JES
Journal of Engineering Sciences

Vol 13, Issue 03, MARCH /2022

ISSN NO:0377-9254

Image Classification of Abnormal Red Blood Cells Using Random Forest Algorithm

Ch. Sai Lakshmi Sadhwika[1], Navya Sree Anagani[2], G. Ramya Devi[3], P. SaiPrakash[4], M. Madhu Gowtham[5]

#1#3#4#5 Student , Dept Of CSE, Gandhi Institute of Technology and Management College, Visakhapatnam, India

#2 Assistant Professor, Dept Of CSE, Gandhi Institute of Technology and Management College, Visakhapatnam, India

Abstract: The study aims to increase the number of abnormal red blood cells that can be detected using image processing. This study used Random Forest Algorithm as a machine learning algorithm in classifying. As a result, the system detected and classified a total of ten abnormal red blood cells. Images used in the system came from hospitals' past patients. In addition, a camera is used to capture the slides. The image was then inserted into the program. The system processed and classified the image. In effect, the results show the name of the abnormal red blood cells detected in the image within the system including the soft copy of the list.

Keywords: Random Forest Algorithm, Abnormal Red Blood Cells, Classifying

1. INTRODUCTION:

Blood is the life-maintaining fluid that flows through the heart, arteries, and other organs and it circulates through the whole body. Moreover, a blood is made up of many parts, mostly the red blood cells, white blood cells, platelets and plasma. Abnormalities of red blood cells vary through size or anisocytosis, through shape or poikilocytosis, in color and even through the presence of inclusion bodies. Detecting these irregularities in the shapes of the red blood cells is significant to one's health as it can determine whether the blood is healthy or not. Medical technicians, Pathologists and Hematologists usually used a manual microscopic method, to classify abnormal shapes of red blood cells. This methodology somehow is difficult and prone to human error. Thus, classifying the abnormal red blood cells using image processing is created using the high technologies.

Mohammad Syahputra Et. Al (2017) said that morphological examination of peripheral blood smears done manually is less efficient and the shapes of the abnormal red blood cells found is not always the same for every analyst because of precision factor, concentration, and lack of knowledge. Radial Basis Function Network was used together with several stages which includes input image, pre-processing, and feature extraction. Research result shows that by using this method, the accuracy to classify abnormal red blood cell types is 83.3%.

The objective of this study is to create a system that can classify 10 abnormal red blood cells and to know the reliability rate of classification of each abnormal red blood cells. Previous studies are usually limited to two to four abnormal red blood cell. Thus, the proponents aimed to create a maximized system.

2.LITERATURE SURVEY

2.1 **Pooja Tukaram Dalvi and Nagaraj Vernekar (2016). Computer Aided Detection of Abnormal Red Blood Cells.**

Red blood cell classification and counting plays a very important role in detecting diseases like iron deficiency anemia, vitamin B12 deficiency anemia etc. In this research we intend to develop a standalone application that can classify the red blood cells into four abnormal types namely elliptocytes, echinocytes, tear drop cells and macrocytes. We will also provide the total red blood cell count .Thirteen Geometric features have been used to classify the red blood cells into the four abnormal types. We have used two data mining classifiers namely Artificial Neural Network and Decision Tree Classifier and we have compared the results of the two classifiers with respect to accuracy in classifying the red blood cells.The proposed method exhibits an accuracy of 95.27% for detecting elliptocytes, 96.06% for echinocytes, 85.82% for tear drop cells 85.82% for macrocytes and 89.76% for normal red blood cells. Red blood cells (RBCs) are the most abundant cells present in human body. Normal RBCs are biconcave and disk shaped. Any abnormality in the shape of RBC indicates presence of disease. The number of RBCs also plays an important role in detecting anemia. A decrease in the number of RBCs and an abnormality in RBC's shape is a clear indicator of presence of blood related disorders. Presence of tear drop cells, echinocytes, elliptocytes, macrocytes indicate presence of diseases like myelofibrosis ,severe iron deficiency ,uremia hereditary elliptocytosis ,haemolytic anaemia etc[14]. Anemia and blood related disorders are prevalent in almost 24.5% of world population.

