yusoff Mashor, and Rosline Hassan	a total of 500 images (200 ALL and 300 AML) were captured from acute leukaemia blood samples by using the Leica microscope	used the Multilayer Perceptron (MLP) and Simplified Fuzzy ARTMAP (SFAM) neural networks. Levenberg-Marquardt and Bayesian Regulation algorithms have been employed to train the MLP network.	the MLP network trained by Bayesian Regulation algorithm has produced the best classification performance with testing accuracy of 95.70% for the overall proposed features.
46.	Analysis of	There are two types of datasets	support vector machine(SVM) and	Leukemia detection with proposed features were

	blood samples for counting leukemia cells using Support vector machine and nearest neighbour Niranjan Chatap, Sini Shibu	are available. The ALL-IDB1 can be used both for testing segmentation capability of algorithms, as well as the classification systems and image pre-processing methods and ALL-IDB2 has segmented WBCs to test the classification of blast cells.	nearest neighbour concept is presented	classified using kNN classifier giving overall accuracy of 93%.
47.	Patch-Based White Blood Cell Nucleus Segmentation Using Fuzzy Clustering Nipon Theera-Umpon	The data set consists of six classes of white blood cells-myeloblast, promyelocyte, myelocyte, metamyelocyte, band, and PMN. There are 20, 9, 139, 33, 45, and 185 handsegmented images for all six cell classes, respectively.	The segmentation is based on the fuzzy C-means clustering alogrithm and mathematical morphology. Bayes classifier	we achieve a good segmentation and promising classification performances compared to an expert's ground truth. Due to the gray-scale inconsistency in each region of a white blood cell image, the proposed patch-based segmentation technique makes more sense than the pixel-based segmentation techniques.
48.	White blood cell segmentation using morphological operators and scale-space analysis Leyza Baldo Dorini Rodrigo Minetto	used grayscale images from the CellAtlas.com, carried out tests on over 100 images	use simple morphological operators and explore the scale-space properties of a toggle operator to improve the segmentation accuracy	the accurate nucleus segmentation results encourage future works which include the classification of WBC using shape descriptors extracted from segmented nucleus.

	Neucimar Jerˆonimo Leite			
49.	A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia & Su-In Lee et.al	30 samples of gene expression of patients, and tested them on two independent data sets	MERGE algorithm, leave-one-out cross validation and multiple regression methods	MERGE algorithm with 83%accuracy and leave-one-out cross validation (LOOCV) with 72% accuracy and multiple regression methods with only 60% accuracy
50.	Support vector machine classification and validation of cancer tissue samples using microarray expression data & Terrence S. Furey et. al	Dataset of 72 patients with 38 samples in training set and 24 in test sample.	Use of SVM algorithm.	An accuracy percent of 70% is obtained through SVM.

6. Screenshots of implementation:

[1]:
```
import numpy as np
import pandas as pd
```

[2]:
```
import matplotlib.pyplot as plt
import seaborn as sns
from mpl_toolkits.mplot3d import Axes3D
%matplotlib inline
```

[3]:
```
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.model_selection import GridSearchCV, cross_val_score
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.cluster import KMeans
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import RandomForestClassifier
```

[4]:
```
from keras.models import Sequential
from keras.layers import Dense
from keras.callbacks import EarlyStopping
```

```
Using TensorFlow backend.
```

[5]:
```
import xgboost as xgb
```

[6]:
```
y = pd.read_csv('Dataset/actual.csv')
```

[7]:
```
y = y.replace({'ALL':0,'AML':1})
labels = ['ALL', 'AML']
```

[8]:
```
df_train = pd.read_csv('Dataset/data_set_ALL_AML_train.csv')


df_test = pd.read_csv('Dataset/data_set_ALL_AML_independent.csv')
```

[9]:
```
train_to_keep = [col for col in df_train.columns if "call" not in col]
test_to_keep = [col for col in df_test.columns if "call" not in col]

X_train_tr = df_train[train_to_keep]
X_test_tr = df_test[test_to_keep]
```

