Comparative Study Between Decision Tree, SVM and KNN to Predict Anaemic Condition

Conference Paper · January 2019

CITATIONS
0

READS
88

3 authors, including:

Nahiyan Bin Noor
Chittagong University of Engineering & Technology
5 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Mohammad saeid Anwar
Chittagong University of Engineering & Technology
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

2019 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON)
28-30 November, 2019, Dhaka, Bangladesh

Comparative Study Between Decision Tree, SVM and KNN to Predict Anaemic Condition

Nahiyan Bin Noor[1*], Md. Saeid Anwar[2*], and Mrinmoy Dey[3]

[1,2,3]Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology, Chattagram-4349, Bangladesh.
[1]nahiyannabil@gmail.com, [2]u1402114@student.cuet.ac.bd, [3]mrinmoycuet@gmail.com

Abstract— **Anemia, a disease which is caused by an inadequacy of hemoglobin or red blood cells in the blood. It is very risky at the time of pregnancy, menstruation and in ICU sometimes causing death. So, it is a need of hemoglobin and detects anemia quickly. Usually, doctors examine the eye conjunctiva color and confirmed by a blood test which is painful, time-consuming and costly. In this study, a total of 104 people (54 males and 50 females) are collected with their clinical blood hemoglobin level, anemic condition and taken palpebral conjunctiva image. The images are captured with a cell phone camera of good resolution. By using the images, the percentage of the red, green and blue pixels are extracted in MATLAB, image processing method. Taking those features, the Hemoglobin level is plotted. A total of 81 data is taken for training purposes and 23 data for testing. For Anemia detection, the 81 data are trained with a used different classifier such as Linear SVM, Coarse Tree, and Cosine KNN and have been got highest accuracy of 82.61% in Decision Tree (Coarse) by testing 23 data.**

Index Terms— **Anemia, Eye Palpebral Conjunctiva, Image Processing, Decision Tree, SVM, KNN.**

I. INTRODUCTION

Anemia is a typical malady or state of the body where a human body needs red platelets to convey an adequate measure of oxygen to our tissues. As indicated by WHO [1], Numerous elements cause weakness, for example, iron inadequacy, nutrient B12, and nutrient A lack, parasitic diseases, and interminable aggravation. World wellbeing association (WHO) reports that [2,3] 24.8% of the all-out populace is influenced by sickliness. Ordinarily, weakness is recognized by estimating hemoglobin. It is realized that 13.5 to 17.5 gm/dl is ok for men and 12 to 15.5 gm/dl is satisfactory for ladies. In any case, clinically under 10 gm/dl is considered as Anemia. In this investigation, as per the clinical report, we have taken the individuals having hemoglobin levels under 10 gm/dl generally weak.

In 2016 Pooja Tukaram et al. [4] used machine learning to detect Anemia pallor. They used five learning methods and four classifiers where K-NN gave higher efficiency. 50 blood smear slides are used to achieve the result.500 instances of the dataset are used where 100 for each elliptocyte, macrocytes, echinocytes, teardrop cells, and normal RBCs. Each of the instances consists of 13 features with a target. In 2012 Magdeeswaran Veluchamy et al. [5] have used 4 geometrical and 16 statistical seven-moment invariant features to illustrate the classification of blood cells which results in 80% efficiency. K S Srinivasan et al. [6] designed a simpler method of diagnosing anemia in which they used non-invasive color analysis. After extracting RGB values from images ANN machine learning is used for hemoglobin measurement. The uniform image has used in this research which has a file format of BMP. The researcher uses a rubber band at the fingertip to put pressure on blood and they take pictures both with and without pressure. Found out a close correlation between the RGB values with and without pressure. Azwad Tamir et al. [7] illustrated a method of anemia detection using images of the conjunctiva. They had 19 subjects with known hemoglobin levels and took conjunctiva images of these subjects by a smartphone. Red and green spectra compared with an ideal value to determine the states anemic or non-anemic and finally got 78.9% accuracy. In 2018 Giovanni Dimauro [8] et al. depicted a system for assessing anemic condition. In this work, non-invasive measurement used to illustrate the results and a microlens associated with a smartphone used to capture the eye conjunctiva images. After taking 113 data of both anemic and healthy people, the KNN classification algorithm is applied to depict the assessment results. Using all 113 data they found a correlation of .745 which is better than the other two modes. Mahammad Firos shaik [9] et al. Present a new approach for diagnosing anemia and types of anemia. Fuzzy logic is used in this research and overall simulation has taken place in the LabVIEW software. A total of nine inputs of blood parameters have been used in the fuzzy logic system and depicted six types of anemia. They got an overall good result from this system which is time-saving. Robert G. Mannino et al. [10] present an eidolon of an overall non-invasive method of perception of anemia. It is a Smartphone app-based research work. They took fingernail images of 100 subjects of both anemic or healthy with the smartphone and analyzed the color values which gives the hemoglobin level instantly. Researchers formed a precision of ±2.4 g dL−1 and a sensitivity of 97% in the overall work. Jahidur Rahman Khan et al. designed a system using machine learning, where they used CART, linear discriminant analysis (LDA), support vector machine (SVM), LR, RF to detect an anemic condition of 600 children which achieved highest 68.53% accuracy in RF system [17]. Manish Jaiswal et al. discussed machine learning for detecting anemic conditions using various algorithms such as random forest, Naive-Bayes, and decision tree. Researchers found the highest 96.05%^ accuracy in NB [18].

