IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Cardio Vascular Anemia Classifier using Supervised Machine Learning Techniques

K. Srinivasa Reddy[1] and K..Vasantha Rao[2] and N.Durga Dheeraj[3] and Dr. B. Prabha[4]
Students, Department of Computer Science and Engineering[1,2,3]
Assistant Professor, Department of Computer Science and Engineering[4]
SRM Institute of Science and Technology, Vadapalani, Chennai, India

Abstract: *Anemia, a condition brought on by a lack of iron, is one of the most severe general health problems. Non-industrialized nations experience higher than normal rates of anemia. To evaluate the frequency of anemia, a blood test known as the total blood count is carried out. Anemia reduces the amount of oxygen that the blood can transport. Anemia occurs when there is either not enough hemoglobin present or when the hemoglobin is worthless. When blood circulates through the body, including the lungs, oxygen binds to hemoglobin and goes to the tissues. Hypoxia of the tissues happens when there is insufficient hemoglobin to properly oxygenate them.The overall amount of RBCs, In anaemia, both the amount of haemoglobin in the blood and the proportion of RBCs fall. RBCs, hemoglobin, hematocrit, RBC indices, and RDW are all calculated or measured using equipment. Using univariate, bivariate, and multivariate analysis, the machine learning-based technique (SMLT) will offer a machine learning-based way for predicting whether or not the disease has already struck. The metrics of machine learning can be compared to the best in the industry.*

Heart disease ,early detection of the condition can significantly enhance patient outcomes. Taking advantage of the vast amount of data accessible for analysis, Machine learning algorithms have developed into powerful tools for predicting heart illness. In this review, we analyse the present status of the subject and give an overview of recent research .We also discuss recent research that has used ML for heart disease prediction, the use of deep learning algorithms for automated identification of heart illness from ECG data and the use of SVM and ANN algorithms for cardiovascular event prediction

Keywords: Anemia, Machine Learning, Classification.

I. INTRODUCTION

Haemoglobin, a protein transported by red blood cells that is high in iron, binds to oxygen in the lungs and transports it to bodily tissues. When red blood cells are insufficient or are not Functioning correctly, you develop anemia. Children's normal values change with growth. Since different types of anemia require various treatment modalities, classification of anemia is essential for its successful management. Conventional treatments for anemia and a number of cutting-edge machine learning algorithms were used to conduct operations on our dataset; the classification process involved the manual review of blood demographic data. In terms of accuracy, precision, and recall, our approach works better than the ones currently in use.Machine learning techniques have emerged as a promising way for diagnosis and more effective than conventional approaches.In this paper, a brand-new method for categorizing anemia using machine learning methods is proposed. We trained a classifying anemia model using a dataset of blood parameters and demographic data, always producing reliable results.Use of in recent years. Large datasets of blood parameters and demographic data can be analyzed by machine learning algorithms to classify We contrasted the results. This method could be used by medical experts to more precisely identify the anemia type and offer more focused and efficient treatments. Overall, our research offers a promising route for applying machine learning methods to the classification of anemia.

Heart disease is a serious health problem and the top cause of mortality around the world. The ability to diagnose heart disease early and accurately is crucial for

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

successful treatment and prevention of cardiovascular events. With recent advancements in machine learning (ML) techniques, there is a growing interest in using ML algorithms for cardiac disease detection and prediction.

To forecast the chance of a patient developing heart disease, ML systems can analyse massive datasets and discover patterns and correlations between diverse characteristics. These algorithms can also assist in identifying the most critical risk factors for heart disease and providing personalised therapy suggestions based on a patient's unique features.

II. RELATED WORK

[1].Classifying anemia types using artificial learning methods

Nilüfer Yurtay, Birgül Öneç,Tuba Karagül Yıldız Anaemia is a relatively frequent condition that negatively impacts quality of life, but with the right care, the patient's situation will improve. It goes without saying that getting the right diagnosis is the first step in getting better. In this work, ANNs, SVMs, naive Bayes, and ensemble decision trees—are used to categorise the 12 forms of anaemia that are most frequently seen in the Düzce Province. Nine distinct datasets are produced as a result of the feature selection procedure. Consequently, our dataset's initial collection of 25 properties is expanded to additional datasets that include attribute values. Here, our goal is to use the doctor's own data to determine the doctor's decisions. Consequently, the datasets also contain the original dataset. We compared the outcomes of all algorithms after running them on the new datasets as well. metrics are used to evaluate the results.