[10]:
```
train_columns_titles = ['Gene Description', 'Gene Accession Number', '1', '2',␣
↪'3', '4', '5', '6', '7', '8', '9', '10',
       '11', '12', '13', '14', '15', '16', '17', '18', '19', '20', '21', '22',␣
↪'23', '24', '25',
       '26', '27', '28', '29', '30', '31', '32', '33', '34', '35', '36', '37',␣
↪'38']

X_train_tr = X_train_tr.reindex(columns=train_columns_titles)
```

[11]:
```
test_columns_titles = ['Gene Description', 'Gene Accession Number','39', '40',␣
↪'41', '42', '43', '44', '45', '46',
       '47', '48', '49', '50', '51', '52', '53',  '54', '55', '56', '57', '58',␣
↪'59',
       '60', '61', '62', '63', '64', '65', '66', '67', '68', '69', '70', '71',␣
↪'72']

X_test_tr = X_test_tr.reindex(columns=test_columns_titles)
```

[12]:
```
X_train = X_train_tr.T
X_test = X_test_tr.T
```

[13]:
```
X_train.columns = X_train.iloc[1]
X_train = X_train.drop(["Gene Description", "Gene Accession Number"]).apply(pd.
↪to_numeric)

X_test.columns = X_test.iloc[1]
X_test = X_test.drop(["Gene Description", "Gene Accession Number"]).apply(pd.
↪to_numeric)
```

[14]:
```
X_train = X_train.reset_index(drop=True)
y_train = y[y.patient <= 38].reset_index(drop=True)


X_test = X_test.reset_index(drop=True)
y_test = y[y.patient > 38].reset_index(drop=True)
```

[15]:
```
X_train_fl = X_train.astype(float, 64)
X_test_fl = X_test.astype(float, 64)
```

```
scaler = StandardScaler()
X_train_scl = scaler.fit_transform(X_train_fl)
X_test_scl = scaler.transform(X_test_fl)
```

[16]:
```
pca = PCA()
pca.fit_transform(X_train)
```

[16]:
```
array([[-4.12032149e+03,  8.43574289e+03, -1.39441668e+04, …,
         2.51106855e+03,  3.92187680e+03,  1.22642865e-11],
       [ 1.86283598e+04,  1.44078238e+04,  1.66177453e+04, …,
        -2.30960132e+02, -1.04099055e+03,  1.22642865e-11],
       [-1.58238732e+04,  1.40484268e+04,  4.73320627e+04, …,
         5.48675197e+02, -2.26227734e+03,  1.22642865e-11],
       …,
       [ 6.50848905e+04, -5.49595793e+04,  1.67854688e+04, …,
         1.18708820e+01, -1.47894896e+03,  1.22642865e-11],
       [ 4.97670530e+04, -3.81956823e+04,  2.93511865e+03, …,
         2.66462156e+03,  7.99461277e+02,  1.22642865e-11],
       [ 1.08241948e+04, -1.68550421e+04, -9.46017931e+02, …,
        -2.04773331e+03, -1.96917341e+03,  1.22642865e-11]])
```

[17]:
```
total = sum(pca.explained_variance_)
k = 0
current_variance = 0
while current_variance/total < 0.90:
    current_variance += pca.explained_variance_[k]
    k = k + 1


pca = PCA(n_components=k)
X_train.pca = pca.fit(X_train)
X_train_pca = pca.transform(X_train)
X_test_pca = pca.transform(X_test)
```

[18]:
```
#1st implementation algorithm- k means
kmeans = KMeans(n_clusters=2, random_state=0).fit(X_train_scl)
km_pred = kmeans.predict(X_test_scl)

print('K-means accuracy:', round(accuracy_score(y_test.iloc[:,1], km_pred), 3))

cm_km = confusion_matrix(y_test.iloc[:,1], km_pred)

ax = plt.subplot()
sns.heatmap(cm_km, annot=True, ax = ax, fmt='g', cmap='Greens')


ax.set_xlabel('Predicted labels')
```

```
ax.set_ylabel('True labels')
ax.set_title('K-means Confusion Matrix')
ax.xaxis.set_ticklabels(labels)
ax.yaxis.set_ticklabels(labels, rotation=360);
```

```
K-means accuracy: 0.765
```