Apart from all, in this study total, 104 data are collected from Chittagong Medical College Hospital, Chittagong and Cox's Bazar Medical College Hospital, Bangladesh with the clinical anemic condition and palpebral conjunctiva image which contained 54 males and 50 females. All data are taken from the patient after taking their consent. Among them 25 people are anemic and 79 people are non-anemic. From the total dataset 81 data are taken to train the model and 23 data kept separate for further testing the system. Basically, the image of their eye conjunctiva is captured and applied image processing to extract the percentage of red, green and blue. The aim of this work is to predict anemia using different

classifiers. Different classifiers are also used such as Linear SVM, Coarse Tree Cosine KNN to train the model that will predict the anemic condition of the patient. These three classifications also compared with themselves and the best algorithm is taken.

II. Methodology

The whole study has been done in MATLAB software with different built-in apps such as Color thresholding, classification learner. A flow chart is showing the overall procedure which is illustrated in fig. 1. The main part of the work is in the flow chart which will be described in detail below.

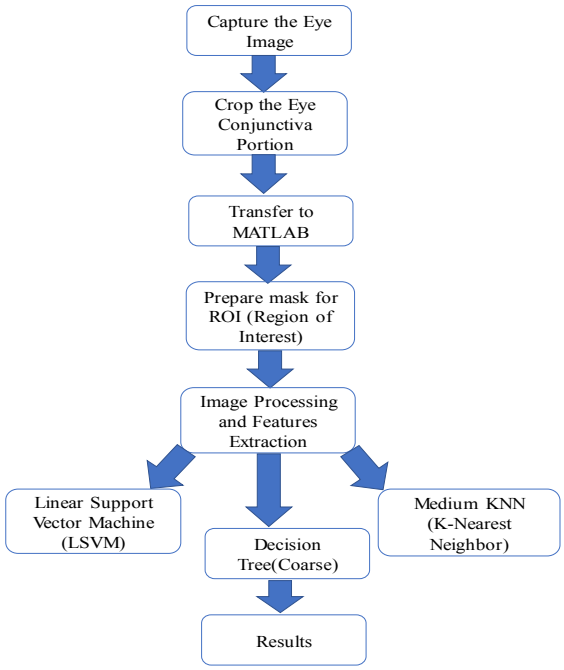

Fig. 1 The flowchart describing the overall procedure

A. *Capture the eye conjunctiva image and crop the eye conjunctiva portion*

The palpebral eye conjunctiva picture has been taken utilizing a 12-megapixel cell phone camera. It has been finished by pulling the lower eyelid and snapping the photo of the eye conjunctiva. At that point centering front conjunctiva, the picture has been taken.The steady light is required for taking the picture so a device represented in fig. 2 is made. Utilizing that the picture is caught by keeping on the flashlight of mobile. In fig. 3 the picture appears. The device is shown in fig. 2 basically a very low-cost box that is handmade spending only $1. It creates a lightless medium where any light can to enter as a result a constant medium of light is created by this box. This box is designed in such a way where a smartphone can easily be placed and by using that smartphone palpebral eye conjunctiva image can be captured. Then the image is cropped and taken only the conjunctiva portion. After cropping the image, it is taken to MATLAB for further processing. As everyone uses a smartphone nowadays, the cost of a smartphone camera is neglected as a part of our device. It is very user-friendly as there is no complication in wearing the box shown in fig. 2 and anyone can take the picture with their smartphones.