[2].Deep Learning Models for Classification of Red Blood Cells in Microscopy Images to Aid in Sickle Cell Anemia Diagnosis.

Ye Duan,Jinglan Zhang,,Laith Alzubaidi,Omran Al-Shamma,Mohammed A. Fadhel. Anaemia can result from a variety of factors, including the creation of RBCs with aberrant shapes, water retention during pregnancy, a decrease in erythropoietin production, insufficient iron consumption, and menstrual blood loss. Additionally, sickle cell anaemia (SCA) is the term used to describe a condition brought on by the presence of RBCs with a sickle shape. These cells are produced as a result of the haemoglobin gene changing. SCA is a hereditary disease because if both parents have two faulty genes, those genes are passed on to the child, who will subsequently develop the syndrome. Haemoglobin S is primarily responsible for the diversity in RBC shape.These cells will become caught in the blood arteries as a result of this irregular shape variation that decreases the oxygen transport. Additionally, these cells have a propensity to fragment . The countries most affected by SCA are South Africa, Italy, Greece, Turkey, and India, according to published data and statistics .

[3].Machine Learning Algorithms for Anemia Disease Prediction. Anima Srivastava,Siddiqui,Manish Jaiswal,Tanveer J.In a variety of fields, including healthcare, stock price prediction, and product suggestion, The forecasting of different diseases and the factors that contribute to them is a crucial component of medical science study. Healthcare data used in the medical field to diagnose illness, anticipate epidemics, quality of life, and prevent premature death. For its prediction, we look into three different classification techniques in this paper. Fatigue and reduced productivity are symptoms of anaemia, which can also raise the risk of maternal and neonatal death if it develops during pregnancy. The World Health Organisation (WHO) estimates that 3.0 million fatalities in poor nations occurred in 2013 as a result of maternal and neonatal mortality.The ability to forecast anaemia disease is crucial for identifying other linked disorders. Anaemia disease is categorised according to its morphology or underlying aetiology. Anaemia is classified into three groups based on morphology: normocytic, microcytic, and macrocytic. Anaemia can have three different causes: blood loss, insufficient synthesis of healthy blood, and excessive blood cell apoptosis.

[4].Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting.Jankisharan Pahareeya,Abir Hussain,,Rajan VohraI,Anil Kumar DudyalaThe statistical significance of a relationship between categorical variables was examined using the Chi square test. In reality, this study paper expands on that work by using the data set that was gathered as a base and applying classification algorithms to categorise the identified patients as Mild, Moderate, and Severe. Numerous factors, including socioeconomic, demographic, behavioural, and others connected to the health care system, have an impact on public health. Social factors including wealth, income, and education can have an impact on a person's body mass index,waist size, and other health indicators.Several data mining approaches and data sets have been used to forecast the frequency of malaria and anaemia in children. Rosangela et al.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

used regression analysis in their study. These factors include age, food, affluence, parental education status, the child's surroundings, and past illnesses.

[5].Machine learning based Diagnosis and Classification Of Sickle Cell Anemia in Human RBC.Anupama Bhan,Adarsh Ganesh , Ayush Goyal,Bheem Sen,Shubhra Dixit An inadequate number of blood cells causes anaemia, the most common blood ailment, which stops the body from getting enough oxygen. Chronic anaemia results from a steady fall in red blood cells and is frequently associated with inflammatory illnesses. RBC production is not as it should be in people with sickle cell disease. RBCs often resemble disc-shaped objects. However, with sickle cell disease, it resembles a crescent moon or an old farming implement called a sickle. It is a hereditary RBC disorder in which the body lacks enough healthy RBCs to deliver oxygen. Typically, SCA symptoms and signs first appear around five months of age. Sickle cells quickly disintegrated and died, leaving just a small number of RBC remained. Sickle cells frequently die within two to three weeks, creating an RBC deficiency. RBC generally lasts for around four months before it has to be replenished.

[6].Heart disease prediction using machine learning techniques Bhartendu Sharma ,Apurv Garg,Rijwan Khan Machine Learning (ML) has had a significant positive impact on the scientific community. This study uses machine learning . Cardiovascular diseases (CVDs) are a widespread condition that can be fatal and affect many people. By considering a number of factors, including age, cholesterol levels, and other factors, as well as chest pain, these can be used to determine whether a person has a cardiovascular disease.Cardiovascular diseases can be more accurately diagnosed using classification techniques that incorporate supervised learning, a type of machine learning. Examples of supervised machine learning algorithms used in this work include Random Forest and K-Nearest Neighbour. In comparison to K-Nearest Neighbour, Random Forest has a prediction accuracy of 81.967% as opposed to 86.885%.

