Z-Score Normalized Features with Maximum Distance Measure Based k-NN Automated Blood Cancer Diagnosis System

Article *in* ECS Transactions · April 2022

DOI: 10.1149/10701.11945ecst

2 authors, including:

Viswanathan Perumal
Vellore Institute of Technology
44 PUBLICATIONS **1,248** CITATIONS

Z-Score Normalized Features with Maximum Distance Measure Based k-NN Automated Blood Cancer Diagnosis System

ECS Transactions, 107 (1) 11945-11956 (2022)
10.1149/10701.11945ecst ©The Electrochemical Society

Z-Score Normalized Features with Maximum Distance Measure Based k-NN Automated Blood Cancer Diagnosis System

Umarani P[a] and Viswanathan P[b]

[a]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu-632 014, India
[b]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu-632 014, India

Leukemia is a blood-forming cancer disease characterized by the abnormal growth of White Blood Cells. Early detection and treatment of cancer reduce mortality and increase the survival rate. Most of the k-NN approaches used a different combination of statistical and geometrical features of the nucleus and cytoplasm with and without normalization, resulting in an uncertain range of features. Minkowski, Euclidean, City-block, Correlation, and Cosine distance metrics discover the average similarity between features predicts Leukemia with less accuracy. This paper proposes a z-score normalized feature set combined with a k-NN maximum distance measure-based automated blood cancer diagnosis system to address this problem. Initially, the features are normalized using the z-score normalization method, which eliminates the mis-classification outcomes. The Chebyshev distance measure, which finds the maximum similarities between the features when k=1, has the highest accuracy of 97.92%. Furthermore, tuning the parameters by grid-search improves the performance rate by 98.65% .

1. Introduction

Leukemia was discovered in the blood and bone marrow and was difficult to find earlier. The immature growth of blood cells affects the immune system leads to other diseases. Depending on the development, it can be classified into two types: acute and chronic. Acute Leukemia multiple the blast cells very fast, and in chronic Leukemia, it multiple very slowly (1). According to French-American-British, Acute Leukemia is further classified into Acute Myeloid Leukemia (AML) common in children and Acute Lymphoblastic Leukemia (ALL) among children and adults, and its growth is rapid. Chronic Leukemia is subclassified into Chronic Myeloid Leukemia (CML) and Chronic Lymphoblastic Leukemia (CLL). 61,090 people will be diagnosed with Leukemia, and 23,660 will die from the disease in the USA. A haematologist will use a light microscope to look for abnormal white blood cells and classify them into types and subtypes. Manual detection in pathology is time consuming and costly due to expensive pathology instruments. It is therefore automated for quick and accurate results (2).

The database images of ALL-IDB (ALL-IDB1 and ALL-IDB2) stimulate new studies about Leukemia research, mainly Acute Lymphoblastic Leukemia, a publicly available

dataset (3). Many studies concentrated on ALL-IDB2 have cropped the area of interest of normal leukocytes and leukemic cells belonging to the ALLIDB1 dataset (4-7). Acute Myeloid Leukemia is characterized by the existence of myeloid blast cells in WBC's (5,15). Identifying the normal and abnormal cells depends on the characteristics of the white blood cells and the quantitative features of the nucleus and the cytoplasm, such as shapes, size, color, and texture features (5-8). Shape features like area, perimeter, eccentricity, circularity, solidity, major axis, minor axis, etc., specify the geometry or shape of the nucleus and the cytoplasm. In addition, the ratio between the area of the cytoplasm and the nucleus indicates the maturity of a cell (2). Texture features provide details about the intensity of difference between the pixels in binary and gray-level images, such as energy, contrast, correlation, homogeneity, etc., in statistical terms (6). Color features extracted from the color channels of the RGB image indicate the abnormalities. This specific information differentiates the leukocytes from other white blood cells.

A combination of two or more features of the nucleus and the cytoplasm has been considered for classification. The choice of the features has been varied, emphasizing size, shapes, and color to differentiate the types of acute leukemia (5). Texture-based features are carried out separately, and statistical and geometrical features of the nucleus region are based on its visible and invisible characters instead of cytoplasm (6, 10). This prominent extraction of features plays an essential part in increasing the performance and reducing the complexity of a classifier; hence 22 statistical and geometrical features of both nucleus and cytoplasm are utilized.