[19]:
```
#2nd implementation algorithm-naive bayes
nb_model = GaussianNB()

nb_model.fit(X_train, y_train.iloc[:,1])

nb_pred = nb_model.predict(X_test)

print('Naive Bayes accuracy:', round(accuracy_score(y_test.iloc[:,1], nb_pred),␣
↪3))

cm_nb =  confusion_matrix(y_test.iloc[:,1], nb_pred)

ax = plt.subplot()
sns.heatmap(cm_nb, annot=True, ax = ax, fmt='g', cmap='Greens')
```

```
ax.set_xlabel('Predicted labels')
ax.set_ylabel('True labels')
ax.set_title('Naive Bayes Confusion Matrix')
ax.xaxis.set_ticklabels(labels)
ax.yaxis.set_ticklabels(labels, rotation=360);
```

```
Naive Bayes accuracy: 0.912
```

[20]:
```
#3rd implementation algorithm-support vector machines

svm_param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001, 0.
↪00001, 10], "kernel": ["linear", "rbf", "poly"], "decision_function_shape" :␣
↪["ovo", "ovr"]}


svm_grid = GridSearchCV(SVC(), svm_param_grid, cv=3)

svm_grid.fit(X_train_pca, y_train.iloc[:,1])

print("Best Parameters:\n", svm_grid.best_params_)


best_svc = svm_grid.best_estimator_
```

```
svm_pred = best_svc.predict(X_test_pca)
print('SVM accuracy:', round(accuracy_score(y_test.iloc[:,1], svm_pred), 3))

cm_svm =  confusion_matrix(y_test.iloc[:,1], svm_pred)

ax = plt.subplot()
sns.heatmap(cm_svm, annot=True, ax = ax, fmt='g', cmap='Greens')


ax.set_xlabel('Predicted labels')
ax.set_ylabel('True labels')
ax.set_title('SVM Confusion Matrix')
ax.xaxis.set_ticklabels(labels)
ax.yaxis.set_ticklabels(labels, rotation=360);
```

```
Best Parameters:
 {'C': 0.1, 'decision_function_shape': 'ovo', 'gamma': 1, 'kernel': 'linear'}
SVM accuracy: 0.941
```

```
C:\Anaconda\lib\site-packages\sklearn\model_selection\_search.py:814:
DeprecationWarning: The default of the `iid` parameter will change from True to
False in version 0.22 and will be removed in 0.24. This will change numeric
results when test-set sizes are unequal.
  DeprecationWarning)
```

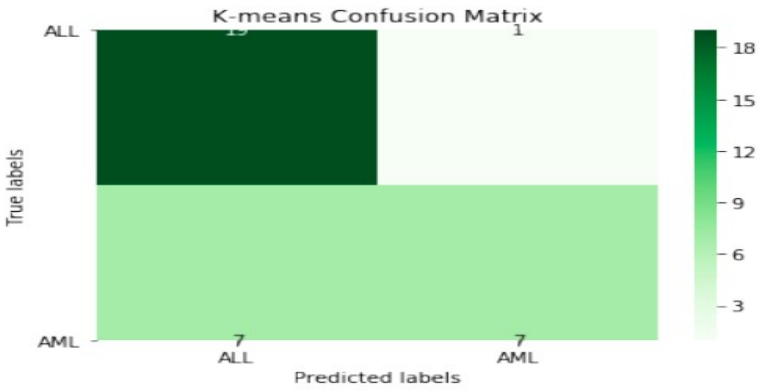
[21]:
```
#4th implementation alogrithm-logistic regression
log_grid = {'C': [1e-03, 1e-2, 1e-1, 1, 10],
                  'penalty': ['l1', 'l2']}

log_estimator = LogisticRegression(solver='liblinear')

log_model = GridSearchCV(estimator=log_estimator,
                  param_grid=log_grid,
                  cv=3,
                  scoring='accuracy')

log_model.fit(X_train, y_train.iloc[:,1])

print("Best Parameters:\n", log_model.best_params_)


best_log = log_model.best_estimator_


log_pred = best_log.predict(X_test)

print('Logistic Regression accuracy:', round(accuracy_score(y_test.iloc[:,1],␣
 ↪log_pred), 3))

cm_log =  confusion_matrix(y_test.iloc[:,1], log_pred)

ax = plt.subplot()
sns.heatmap(cm_log, annot=True, ax = ax, fmt='g', cmap='Greens')

ax.set_xlabel('Predicted labels')
ax.set_ylabel('True labels')
ax.set_title('Logistic Regression Confusion Matrix')
ax.xaxis.set_ticklabels(labels)
ax.yaxis.set_ticklabels(labels, rotation=360);
```

```
Best Parameters:
 {'C': 0.1, 'penalty': 'l1'}
Logistic Regression accuracy: 1.0

C:\Anaconda\lib\site-packages\sklearn\model_selection\_search.py:814:
DeprecationWarning: The default of the `iid` parameter will change from True to
False in version 0.22 and will be removed in 0.24. This will change numeric
results when test-set sizes are unequal.
  DeprecationWarning)
```