2.2 Vishwas Sharma Et. Al (2010). Detection of Sickle Cell Anaemia and Thalassaemia Causing Abnormalities in Thin Smear of Human Blood Sample Using Image Processing.

Blood is a connective tissue in which Red blood cells function to transport oxygen and it is normally in disk shape. The inherited disorder of blood includes hemoglobin opathies which are major public health problem in India. Sickle cell disease refers to a group of genetic disorders characterized by presence of sickle hemoglobin, anemia, acute and chronic tissue injury to blockage of blood flow by abnormally shaped red cell. Sickle cell disease is Sickle cell anemia. It is a disorder in which the body makes sickle-shaped red blood cells. "Sickle-shaped means the red blood cells are crescent shaped. Sickle cell anemia is also a serious disorder problem in chhattisgarh state. It is highly prevalent among scheduled caste, scheduled tribe and other back-word class. In Chhattisgarh the highest percentage of sickle cell diseases found in Sahu, Mahar, Gond, Devangan, Kurmi and Halba etc. This paper proposed a method to recognize the sickle shaped red blood cells present in the blood smear by using fractal dimension. Fractal Dimension is used to recognize the shape of the red blood cells and segmentation the sickle shaped red blood cells for shape analysis to find the percentage of sickle cell anemia. Results exhibit the future aspect of the technique, which overcome traditional shape recognition and analysis methods found in various literatures. ickle cell disease (SCD), also known as sickle cell anemia, is a serious disease in which the body makes an altered form of hemoglobin, the protein in red blood cells that carries oxygen throughout the body. This genetic alteration causes the body to produce abnormal sickle- or crescent-shaped red blood cells.

3.PROPOSED SYSTEM

The proponents proposed a system that automatically classifies 3 abnormal red blood cells using the Random Forest Algorithm. The algorithm is commonly used in classification and regression analysis. A decision tree is a simple representation for classifying example that splits on its nodes. It uses question on an attribute and splits the node that results to a branch or to end up with an output. The idea of

the whole system of the decision tree is to divide the data set into smaller data set based on the descriptive features or the attributes until it reaches a specific abnormal red blood cell.

3.1 IMPLEMENTATION

1. **Data Collection:**Collect sufficient data samples and legitimate software samples.
2. **Feature Extraction**:For each image extract the features ubsing image processing and save in '.csv' extension
3. **Train and Test Modelling:** Split the data into train and test data Train will be used for trainging the model and Test data to check the performace

3.2 ALGORITHM

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** *which is a process of* combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, **"Random Forest is a classifier that contains a number of decision** trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that **dataset."** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