Fig. 2 Device creating constant light

Fig. 3 Palpebral Conjunctiva

B. *Prepare mask for ROI using MATLAB image thresholding app*

The pictures are edited, kept the portion of Interest and disposed of the white foundation shade of the picture by utilizing the MATLAB Color thresholding application. Some white shade made because of the flashlight of smartphones that additionally discarded utilizing RGB Thresholding appeared in Fig. 4.1, 4.2, 4.3. The color thresholding app of MATLAB can eliminate the unwanted portions of an image efficiently using RGB pixel color thresholding technique. At the point when it turned thoroughly dark this is called concealing where undesirable pixels are wiped out. After the picture looks like fig. 4.2. Normally white foundation shading would influence the red, green and blue pixel rate due to matrix value [255, 255, 255] for three pixels as it is realized that white is a full combination of Red, Green, and Blue. For black color, the matrix is [0 0 0] which has no effect in the total image and this is the reason the unwanted and defected portion is eliminated using color thresholding app.

Fig 4.1 White portion created due to flashlight

Fig 4.2 Unwanted white portion is eliminated using color thresholding

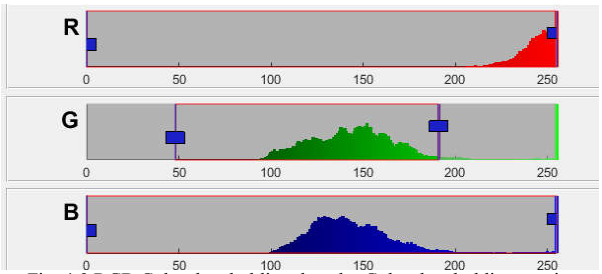

Fig. 4.3 RGB Color thresholding done by Color thresholding app in MATLAB

C. *Image processing and feature extraction in MATLAB*

Suner et al. [11] outlined that shading attributes of palpebral eye conjunctiva are compared to hemoglobin fixation that is paleness and that the portrayal of these qualities in an advanced arrangement depends on relative shades of red, green and blue for every pixel. Along these lines, from that eye conjunctiva pictures can be used as anemia indicator. Features can be extracted utilizing

MATLAB and image processing methods. The percentage of RGB pixel from eye image is extracted which is as follows.

The percentage of red pixel=

$$\frac{\text{Total red pixel}}{\text{Total Red pixel+Total Green pixel+Total blue Pixel}} * 100 \quad (1)$$

The percentage of green pixel=

$$\frac{\text{Total Green pixel}}{\text{Total Red pixel+Total Green pixel+Total blue Pixel}} * 100 \quad (2)$$

The percentage of blue pixel=

$$\frac{\text{Total Blue pixel}}{\text{Total Red pixel+Total Green pixel+Total blue Pixel}} * 100 \quad (3)$$

Using equation (1). (2) and (3) the most important three features are extracted and the features are the percentage of red, green and blue pixels.

TABLE I: LIST OF PEOPLE WITH THEIR PERCENTAGE OF RED, GREEN AND BLUE PIXEL IN EYE CONJUNCTIVA IMAGE ALONG WITH THEIR ANEMIC CONDITION

Name	%Red Pixel	%Green Pixel	%Blue Pixel	Anemic Condition
Shahalam	47.6372	26.7953	25.5675	Non Anemic
Rumana	45.0693	29.8506	25.0801	Anemic
Ajmal	46.7979	28.1768	25.0253	Non Anemic
Haradhan	42.2324	30.6757	27.0919	Anemic
Nujhat	43.1425	30.289	26.5686	Anemic
Noor	42.5314	30.9742	26.4944	Anemic

In Table I. 6 people's data are given out of 104 people with their age, gender, red pixel percentage, Green Pixel percentages, blue pixel percentage along with their hemoglobin value and Anemic condition.