Each feature or attribute of normal and leukemic cells has a different range of values, resulting in uncertainty of each attribute of varying magnitudes leading to a poor learning model. Therefore, it is necessary to transform with normalization to equalize the range of values in -1.0 to 1.0 and 0.0 to 1.0. It will positively affect the classifier performance since some of the features have large values, and without normalization, these large values may harm the classifier. To normalize features, methods such as min-max normalization, z-score standardization, and decimal scaling are used (9,11). Min-Max normalization scales variables between 0 and 1 on the basis of their difference between their minimum and maximum values. Whereas Z-Score normalization keeps the mean constant at 0 and the variance constant at 1, decimal scaling divides each value of data by the feature's maximum absolute value (12,13). The direct application of distance measures such as Euclidean distance, city block, Minkowski, Chebyshev, correlation, etc., in k-NN without normalization affect the results (14). Hence, compared to these normalizations, Z-score normalization has more advantages of handling outliers applied with various proportions of training samples used to improve classification performance.

The k-NN learning algorithm classifies the normalized data samples into cancerous and non-cancerous shown in Fig (1). The similarity between the features Measured by distance measure and GridSearch cross-validation is used to obtain the k value with the best accuracy (15). The similarity of each test sample is compared with training samples among the features are done by distance measures. Among these, Euclidean distance is being the most extensively used in many studies. Even though it performs well, choosing another best distance measure is essential in calculating the distances. All distance functions should perform differently over the different features. Compared to other distance measures, the maximum value distance, i.e., Chebyshev distance, gives the

highest accuracy in the biomedical data. Therefore, choosing which distance function is used in k-NN is the best choice for different proportions of training samples both in minimum and maximum distance.

The GridSearch cross-validation selects the best parameters for k processing with the Chebyshev distance utilized in the k-nearest neighbor achieves the highest accuracy compared to other distance measures with k= 1. The k values obtained will be used to determine cell types using k-NN, which will be searched for the majority of cell types from a number k which has the closest resemblance (13,15). This model is tested with a value of k starts from 1 to 10 of performance metrics such as accuracy, precision, recall, and f1-score to predict the best model with different distance measures in k-NN for predicting cancerous and non-cancerous cells.

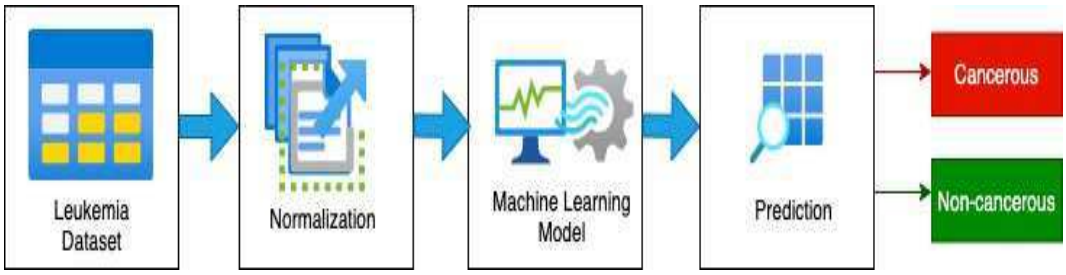

Figure 1. Automated Blood Cancer Diagnosis System

Z-Score normalized features with maximum distance measure-based k-NN automated blood cancer diagnosis system is applied with 22 statistical and geometrical features of the nucleus and the cytoplasm of the white blood cells used to improve accuracy and overcome the computation problem. This paper is organized as follows. In section 2, the related work has been explored. The proposed approach with the context of an experiment is in section 3. Whereas section 4 elaborates the selected input features and the experimental outcomes and Section 5 concludes the paper.