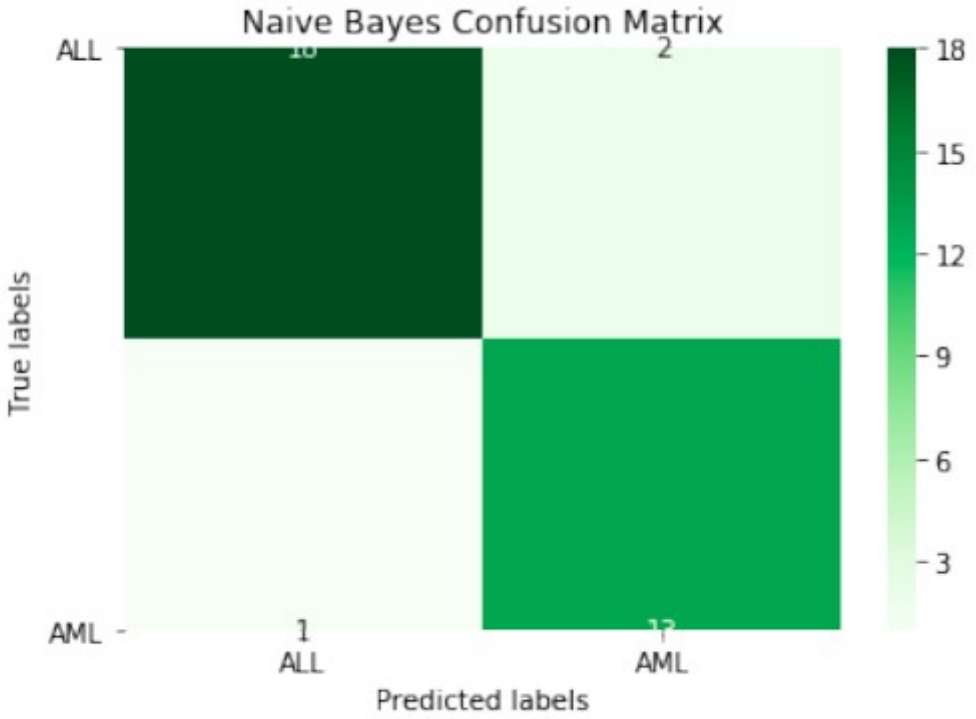
[22]:
```
#5th implementation-XG Boost
xgb3_model = xgb.XGBClassifier()
xgb3_model.fit(X_train, y_train.iloc[:,1])

xgb3_pred = xgb3_model.predict(X_test)

print('XGB (no PCA or Grid Search) accuracy:', round(accuracy_score(y_test.
 ↪iloc[:,1], xgb3_pred), 3))

cm_xgb3 = confusion_matrix(y_test.iloc[:,1], xgb3_pred)

ax = plt.subplot()
sns.heatmap(cm_xgb3, annot=True, ax = ax, fmt='g', cmap='Greens')


ax.set_xlabel('Predicted labels')
ax.set_ylabel('True labels')
```