[7].A novel approach for heart disease prediction using strength scores with significant predictors Yin Kia Chiam, Wan Azman Wan Ahmad,Armin Yazdani,Kasturi Dewi Varathan,Asad Waqar MalikThis study contributed to the highest level of confidence when predicting heart disease using key WARM components. It has been demonstrated that giving weight ratings that are appropriate improves the performance of the prediction's confidence level. Heart disease was predicted using a set of key characteristics with weights that varied according to how severe each feature was.The method to measure weight can be further investigated for use with various datasets to enable more weighted prediction models.Future research should examine available machine learning methods for identifying the crucial traits.

[8].Heart Disease Prediction Using Machine Learning Heba Yahia Youssef, Chaimaa Boukhatem,Ali Bou Nassif Cardiovascular disease refers to any severe cardiac condition. Researchers are concentrating on developing smart systems that precisely identify heart conditions based on electronic health data and using machine learning algorithms because heart conditions can be fatal. This paper provides machine learning methods for predicting heart disease that leverage patient data on important health metrics. The study demonstrated four classification techniques in order to build prediction models.Processes for feature selection and data preprocessing were used before the models were created. The models were evaluated using the criteria of precision, recall, recall accuracy, and F1-score. The SVM model's accuracy score of 91.67% was the highest

[9].Detecting Clinical Signs of Anaemia From Digital Images of the Palpebral Conjunctiva Attilio Guarini,Giovanni Dimauro, Angela Iacobazzi,Francesco Girardi,Crescenza Pasciolla,Danilo Caivano.This work is a component of a larger investigation of a non-invasive technique for detecting anemia risk. The palpebral conjunctiva is the exposed tissue that the method used in this research uses to identify clinical indicators of anaemia from digital photographs. The new acquisition tool , which substantially facilitates digital image acquisition for this analysis, is affordable and easy to use. In reality, It was found that the fundamental issue with the acquisitions was the lack of ambient light., rendering unnecessary any preventative standardisation of the photos. Images can now be abstracted from pointless details by applying colour clustering. The findings collected validate the system's dependability and its suitability for use as a personal monitoring tool.

[10].Anemia Prediction Based on Rule Classification Amjad A. Ahmed,Sahar J. Mohammed MOHAMMED,Mohammed Sami MOHAMMED,Arshed A. Ahmad. Anaemia and other disorders like it may be prevented by making sure the body doesn't lack iron or vitamins. If not diagnosed at an early stage, anaemia can result in serious health difficulties like pregnancy troubles or even cardiac problems. 539 participants' data, with 10 relevant attributes for each, were gathered for this study. To create a curated anaemia prediction system, three based rule

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

classification approaches are used: ZeroR, OneR, and PART to collect pertinent anaemia datasets connected with "If" and "Then" procedures. In terms of the strategies used, PART offered 85% more accuracy than ZeroR and OneR. These methods provide a standard for other methods that were used to explain the necessary understanding of anemia data standards.

III. METHODOLOGY

A generalised dataset would be created by combining datasets from several sources. Before being loaded, the supplied dataset will be cleaned and trimmed. It will then be checked for accuracy before being utilised for analysis in this portion of the report. Two setswere constructed using the data collected. It is common practice to divide the dataset in a 7:3 ratio. The Training set receives the application of the DataModel produced by machine learning algorithms, and the prediction of the Test set is dependent on the accuracy of the test results. The great preprocessing capability of the ML prediction model enables it to forecast the presence or absence of the anaemia illness.