2. Related Work

The k-NN automated system extracts and analyzes white blood cells and discriminate blast cells from healthy cells based on features, distance measure, and k value. The k-NN classifier (5) uses 12 features like area of the blast, radius, and perimeter, standard deviation and mean of RGB color, and shape features. This feature is evaluated by initializing k value 1-10 with the four-distance metrics are Euclidean $E\_d(x,y)$, city-block or Manhattan $Cb\_d(x,y)$, cosine $Co\_d(\mathrm{x},\mathrm{y})$ and correlation $\mathrm{Cor\_d(x,y)}$.

$$E\_d(x,y) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - y_i)^2}{m}} \qquad [1]$$

$$Cb\_d(x,y) = \sum_{i=1}^{m} |x_i - y_i| \quad [2]$$

$$Co\_d(\mathrm{x},\mathrm{y}) = 1 - \frac{\sum_{i=1}^{m} x_i y_i}{\sqrt{\sum_{i=1}^{m} {x_i}^2}\sqrt{\sum_{i=1}^{m} {y_i}^2}} \quad [3]$$

$$\mathrm{Cor\_d}(\mathrm{x},\mathrm{y}) = \frac{1}{2}\left(1 - \frac{\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i-\bar{x})^2}\sqrt{\sum_{i-1}^{m}(y_i-\bar{y})^2}}\right) \quad [4]$$

The *Co _ d* (*x, y*) resulted in higher prediction at k=4 compared to other distance metrics. This approach is manual, and features are not normalized, requiring complex calculation to improve the accuracy, which may take more time to classify.

The customized k-NN (6) system based on the modification of distance metric in Eqn. [5]. The GLCM texture, statistical and geometric features of nucleus is evaluated using a customized Euclidean distance metric shown in Eqn. [6]. kNN with *E _ dm* (*x, y*) achieved more accuracy than the *E _ d* (*x, y*). This approach attempts only centroid distance metric and is not focused on normalization leads to less accuracy.

$$E\_dm(x,y) = \sqrt{(x_i + y_i)^2 + (x_i - y_i)^2} \quad [5]$$

The k-NN Classifier (1) used the distance measure with weighted distance metric in Eqn. [8] and unweighted distance metrics in Eqn. [1], [6] and [7]. The six features such as area, nucleus ratio, circularity, perimeter, mean, and standard deviation are used for classification (17). Then k-fold cross-validation was conducted to avoid overfitting in classification. The unweighted minkowski in Eqn. [8] provided higher accuracy compared to other weighted and unweighted distance metrics at k =19. However, the moderate recall and precision values obtained without normalization make it less suitable for practical purposes.

$$C_{d(x,y)} = \overset{max}{i}|x_i - y_i| \quad [6]$$

$$M\_du(x,y) = \left(\sum_{i=1}^{m} |x_i - y_i|^p\right)^{\frac{1}{p}} \quad [7]$$

$$\hat{d}(x_i, y_i) = \frac{\sum_{i=1}^{m} w(x_i,y_i) * d(x_i,y_i)}{\sum_{i=1}^{m} w(x_i,y_i)} \quad [8]$$

where, $x_i = (x_1, x_2, \ldots, x_m)$, $y_i = (y_1, y_2, \ldots, y_m)$, $w(x_i, y_i) = \frac{1}{d(x_i,y_i)}$, m is the dimensionality of the feature space, p = 2, the Minkowski distance in Eqn. [7] gives the Euclidean distance shown in Eqn. [1] and p = ∞, gives the Chebyshev distance shown in Eqn.[6].

3. k-NN Automated Z-Score Normalized Leukemia Diagnosis System

Different features or attributes in Leukemia dataset having different ranges of value are pre-processed using normalization to make it on a suitable range. Then, the normalized

features are split into training and testing samples. Training samples as input to the model by checking the closest resemblances among features by k-NN Chebyshev distance measure are validated with the testing samples for prediction and accurate classification of cancerous and non-cancerous cells as shown in Fig. [2].