7. Results and discussion:

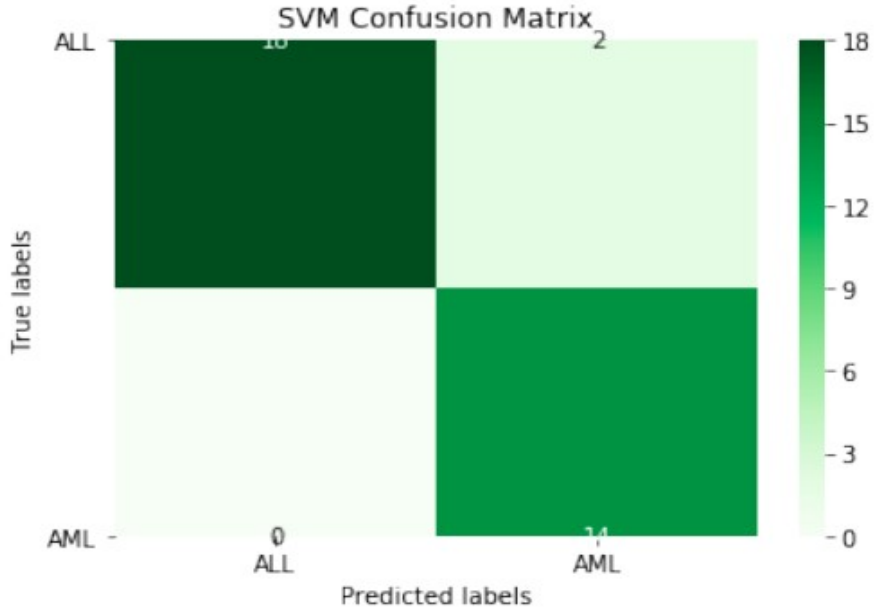
```
K-means accuracy: 0.765
```



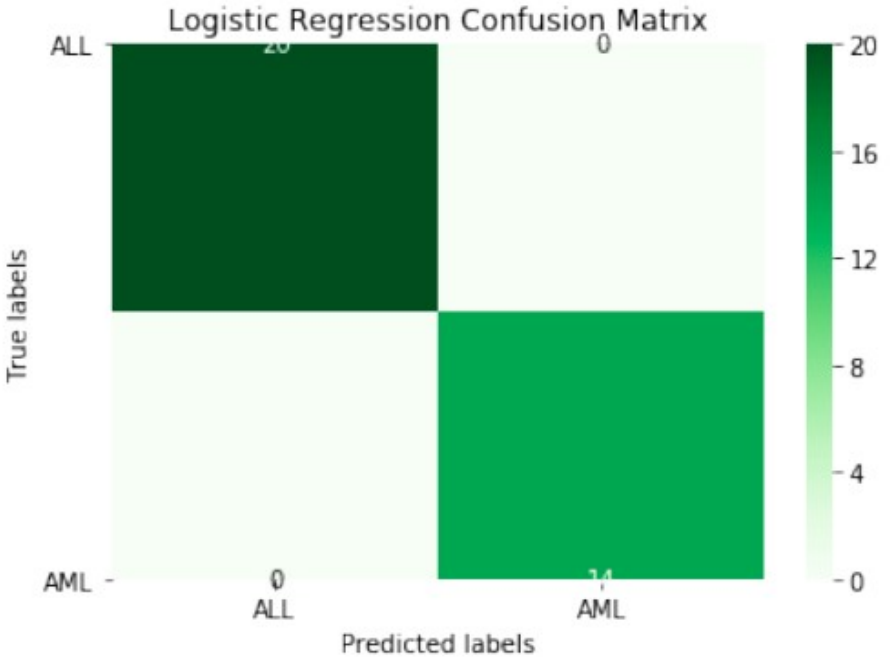
```
Naive Bayes accuracy: 0.912
```



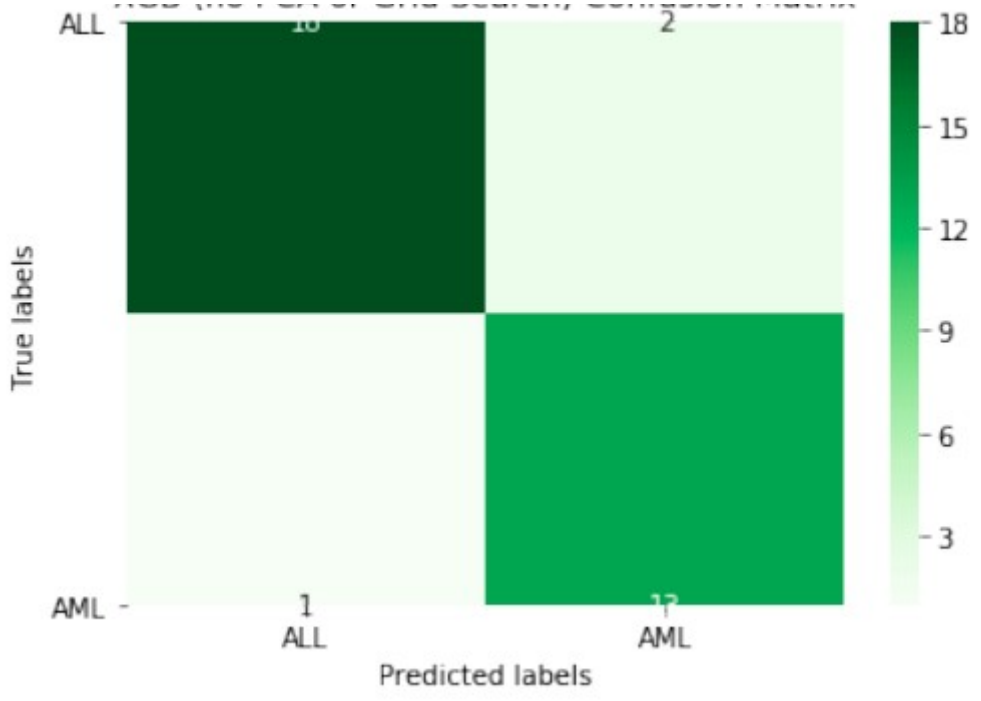

```
SVM accuracy: 0.941

C:\Anaconda\lib\site-packages\sklearn\model_selection\_search.py:814:
DeprecationWarning: The default of the `iid` parameter will change from True to
False in version 0.22 and will be removed in 0.24. This will change numeric
results when test-set sizes are unequal.
  DeprecationWarning)
```




```
Logistic Regression accuracy: 1.0
```




```
XGB accuracy: 0.912
```



We obtained the following accuracy value for the different algorithms we used.

1)k-Means: 0.765

2)Naive Bayes:0.912

3)SVM:0.941

4)Logistic regression:1.0

5)XG-Boost:0.912

From above, we can see the highest value of accuracy is of Logistic regression. Hence we can conclude that the logistic regression algorithm prove to be the best one for our dataset.

8. Conclusion:

Hereby, we can conclude from our project that after doing a proper study through various research papers and comparative analysis. Logistic regression has proved to be the best approach for our dataset and that we get the best accuracy through it when compared to other algorithms we had taken under consideration. Also, we can say that the detection of leukaemia patients through this algorithm has helped to sort the patients suffering from acute or chronic leukaemia. Besides this, we know that leukaemia being fatal cancer needs to be rectified at the earliest. Hence our report can help the doctor in identifying the right type of Leukaemia at the earliest since we have been able to show the Logistic regression method to be the best in terms of accuracy.