D. *Classification*

The MATLAB software is utilized where Classification Learner automatically built-in. A Microsoft Excel document is framed containing all information with Red, Green, Blue pixel percentage alongside Anemic condition. The excel file imported as CSV (Comma Separated Value) document to classification learner app where the Red, green and blue pixels percentages are taken as features and the anemic condition is taken as a response. An aggregate of 81 data is taken for training. At that point, this model is tested with the rest 23 data. Applying a distinctive method, the model is prepared. The best model is contrasted and the other two. For anticipating anemic condition Linear SVM, Coarse Tree and Cosine KNN are utilized. These calculations are clarified underneath.

1. Support Vector Machine (SVM)

Support Vector Machine is one of the supervised learning models for classification as well as regression [12,13]. Support Vector Machine or SVM is essentially a classifier which is binary. To execute as a multi-classification two techniques named one against all and one against all are utilized for binary. It relies upon C and γ parameter stands for soft margin cost function and γ which leads to low variance and high biasing. So, as to build a settled SVM model, the most ideal pair (C, γ) must be found [14].

2. Decision Tree

The decision tree fabricates regression and classification models as a tree structure. The entire dataset is broken into littler subsets with an expansion top to the bottom of the tree. The conclusive outcome is a tree with decision nodes and leaf nodes. A decision node has two or more branches. The leaf node represents a classification or decision. The topmost decision node in a tree that corresponds to the best predictor is called the root node. It can handle categorical as well as numerical data [15].

3. *KNN (K- Nearest Neighbour)*

K-Nearest Neighbour algorithm alludes as a notable procedure which is ordered items with respect to nearest preparing models on the issues' is essentially sluggish learning in which capacity is evaluated locally as it were. KNN is the most straightforward calculation among all machine learning algorithms. With the more prominent number of votes, an item is characterized. The item is assigned to the class to its nearest neighbour if k=1 [16].

III. Results & Discussion

In classification, the three algorithms such as Coarse Tree, Linear SVM, Cosine KNN. In Table II, the prediction of an Anemic condition is shown by using different algorithms. In Table II, a comparative analysis is depicted where the 23-test data are tested using the three algorithms as mentioned above. where it is seen that the best method is found using Decision Tree with an accuracy of 82.6%. The other two methods have lower accuracy of where 73.91% accuracy is found in both SVM and KNN technique. It is also seen that predicting anemic condition is lower accuracy and predicting no anemic is more effective that effective one. In Fig. 5 confusion matrix is illustrated where 23 test data are predicted. 14 non-anemic data and 5 anemic data are predicted correctly while 2 anemic and 2 non-anemic data are predicted wrong.

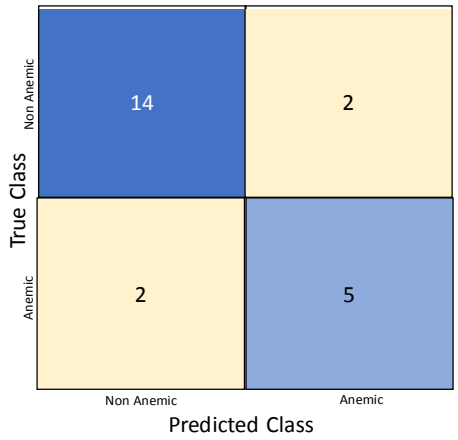

Fig. 5 Confusion matrix of True class vs predicted class

From the above discussion, it can be said that the overall accuracy is found which is better than [5], [7], [17] where accuracy found 80% ,78.9% and 68.53% respectively.

TABLE II: LIST OF PEOPLE WITH THEIR ACTUAL ANEMIC CONDITION AND PREDICTED MODEL USING THREE ALGORITHMS DECISION TREE, SVM AND KNN