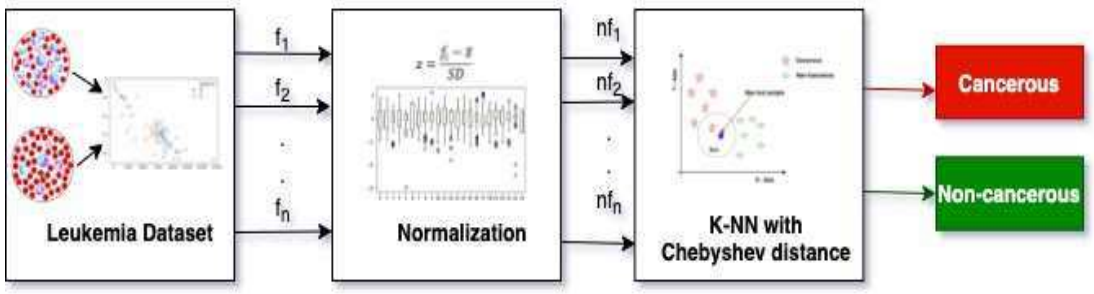

Figure 2. k-NN automated Z-Score normalized leukemia diagnosis system

The Leukemia dataset is composed of 22 statistical and geometrical features, or attributes, of WBCs, the nucleus, and the cytoplasm, which were extracted from fuzzy c-means (4). The features such as WBC area (f_1), WBC convex area (f_2), WBC perimeter (f_3), eccentricity wbc (f_4), solidity wbc (f_5), orientation wbc (f_6), nucleus area (f_7) , nucleus ratio (f_8), perimeter nucleus (f_9), round nucleus (f_{10}), eccentricity nucleus (f_{11}), solidity nucleus (f_{12}), convex area nucleus (f_{13}), average cytoplasm of region (f_{14}), average cytoplasm granularity (f_{15}), average cytoplasm bl (f_{16}), entropy cytoplasm (f_{17}), minor axis of cell (f_{18}), minor axis of nucleus (f_{19}), axis mean ratio (f_{20}), major axis nucleus (f_{21}) and major axis (f_{22}) are taken as input attributes. The 22 input features and additional one target attribute is normalized through z-score normalization used to classify the leukemia cells with maximum distance k-Nearest Neighbor Classifier using Algorithm.1.

Algorithm 1: Pseudocode for Z-Score Normalized Features with Maximum Distance k-Nearest Neighbor Classifier

Given the features $f_i = (f_1, f_2, f_3, \ldots\ldots\ldots, f_{23})$where n = 23, k, training samples x_i and testing samples y_i.

for i = 1, 2, ..., 23

do

Calculate the mean

$\bar{x} = (f_1 + f_2 + f_3 + \cdots \ldots\ldots + f_{22})/23;$

Subtract the mean value with each feature and square them

$(f_1 - \bar{x})^2, (f_2 - \bar{x})^2, (f_3 - \bar{x})^2, \ldots\ldots, (f_{22} - \bar{x})^2;$

Calculate the variance

$Var(x) = [(f_1 - \bar{x})^2 + (f_2 - \bar{x})^2 + (f_3 - \bar{x})^2 + \ldots\ldots + (f_{22} - \bar{x})^2]/23;$

Find out the Standard deviation, σ

$\sigma = \sqrt{Var(x)};$

Normalize with z-score

$nf_i = \frac{f_i - \bar{x}}{\sigma}$;

end

Split the normalized dataset, nf_i into training samples,x_i and testing samples,y_i

for k =1, 2, ..., n

do

Choose the nearest neighbor as k=5;

Calculate the maximum value distance between the testing sample, $y_i$and all training sample, x_i using C_$d(x, y) = {}^{max}_{i}|x_i - y_i|$;

Sort the obtained distances in ascending order and choose the k samples nearest to the similar test samples;

Assign a class to test samples if it has the closest resemblance;

end

Rescale the whole dataset;

Train the k-nearest neighbor model on the whole dataset with all targets;

3.1. Z-Score Normalization

Dataset features $f_i = (f_1, f_2, f_3, \ldots\ldots\ldots, f_{23})$ will not be in the same range; hence, Z-Score Normalization, $nf_i = \frac{f_i - \bar{x}}{\sigma}$ is applied to transform it on a particular range, where f_i is the value is to be normalized, $\bar{x}$ is mean, and σ is Standard Deviation. It provides the equal importance to each attribute or feature and also efficiently handles the outliers compared to other normalization techniques. Normalization also helps to increase the accuracy performance of the classification is further evaluated using k-Nearest Neighbor with chebyshev distance measure for classification of data as shown in Fig. [3].