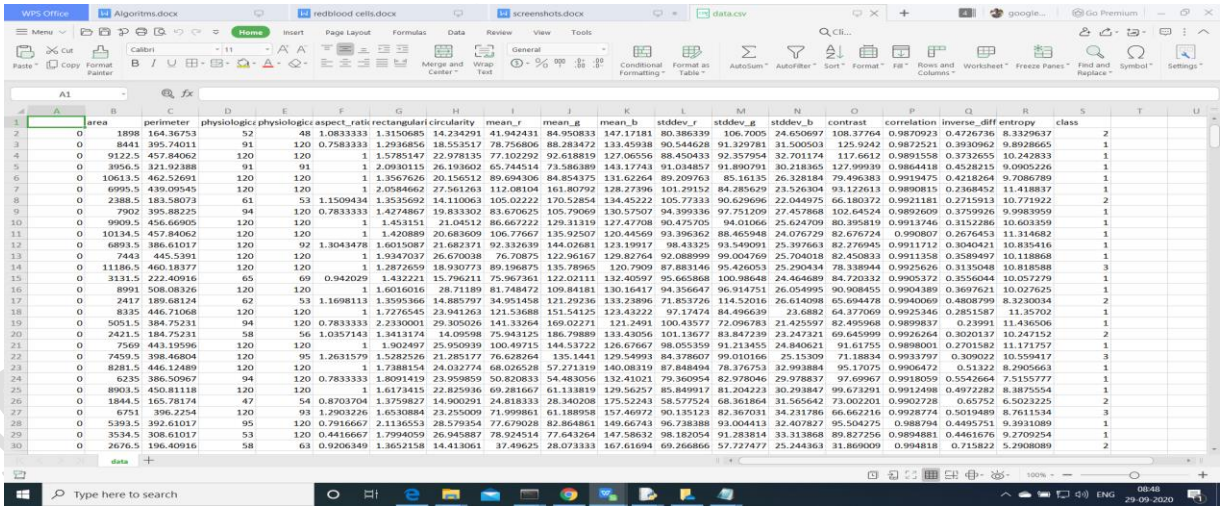
4. DATASET DESCRIPTION

Data Description:

White blood cell data collected form kaggle

For this classification we take datasets. This dataset consists of nineteen columns and 401 records. Out of ninteen columns 18 columns are consists attributes and last column consists of class labels.