Sl.	Name	Gender	Hb (g/dl)	Actual Anemic Condition.	Predicted Anemic Condition (Tree).	Predicted Anemic Condition (SVM).	Predicted Anemic Condition (KNN).
1	Jafor Alam	Male	6.3	Anemic	Anemic	Anemic	Anemic
2	Khadiza1	Female	13.5	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
3	Lalu	Male	11.7	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
5	Mira nath	Female	13.5	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
6	Monoara	Female	12.4	Non-Anemic	Anemic	Anemic	Anemic
7	Munaf	Male	16.2	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
8	Nasima	Female	8.6	Anemic	Anemic	Non- Anemic	Non- Anemic
9	Purnima	Female	10.3	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
10	Rasheda	Female	13	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
11	Rokeya	Female	9.7	Anemic	Non- Anemic	Non- Anemic	Non- Anemic
12	Roksana	Female	12.6	Non-Anemic	Non- Anemic	Non- Anemic	Anemic
13	Roshida	Female	15.4	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
14	Sahidul	Male	14.1	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
15	Shopon	Male	9	Anemic	Anemic	Anemic	Anemic
16	Zidan	Male	14.6	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
17	Shahab	Male	14	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
18	Rehena	Female	10	Anemic	Non- Anemic	Non- Anemic	Non- Anemic
19	Kulsuma	Female	8.3	Anemic	Anemic	Non- Anemic	Anemic
19	Hashina	Female	13.6	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
20	Purnima dey	Female	9.9	Anemic	Anemic	Anemic	Anemic
21	Nur safa	Female	11.6	Non-Anemic	Anemic	Anemic	Anemic
22	Kohinur	Female	12.2	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
23	Khadiza2	Female	12.1	Non-Anemic	Non- Anemic	Non- Anemic	Non- Anemic
Total number of predictions					23	23	23
Correctly predicted					19	17	17
Incorrectly predicted					4	6	6
Percentage of Accuracy predicting anemic condition					71.43%	42.85%	57.14%
Percentage of Accuracy predicting non anemic condition					87.5%	87.5%	81.25%
Overall percentage of accuracy					82.61%	73.91%	73.91%
Overall percentage of Error					17.39%	26.09%	26.09%

Table III illustrates the comparison between recent works based on machine learning and this work where different dataset number along with technique and accuracy is shown.

TABLE III: COMPARISON BETWEEN OTHER WORKS AND THIS WORK

Reference No.	Technique	Data Set	Accuracy
[4]	K-NN, Decision tree, ANN, Naïve-Bayes	500	Overall 92.122%
[18]	Naive-Bayes, random forest, and decision tree	200	96.05% with NB
[19]	Machine learning (Decision tree)	96	Fully matched with physicians
[20]	Decision tree(J48), SVM	514	88.13 % with SVM
[19]	Machine learning (K-NN, LDA, CART, SVM, LR, RF	600	Highest 68.53% random forest (RF)
This work	Decision Tree, SVM, KNN	104(84 train data 23 test data)	Accuracy of 82.61%

IV. Conclusion

In the entire work, a non-intrusive technique for an anemic condition forecast is effectively created where a comparison between three algorithms is done. By utilizing the best calculation anemia can be anticipated inside a brief time frame alongside easily and cost-viably. As a rule, to identify anemia one needs to conduct CBC (Complete Blood Count) test where Hemoglobin level is represented. That cost by and large USD 5 (400 BDT). In this non-intrusive procedure, the sickly condition can be effectively anticipated without leading a CBC test and with no expense. The destitute individuals and individuals living in a remote region can undoubtedly identify anemia utilizing this calculation where just the eye conjunctiva picture is taken, from that anemic condition is anticipated adequately. In this work, we have about 82.61% which is not exactly the clinical test. This work has anticipated 23 patients' anemic condition and 4 forecasts aren't right which is the shortcoming of this algorithm. In our dataset, we have incorporated a portion of the information which are abnormal and exceptional. About 10 to 15 patients are influenced by dengue fever. The eye conjunctiva becomes reddish and has an extremely awful relationship with the anemic condition and which drives our outcome to less precise.

Acknowledgment

The authors would like to thank the authority, doctors and intern doctors of Chittagong Medical College Hospital and Cox's Bazar Medical College Hospital for their support to collect data. The authors also would like to thank the Department of Electrical and Electronic Engineering of Chittagong University of Engineering and Technology for the help in this research work.

References

[1] WHO (world health organization) statistics https://www.who.int/topics/anaemia/en/, 16 September 2019.

[2] World Health Organization "Nutritional Anemias. Report of a WHO scientific group," World Heal Organ—Tech Rep Ser, 405:5-37, 1968.