3.2. k-NN Based on Chebyshev Distance Measure

The k-NN model is applied to the normalized training features x_i chooses the closest resemblance among the number of neighbors determined by k. The normalized testing feature, y_i are assigned to a particular class whether cancerous or non-cancerous cells by finding out the largest difference among the features using Chebyshev distance ${}^{max}_{i}|x_i - y_i|$ with a minimum time of classifying the cells where m is the dimensionality of the feature space. The distance function must satisfy with the properties of $d(x, y) \geq 0$, $d(x, y) = 0\ iff\ x = y$, $d(x, y) = d(y, x)$ and $d(x, y) \leq d(x, z) + d(z, y)$. Then, to calculate the similarity measure $s(x, y)$ for the distance is in the range [0,1] is $s(x, y) = 1 - d(x, y)$ as shown in Fig. [3].

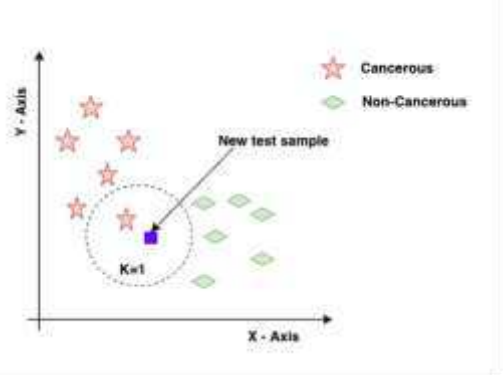

Figure 3. k-Nearest Neighbor Algorithm

4. Experimental Evaluation

The statistical and geometric features of 96 blood smear images are taken and pre-processed to transform the raw data into a useful and efficient format. We will use python v.3 and the sci-kit package to develop this k-NN model. The features such as WBC area (f_1), WBC convex area (f_2), WBC perimeter (f_3), eccentricity wbc (f_4), solidity wbc (f_5), orientation wbc (f_6), nuclueus area (f_7) , nucleus ratio (f_8), perimeter nucleus (f_9), round nucleus (f_{10}), eccentricity nucleus (f_{11}), solidity nucleus (f_{12}), convex area nucleus (f_{13}), average cytoplasm of region (f_{14}), average cytoplasm granularity (f_{15}), average cytoplasm bl(f_{16}), entropy cytoplasm (f_{17}), minor axis of cell (f_{18}), minor axis of nucleus (f_{19}), axis mean ratio (f_{20}), major axis nucleus (f_{21}) and major axis (f_{22}) are taken as input attributes. The attribute 'target' is added for predicting whether it is cancerous(M) or non-cancerous(B), depending upon the characteristics of the attributes.

A total of 23 attributes are utilized for the construction and processing of the model. Split the dataset into the training set and testing set to avoid a correlation among the features and evaluate the model accurately. This method mainly split in 50:50 ratio with 50 percent training set and 50 percent testing set and also analyzed the results in 70:30. Each attribute has a different range of values, i.e., WBC area, WBC convex area, nuclueus area, convex area nucleus are in a very large difference in magnitude. This impact creates the problem of classification and is solved by standardizing the attributes with z-score normalization. It gives equal weights or importance to each attribute based on the mean and standard deviation and shows the results of most attributes ranging from -2.00 to +2.00 because the input is normally distributed. Thus, normalization improves the accuracy and efficiency of the model.