Contains 4 categories



	area	perimeter	physiologica	physiologica	aspect_ratio	rectangularity	circularity	mean_r	mean_g	mean_b	stddev_r	stddev_g	stddev_b	contrast	correlation	inverse_diff	entropy	class
2	0	1898	164.36753	52	48	1.0833333	1.3150685	14.234291	41.942431	84.950833	147.17181	80.386339	106.7005	24.650697	108.37764	0.9870923	0.4726736	8.3329637	2
3	0	8441	395.74011	91	120	0.7583333	1.2936856	18.553517	78.756806	88.283472	133.45938	90.544628	91.329781	31.500503	125.9242	0.9872521	0.3930962	9.8928665	1
4	0	9122.5	457.84062	120	120	1	1.5785147	22.978135	77.102292	92.618819	127.06556	88.450433	92.357954	32.701174	117.6612	0.9891558	0.3732655	10.242833	1
5	0	3956.5	321.92388	91	91	1	2.0930115	26.193602	65.744514	73.586389	143.11743	91.034857	91.890791	30.218365	127.99939	0.9864418	0.4528215	9.0905226	1
6	0	10613.5	462.52691	120	120	1	1.3567626	20.156512	89.694306	84.854375	131.62264	89.209763	85.16135	26.328184	79.496383	0.9919475	0.4218264	9.7086789	1
7	0	6995.5	439.09545	120	120	1	2.0584662	27.561263	112.08104	161.80792	128.27396	101.29152	84.285629	23.526304	93.122613	0.9890815	0.2368452	11.418837	1
8	0	2388.5	183.58073	61	53	1.1509434	1.3535692	14.110063	105.02222	170.52854	134.45222	105.77333	90.629696	22.044975	66.180372	0.9921181	0.2715913	10.771922	2
9	0	7902	395.88225	94	120	0.7833333	1.4274867	19.833302	83.670625	105.79069	130.57507	94.399336	97.751209	27.457868	102.64524	0.9892609	0.3759926	9.9983959	1
10	0	9909.5	456.66905	120	120	1	1.453151	21.04512	86.667222	129.31319	127.47708	90.475705	94.01066	25.624709	80.395819	0.9913746	0.3152286	10.603359	1
11	0	10134.5	457.84062	120	120	1	1.420889	20.683609	106.77667	135.92507	120.44569	93.396362	88.465948	24.076729	82.676724	0.990807	0.2676453	11.314682	1
12	0	6893.5	386.61017	120	92	1.3043478	1.6015087	21.682371	92.332639	144.02681	123.19917	98.43325	93.549091	25.397663	82.276945	0.9911712	0.3040421	10.835416	1
13	0	7443	445.5391	120	120	1	1.9347037	26.670038	76.70875	122.96167	129.82764	92.088999	99.004769	25.704018	82.450833	0.9911358	0.3589497	10.118868	1
14	0	11186.5	460.18377	120	120	1	1.2872659	18.930773	89.196875	135.78965	120.7909	87.883146	95.426053	25.290434	78.338944	0.9925626	0.3135048	10.818588	3
15	0	3131.5	222.40916	65	69	0.942029	1.432221	15.796211	75.967361	122.02111	132.40597	95.665868	100.98648	24.464689	84.720332	0.9905372	0.3556044	10.057279	1
16	0	8991	508.08326	120	120	1	1.6016016	28.71189	81.748472	109.84181	130.16417	94.356647	96.914751	26.054995	90.908455	0.9904389	0.3697621	10.027625	1
17	0	2417	189.68124	62	53	1.1698113	1.3595366	14.885797	34.951458	121.29236	133.23896	71.853726	114.52016	26.614098	65.694478	0.9940069	0.4808799	8.3230034	2
18	0	8335	446.71068	120	120	1	1.7276545	23.941263	121.53688	151.54125	123.43222	97.17474	84.496639	23.6882	64.377069	0.9925346	0.2851587	11.35702	1
19	0	5051.5	384.75231	94	120	0.7833333	2.2330001	29.305026	141.33264	169.02271	121.2491	100.43577	72.096783	21.425597	82.495968	0.9899837	0.23991	11.436506	1
20	0	2421.5	184.75231	58	56	1.0357143	1.3413174	14.09598	75.943125	186.79889	133.43056	101.13677	83.847239	23.247321	69.645999	0.9926264	0.3020137	10.247152	2
21	0	7569	443.19596	120	120	1	1.902497	25.950939	100.49715	144.53722	126.67667	98.055359	91.213455	24.840621	91.61755	0.9898001	0.2701582	11.171757	1
22	0	7459.5	398.46804	120	95	1.2631579	1.5282526	21.285177	76.628264	135.1441	129.54993	84.378607	99.010166	25.15309	71.18834	0.9933797	0.309022	10.559417	3
23	0	8281.5	446.12489	120	120	1	1.7388154	24.032774	68.026528	57.271319	140.08319	87.848494	78.376753	32.993884	95.17075	0.9906472	0.51322	8.2905663	1
24	0	6235	386.50967	94	120	0.7833333	1.8091419	23.959859	50.820833	54.483056	132.41021	79.360954	82.978046	29.978837	97.69967	0.9918059	0.5542664	7.5155777	1
25	0	8903.5	450.81118	120	120	1	1.6173415	22.825936	69.281667	61.133819	129.56257	85.849917	81.204223	30.293847	99.673291	0.9912498	0.4972282	8.3875554	1
26	0	1844.5	165.78174	47	54	0.8703704	1.3759827	14.900291	24.818333	28.340208	175.52243	58.577524	68.361864	31.565642	73.002201	0.9902728	0.65752	6.5023225	2
27	0	6751	396.2254	120	93	1.2903226	1.6530884	23.255009	71.999861	61.188958	157.46972	90.135123	82.367031	34.231786	66.662216	0.9928774	0.5019489	8.7611534	3
28	0	5393.5	392.61017	95	120	0.7916667	2.1136553	28.579354	77.679028	82.864861	149.66743	96.738388	93.004413	32.407827	95.504275	0.988794	0.4495751	9.3931089	1
29	0	3534.5	308.61017	53	120	0.4416667	1.7994059	26.945887	78.924514	77.643264	147.58632	98.182054	91.283814	33.313868	89.827256	0.9894881	0.4461676	9.2709254	1
30	0	2676.5	196.40916	58	63	0.9206349	1.3652158	14.413061	37.49625	28.073333	167.61694	69.266866	57.727477	25.244363	31.869009	0.994818	0.715822	5.2908089	2