9. Reference:

1. Machine learning applications in the diagnosis of leukemia: Current trends and future directions. Haneen T. Salah Ibrahim N. Muhsen Mohamed E. Salama Tarek Owaidah Shahrukh K. Hashmi.
2. Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization. Worawut Srisukkham, Li Zhang, Siew Chin Neoh, Stephen Todryk, Chee Peng Lim
3. Acute leukemia classification using convolution neutral network in clinical decision support system. Thanh.TTP, Giao N. Pham, Jin-Hyeok Park, Kwang-Seok Moon, Suk-Hwan Lee, and Ki-Ryong Kwon.
4. Image processing and machine learning in the morphological analysis of blood cells J. Rodellar S. Alférez A. Acevedo A. Molina A. Merino
5. Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks. Sarmad Shafique MS, Samabia Tehsin, PhD
6. Recognition and prediction of leukemia with Artificial Neural Network (ANN) Saeid Afshar, Fahimeh Abdolrahmani, Fereshte Vakili Tanha, Mahin Zohdi Seif, Kobra Taheri
7. White Blood Cells Segmentation and Classification to Detect Acute Leukemia Ms. Minal D. Joshi, Prof. Atul H. Karode, Prof. S.R.Suralkar
8. Automatic Recognition of Acute Myelogenous Leukemia in Blood Microscopic Images Using K-means Clustering and Support Vector Machine. Fatemeh Kazemi, Tooraj Abbasian Najafabadi, and Babak Nadjar Araabi1,
9. Automated Detection of Acute Leukemia using K-mean Clustering Algorithm. Sachin Kumar, Sumita Mishra, Pallavi Asthana, Pragya
10. Application of Support Vector Machine and Genetic Algorithm for Improved Blood Cell Recognition. Stanislaw Osowski, Robert Siroí c, Tomasz Markiewicz, and Krzysztof Siwek
11. Detection of Leukemia in Human Blood Sample based on Microscopic Images. Mishra, S. J., & Deshmukh,A.P.
12. Classification of Acute Leukaemia Cells Using Multilayer Perceptron and Simplified Fuzzy ARTMAP Neural Networks. Nasir A. A., Mashor M. Y., Hassan R.
13. Chatap N., Shibu S., “Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbour
14. N. T. Umpon. Patch based white blood cell nucleus segmentation using fuzzy clustering.
15. L.B. Dorini, R. Minetto, and N.J. Leite. White blood cell segmentation using morphological operators and scale-space analysis
16. Identification of Leukemia Subtypes from Microscopic Images Using Convolutional Neural Network Nizar Ahmed , Altug Yigit, Zerrin Isik and Adil Alpkocak
17. Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks Sarmad Shafique, MS1 and Samabia Tehsin, PhD1
18. Detection and Classification of Blood Cancer from Microscopic Cell Images Using SVM KNN and NN Classifier,Sachin Paswan,Yogesh Kumar Rathore
19. Prediction of cancer using customised fuzzy rough machine learning approache Chinnaswamy Arunkumar , Srinivasan Ramakrishnan

20. On the Feature Selection Methods and Reject Option Classifiers for Robust Cancer Prediction MUHAMMAD HAMMAD WASEEM , MALIK SAJJAD AHMED NADEEM , ASSAD ABBAS ALIYA SHAHEEN , WAJID AZIZ , ADEEL ANJUM , UMAR MANZOOR , MUHAMMAD A. BALUBAID, AND SEONG-O SHIM
21. Leukemia Detection from Blood Smears Lokman Faivdullah, Farid Azahar, Zaw Zaw Htike, and Wei Yan Nyein Naing
22. Bayesian clustering of flow cytometry data for the diagnosis of B-Chronic Lymphocytic Leukemia q John Lakoumentas a , John Drakos a,b , Marina Karakantza b , George C. Nikiforidis a , George C. Sakellaropoulos a
23. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT R Shouval 1,2 , O Bondi 3 , H Mishan 3 , A Shimoni 1 , R Unger 3 and A Nagler 1
24. Cell Population DataeDriven Acute Promyelocytic Leukemia Flagging Through Artificial Neural Network Predictive Modeling Rana Zeeshan Haider , Ikram Uddin Ujjan and Tahir S. Shamsi
25. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia Liyan Pan , Guangjian Liu , Fangqin Lin , Shuling Zhong , Huimin Xia , Xin Sun & Huiying Liang
26. Scalable Prediction of Acute Myeloid Leukemia Using High-Dimensional Machine Learning and Blood Transcriptomics
27. Cell Population DataeDriven Acute Promyelocytic Leukemia Flagging Through Artificial Neural Network Predictive Modeling,Volume 13 Number 1 ,January 2020 ,pp. 11–16 Authors:Rana Zeeshan Haider, Ikram Uddin Ujjan and Tahir S. Shamsi
28. Computer Aided Diagnostic System for Detection of Leukemia using Microscopic Images Authors: Jyoti Rawat, Annapurna Singh, H S Bhadauria,Jitendra Virmani
29. Leukemia Blood Cell Image Classification Using Convolutional Neural Network T. T. P. Thanh, Caleb Vununu, Sukhrob Atoev, Suk-Hwan Lee, and Ki-Ryong Kwon
30. Classification of Leukemia Blood Samples Using Neural Networks MALEK ADJOUADI, MELVIN AYALA, MERCEDES CABRERIZO, NUANNUAN ZONG, GABRIEL LIZARRAGA, and MARK ROSSMAN
31. A Hybrid Deep Learning Architecture for Leukemic B-lymphoblast Classification (Sara Hosseinzadeh Kassani,Peyman Hosseinzadeh Kassani,Michal J. Wesolowski,Kevin A. Schneider,Ralph Deters)
32. High Accuracy Classification of White Blood Cells using TSLDA Classifier and Covariance Features Hamed Talebi1, Amin Ranjbar2, Alireza Davoudi1, Hamed Gholami1, Mohammad Bagher Menhaj
33. Automated AML Detection from Complete Blood Smear Image Using KNN Classifier Nayana B. Sen, Mercy Mathew
34. Automated Detection of Acute Leukemia using K-mean Clustering Algorithm Sachin Kumar, Sumita Mishra, Pallavi Asthana, Pragya
35. Cell Population Data–Driven Acute Promyelocytic Leukemia Flagging Through Artificial Neural Network Predictive Modeling Author links open overlay panelRana ZeeshanHaider∗‡Ikram UddinUjjan†Tahir S.Shamsi
36. Classification of acute leukemia using medical-knowledge-based morphology and CD marker Author links open overlay panelJakkrichLaosaiKosinChamnongthai
37. A decision support system for Acute Leukaemia classification based on digital microscopic images Author links open overlay panelAhmed S.NegmaOsama A.HassanaAhmed H.Kandilb