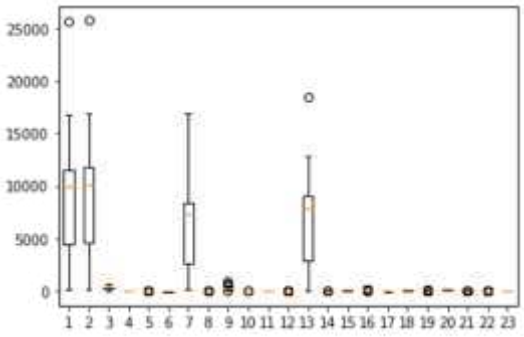

Figure 4. Original Features, f_i

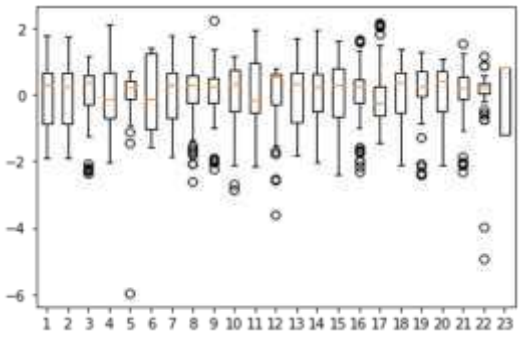

Figure 5. Normalized Features, nf_i

After scaling the features, it works with k-NN with all distance metrics and varying k values. When the k value is 1, 3, and 5, the accuracy is very high. Then it decreases when the k value increases with a higher error rate. The most frequent test samples are assigned to class label cancerous or non-cancerous by a majority vote by measuring the distance between the training and testing samples using distance metrics.

The performance criterion such as accuracy, precision, recall, and F1 Score of classifiers has been calculated for all the distance metrics from the Confusion Matrix. Confusion Matrix displays the frequency of correct and incorrect predictions. True Positive (TP) measures the number of correctly predicted cancerous cells. True Negative (TN) calculates the number of non-cancerous cells that are correctly predicted. False Positive (FP) specifies the number of non-cancerous cells is identified as cancerous. False Negative (FN) estimates the number of cancerous cells that are recognized as non-cancerous. Accuracy is the ratio of correctly predicts values over the total predicted values (18).

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad [9]$$

$$Precision = \frac{TP}{(TP+FP)} \quad [10]$$

$$Recall = \frac{TP}{(TP+FN)} \quad [11]$$

$$F1\_Score = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \quad [12]$$

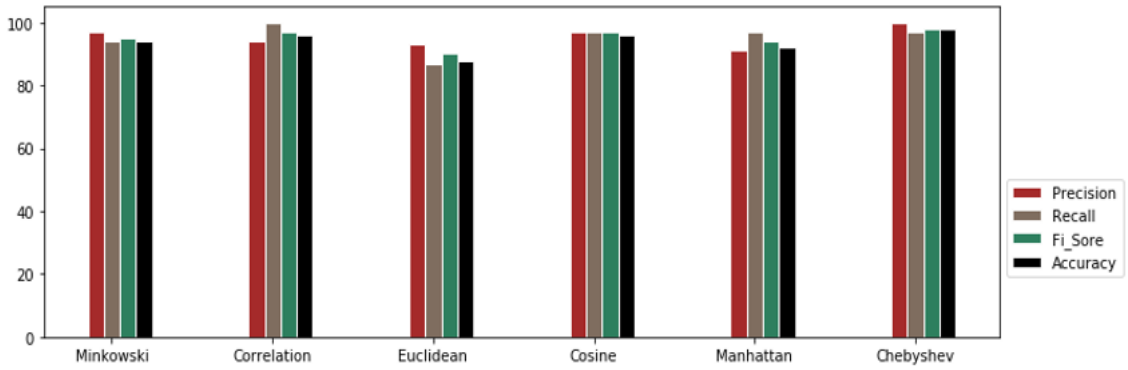

Figure 6. Performance measure of various distance measures

In this approach, the maximum distance measure, i.e., Chebyshev distance, gives better accuracy than other distance measures like Euclidean, Minkowski, city-block or manhattan, correlation, and cosine distance measure are shown in Fig. 6 and 7. It also shows that Minkowski achieves accuracy of 93.75% correlation achieves 95.83%, euclidean achieves 88%, cosine achieves 95.83%, Manhattan achieves 91.67%, and Chebyshev reaches 97.92% accuracy with a reasonable rate. The highest performance on this dataset is the Chebyshev distance measure. Furthermore, selecting the best parameters using grid search cross-validation can improve the accuracy of the model. 5-fold Cross-validation of multiple K values are checked with best-performing parameters score values of mean and standard deviation.