Fig 1:Dataset Values

4. RESULTS AND DISCUSSIONS

Journal of Engineering Sciences

Vol 13, Issue 03, MARCH /2022

ISSN NO:0377-9254

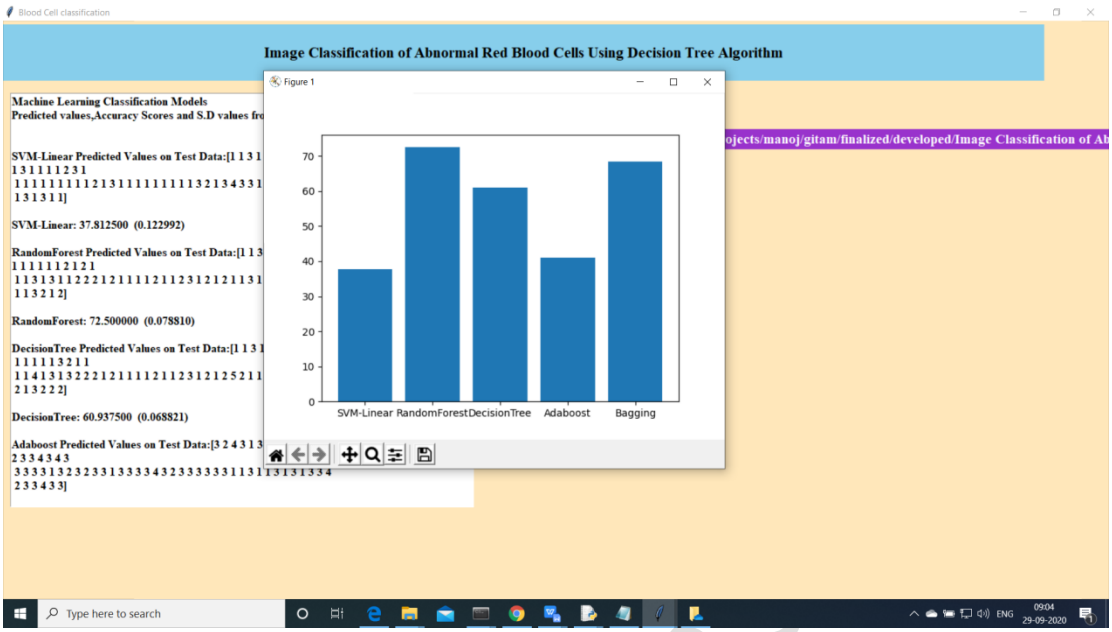


Fig 2: Accuracy Comparision for all the models

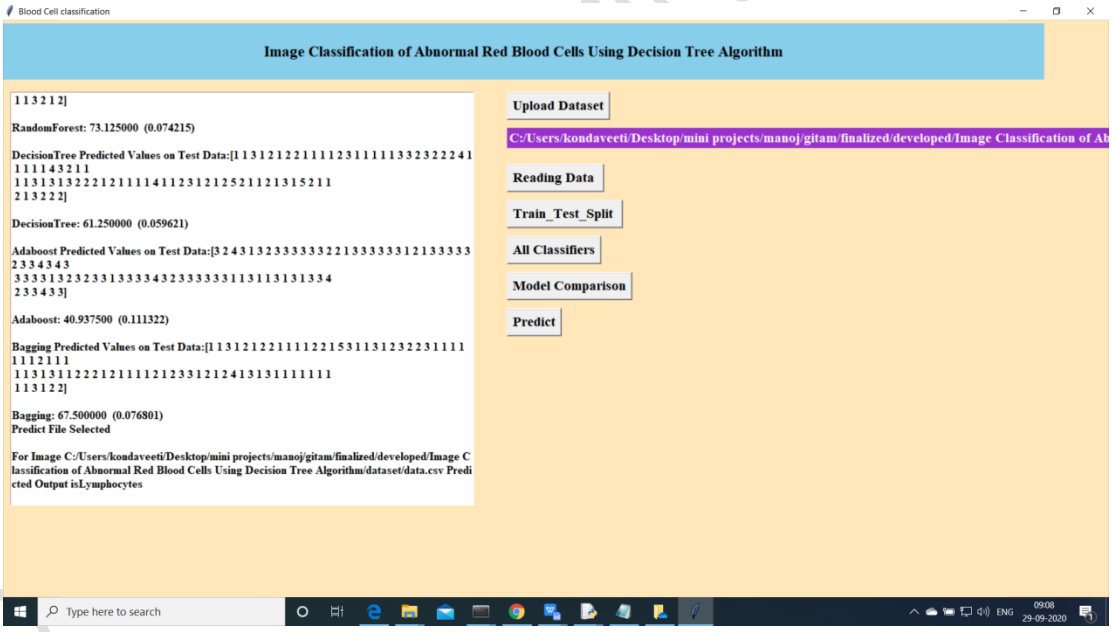


Fig 3: Predict Results

Extension Random Forest salgorithm is performed well compared other ml algortithm

s

5. CONCLUSION

Random Forest can be used as a classification for the unusual pink blood cells observed in the blood. The device used to be in a position to classify the peculiar crimson blood cells the use of Random Forest Algorithm primarily based on the facts gathered from the forty snap shots that have been composed of 600 pattern cells. Errors in classifications have been end result of small variations between the attributes used. Abnormal purple blood cells like elliptocytes and ovalocytes nearly have the identical parameters and attributes ensuing to challenge in classifying the two odd pink blood cells.

The common reliability fee is 72.18%. The common error price of 10.69% was once encountered generally from node H.

REFERENCES

[1] Mohammad Syahputra Et. Al (2017). Abnormal Classification on the Shape of Red Blood Cells Using Radial Basis Function Network.

[2] Pooja Tukaram Dalvi and Nagaraj Vernekar (2016). Computer Aided Detection of Abnormal Red Blood Cells.

[3] Vishwas Sharma Et. Al (2010). Detection of Sickle Cell Anaemia and Thalassaemia Causing Abnormalities in Thin Smear of Human Blood Sample Using Image Processing.