38. .Prediction and Diagnosis of Leukemia Using Classification Algorithms Khaled A. S. Abu Daqqa,Ashraf Y. A. Maghari,Wael F. M. . Al Sarraj
39. Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers Jaroonrut Prinyakupt & Charnchai Pluempitiwiriyawej
40. Neutrosophic Sets and Fuzzy C-Means Clustering for Improving CT Liver Image Segmentation Authors Authors and affiliations Ahmed M. AnterAboul Ella HassanienMohamed A. Abu ElSoudMohamed F. Tolba
41. Applying Decision Trees to Gene Expression Data from DNA Microarrays: A Leukemia Case Study Oscar Picchi Netto, Sergio Ricardo Nozawa , Rafael Andres Rosales Mitrowsky ´, Alessandra Alaniz Macedo1, Jose Augusto Baranauskas ´
42. Image processing and machine learning in the morphological analysis of blood cells J. Rodellar S. Alférez A. Acevedo A. Molina A. Merino
43. Leukemia identification from bone marrow cells images using a machine vision and data mining strategy Gonzalez, Jesus A.a | Olmos, Ivanb | Altamirano, Leopoldoa | Morales, Blanca A.a | Reta, Carolinaa | Galindo, Martha C.a | Alonso, Jose E.c | Lobato, Rubenc
44. Analysis of clinical prognostic variables for Chronic Lymphocytic Leukemia decision-making problems J.deAndrés-Galianaad1Juan L.Fernández-Martíneza1OscarLuacesdJuan J.del CozdLeticiaHuergo-ZapicocAndreaAcebes-HuertacSegundoGonzálezcAna P.González-Rodríguezb
45. Discrimination of malignant neutrophils of chronic myelogenous leukemia from normal neutrophils by support vector machine panelWanmaoNiaXiangminTongaWenbinQianaJieJinaHongchanZhaob
46. Classification of acute lymphoblastic leukemia using deep learning Amjad Rehman Naveed Abbas Tanzila Saba Syed Ijaz ur Rahman Zahid Mehmood Hoshang
47. Uncovering Differential Multi-microRNA Signatures of Acute Myeloid and Lymphoblastic Leukemias with a Machine-Learning-Based Network Approach Julián Candia, Srujana
48. A fuzzy neural approach for leukemia cancer classification Dr b b m krishna kanth
49. Classification of wbc for blood cancer diagnosis using deep learning neural networks Dhvani kansara,shaunak sompura,saifil momin, Mitchell d'silva
50. Classifying White Blood Cells in Blood Smear Images using a Convolutional Neural Network Gulshan Sharma, Rakesh Kumar
51. Unsupervised Blood Microscopic Image Segmentation and Leukemia Detection using Color based Clustering Subrajeet Mohapatra, Dipti Patra1 and Sanghamitra Satpathy
52. Efcient Classifcation of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features Ahmed T. Sahlol ,Philip Kollmannsberger & Ahmed A. Ewees