It chooses k=1 as the best value, and the results have better accuracy. Minkowski and Euclidean distances achieve improvement in the performance of the model. Minkowski distance reaches 96% accuracy, which is 2% higher than the earliest accuracy and Euclidean distance gets 6% more results in 94% accuracy and 0.73% improvement in Chebyshev distance through grid search cross-validation. After cross-validation, the performance will go high on specific distance measures. But the performance of correlation, cosine, and Manhattan remains the same after tuning the parameters. Compared with all techniques, the Chebyshev distance metric works well on this medical

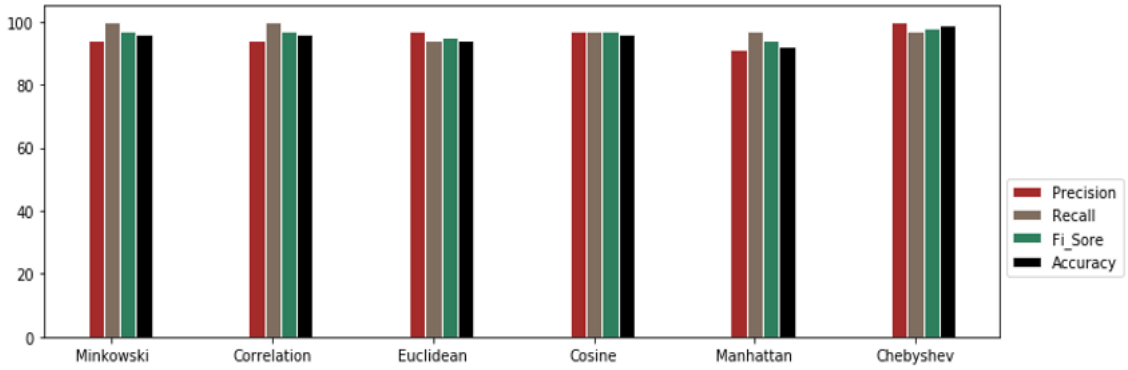

Figure 7. Performance measure of various distance measures with cross-validation

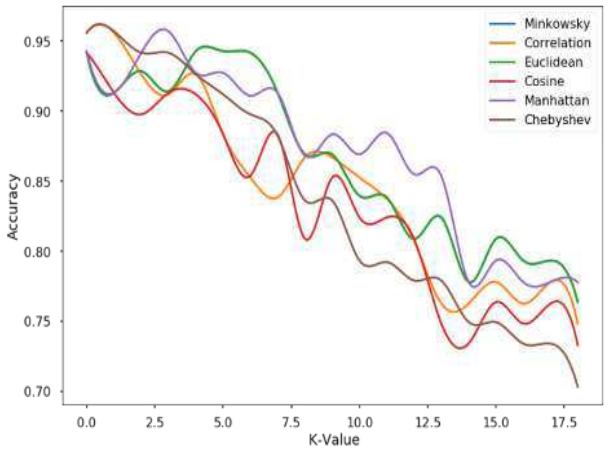

Figure 8. Accuracy based on k of 50:50 proportions

data with the highest accuracy of 98% on normal k-NN and 98.65% on tuned k-NN. The graph depicted in Fig.8. obtained upon completing the testing process in an experiment dealing with 50:50 proportion and also shows that Chebyshev with k=1 achieved the highest accuracy among the other measures.

5. Conclusion

An automated blood cancer diagnosis system using K-Nearest Neighbor has been implemented with z-score normalization and various distance measures. The z-score normalization effectively determined the outliers by fixing the features in a particular range. It is further classified using k-NN with minimum k value and maximum distance measure. Grid search cross-validation improves euclidean, manhattan, and Chebyshev by selecting the best parameters. The overall results of our experiment showed that Chebyshev distance only achieves the highest accuracy of 98%, precision of 100%, recall of 97%, and f1-score of 98%. We observed only initial k values as the highest prediction. Moreover, there is no significant difference among fewer distance measures with larger values of k. Therefore, we will work on weighted distance measures with different training and testing samples in future work.