[4] Ramin Soltanzadeh and Hossein Rabbani (2018). Classification of Three Types of Red Blood Cells in Peripheral Blood Smear Based on Morphology.

[5] Jameela Ali Akrimi ET. Al (2014). Classification Red Blood Cells Using Support Vector Machine.

[6] Pranati Rakshit, Et. Al (2013). Detection of Abnormal Findingss in Human RBC in Diagnosing G-6-P-D Deficiency Haemolytic Anaemia Using Image Processing.

[7] Ahmad Sabry Mohamad ET. Al (2017). Sickle Cell Disease Verification via Sobel Edge Algorithms for Image Processing.

Author's Profile

Name: Ch. Sai Lakshmi Sadhwika
College name: Gandhi Institute of Technology and ManagementCollege address: Visakhapatnam, India
Mail id: sadhwikachavatapalli@gmail.com
B.tech Branch: Computer Science and Engineering

Vol 13, Issue 03, MARCH /2022

ISSN NO:0377-9254

Name: Navya Sree Anagani

Designation: Assistant Professor

College name: Gandhi Institute of Technology and Management College address: Visakhapatnam, India

Mail id: nanagani@gitam.edu

Name: G. Ramya Devi
College name: Gandhi Institute of Technology and ManagementCollege address: Visakhapatnam, India
Mail id: ramyavarma4567@gmail.com
B.tech Branch: Computer Science and Engineering

Name: P. SaiPrakash
College name: Gandhi Institute of Technology and ManagementCollege address: Visakhapatnam, India

Mail id: saiprakashpavuluri12426@gmail.com

B.tech Branch: Computer Science and Engineering

Name: M. Madhu Gowtham
College name: Gandhi Institute of Technology and ManagementCollege address: Visakhapatnam, India

Mail id: gowthamnaidupspk3@gmail.com

B.tech Branch: Computer Science and Engineering