[3] Prevalence of anemia, "Department of Nutrition for Health and Development, World Health Organization., Geneva, Switzerland. WHO Vitamin and Mineral Nutrition Information System, Public Health Nutr. April 2009.World Health Organization. Worldwide prevalence of anemia. WHO Rep. 51, 2005.

[4] Pooja Tukaram and Nagaraj Verekar, "Anemia Detection using ensemble learning technique and statistical models". *IEEE International Conference on recent trends in electronics information communication technology, may 20-21,2016,India*

[5] Magudeeswaran Veluchamy, Karthikeyan Perumal and Thirumurugan Ponuchamy, "Feature Extraction and Classification of Blood Cells Using Artificial Neural Network".

[6] K S Srinivasan, D. Lakshm, H Ranganathan, and N Gunasekaran, "Non-Invasive Estimation Of hemoglobin in Blood Using Color Analysis" *first international Conference on Industrial and Information Systems, ICIIS 26, Srilanka.*

[7] Azwad Tamir, Chowdhury S. Jaha, Mohammad S. Saif, Sums U. Zaman, Md. Mazharul Islam, Asir Intisar Khan, Shaikh Anowarul Fattah, and Celia Shahnaz," Detection of Anemia from Image of the Anterior Conjunctiva of the Eye by Image Processing and Thresholding", *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) 21 - 23 Dec 2017, Dhaka, Bangladesh.*

[8] Giovanni Dimauro,Danilo Caivano and Francesco Girardi, "A new method and a non-invasive device to estimate anaemia based on digital images of the conjunctiva," *Digital Object Identifier,* vol. 6, pp 46968-46975,10.1109/ACCESS.2018.2867110

[9] Mahammad Firose Shaik and Monica Subashini M, "Anemia Diagnosis by Fuzzy Logic Using LabVIEW," *International Conference on Intelligent Computing and Control (I2C2),*2017.

[10] Robert G. Mannino, David R. Myers, Erika A.Tyburski, Christina Caruso, Jeanne Boudreaux, Traci Leong, G. D. Clifford, and Wilbur A. Lam, "Smartphone app for non-invasive detection of anemia using only patient-sourced photos," *Nature Communications,* vol. 9, Article number: 4924 (2018).

[11] S. Suner, G. Crawford, J. McCurdy, and G. Jay," Non-invasive determination of hemoglobin by digital photography of palpebral conjunctiva", J. Emerg, Med,33 (2)(2007): 10

[12] Alex J. Smola and Bernhard Scholkopf," A tutorial on support vector regression", Statistics and Computing, Springer, 2004.

[13] SR Gunn, "Support Vector Machines for Classification and Regression", ISIS technical report, 1998.

[14] Marko Lukic, Eva Tuba and Milan,"Leaf Recognition Algorithm using Support Vector Machine with Hu Moments and Local Binary Patterns", SAMI 2017 *IEEE 15th International Symposium on Applied Machine Intelligence and Informatics*, January 26-28, 2017 • Herl'any, Slovakia.

[15] https://medium.com/@rishabhjain_22692/decision-trees-it-begins-here-93ff54ef134 , 15 September 2019

[16] Lloyd-Williams, M." Case studies in the data mining approach to health information analysis", Knowledge Discovery and Data Mining, 1998/434), IEEEColloquium on, 8 May 1998, 1996 Page(s): 1/1 -1/4

[17] Jahidur Rahman Khan, Srizan Chowdhury, Humayera Islam and Enayetur Raheem, "MACHINE LEARNING ALGORITHMS TO PREDICT THE CHILDHOOD ANEMIA IN BANGLADESH," *Journal of Data Science*, vol. 16, pp. 195-218, 2019

[18] Manish Jaiswal , Anima Srivastava and Tanveer J. Siddiqui, "Machine Learning Algorithms for Anemia disease Prediction," *IEEE Conference on e-Learning, e-Managment and e-Services*, 2018

[19] Dogan, S., & Turkoglu, I. "Iron deficiency anemia detection from hematology parameters by using decision tree," *International Journal of Science and Technology,* vol. 03, no. 01, pp. 85–92, 2008

[20] Shilpa, S. A., Nagori, M., & Kshirsaga, "Classification of anemia using data mining techniques," In *Swarm, evolutionary, and memetic computing* (pp. 113–121), 2011