References

1. Prakisya, Nurcahya Pradana Taufik, Liantoni, Febri, Hatta, Puspanda, Aristyagama, Yusfia Hafid and Setiawan, Andika. ” Utilization of K-nearest neighbor algorithm for classification of white blood cells in AML M4, M5, and M7” Open Engineering, vol. 11, no. 1, pp. **662-668** (2021).
2. Moshavash Z, Danyali H, Helfroush MS, An Automatic and Robust DecisionSupport System for Accurate Acute Leukemia Diagnosis from Blood Microscopic Images. J Digit Imaging. Oct;31(5):**702-717** (2018).

3. R. D. Labati, V. Piuri and F. Scotti, All-IDB: The acute lymphoblastic leukemiaimage database for image processing, 2011 18th IEEE International Conference on Image Processing, pp. **2045-2048** (2011).
4. Perumal, Viswanathan, Fuzzy C Means Detection of Leukemia Based on Morphological Contour Segmentation. Procedia Computer Science. 58. **84-90** (2015).
5. N. Z. Supardi, M. Y. Mashor, N. H. Harun, F. A. Bakri and R. Hassan, Classification of blasts in acute leukemia blood samples using k-nearest neighbour, 2012 IEEE 8th International Colloquium on Signal Processing and its Applications, pp. **461-465** (2012).
6. Umamaheswari, Duraiswamy, and Shanmugam Geetha, A framework for efficient recognition and classification of acute lymphoblastic leukemia with a novel customized-knn classifier. Journal of computing and information technology 26.2: **131-140** (2018).
7. Purwanti, E., Calista, E, Detection of acute lymphocyte leukemia using knearest neighbor algorithm based on shape and histogram features. Journal of Physics: Conference Series, **853(1)** (2017).
8. Chatap, Niranjan J. and S. Shibu, Analysis of blood samples for counting leukemia cells using Support vector machine and nearest neighbour." IOSR Journal of Computer Engineering 16: **79-87** (2014).
9. Chawla, Shagun and Kumar, Rajat and Aggarwal, Ekansh and Swain, Sarthak,Breast Cancer Detection Using K-Nearest Neighbour Algorithm, International Journal of Computational Intelligence IoT, Vol. 2, No. **4** (2018).
10. Subhan, Ms. Parminder Kaur, Significant Analysis of Leukemic Cells Extraction and Detection Using KNN and Hough Transform Algorithm, International Journal of Computer Science Trends and Technology (IJCST), Vol.- 3 Issue 1, pp. **27-33**, (2015).
11. L. Al Shalabi and Z. Shaaban," Normalization as a Preprocessing Engine forData Mining and the Approach of Preference Matrix," 2006 International Conference on Dependability of Computer Systems, pp. **207-214** (2006).
12. Abu Alfeilat HA, Hassanat ABA, Lasassmeh O, Tarawneh AS, Alhasanat MB,Eyal Salman HS, Prasath VBS. Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. Big Data. 7(4):**221-248** (2019).
13. Bhanja, Samit, and Abhishek Das. Impact of data normalization on deepneural network for time series forecasting (2018).
14. Singh, Dalwinder Singh, Birmohan, Investigating the impact of data normalization on classification performance. Applied Soft Computing. **105524** (2019).
15. E. S. Wiharto, S. Palgunadi, Y. R. Putra and E. Suryani," Cells identificationof acute myeloid leukemia AML M0 and AML M1 using K-nearest neighbour based on morphological images," 2017 International Conference on Data and Software Engineering (ICoDSE), pp. **1-6** (2017).
16. Talaat, Ahmed Abdeldaim, Ahmed Hassanien, Aboul Ella. Automatic acutelymphoblastic leukemia classification model using social spider optimization algorithm. Soft Computing. 23. **1-16**, (2019).

17. Ali, N., Neagu, D. Trundle, P. Evaluation of k-nearest neighbour classifierperformance for heterogeneous data sets. SN Appl. Sci. 1, 1559 (2019). 18. S. Chand and V. P. Vishwakarma," Leukemia Diagnosis using Computational Intelligence," 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. **1-7**, (2019).
18. Sukhia, K.N., Ghafoor, A., Riaz, M.M., Iltaf, N. Automated acute lymphoblastic leukaemia detection system using microscopic images. IET Image Process., 13, **2548-2553** (2019).