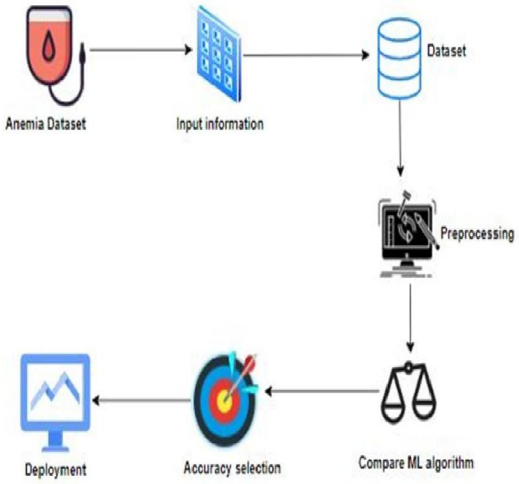

Fig 1.1 - System Architecture

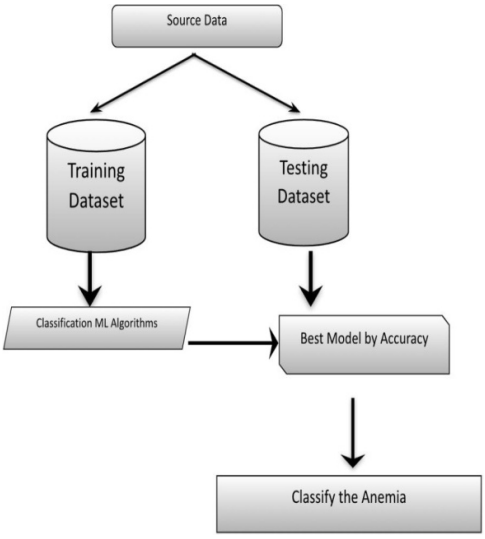

Fig 1.2 - Training and testing

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

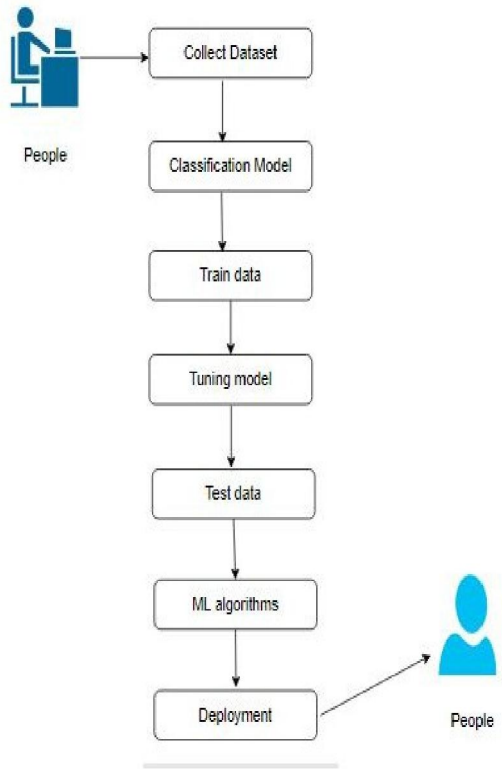

Fig 1.3 -WorkFlow

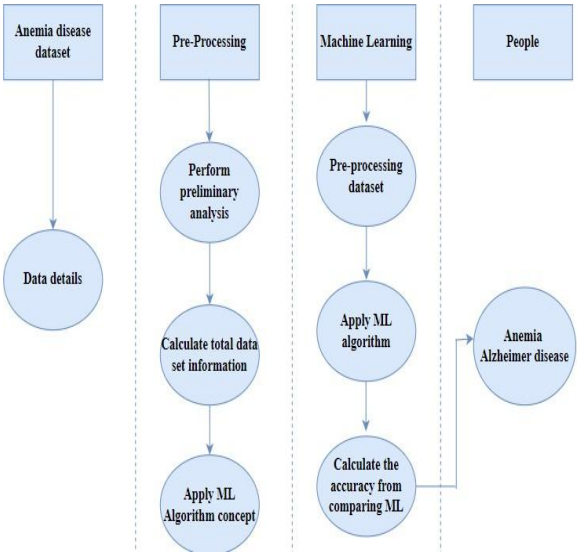

Fig1.4- Activity Diagram

The specifics of the image, the pre-processing of the data, the first analysis, and the algorithms are shown in the figure below.

The forward and reverse engineering processes used to build the executable system are depicted in the activity diagram.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

Preparing the dataset:

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Gender	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	MCV	Hemoglob	MCH	MCHC	target
2	1	0	125	212	0	1	168	0	1	2	2	3	71.2	9	21.5	29.6	0
3	1	0	140	203	1	0	155	1	3.1	0	0	3	74.5	11.6	22.3	30.9	0
4	1	0	145	174	0	1	125	1	2.6	0	0	3	82.9	12.7	19.5	28.9	0
5	1	0	148	203	0	1	161	0	0	2	1	3	92.3	12.7	28.5	28.2	0
6	0	0	138	294	1	1	106	0	1.9	1	3	2	87.9	13	18.3	29.6	0
7	0	0	100	248	0	0	122	0	1	1	0	2	29.1	83.7	14.9	22.7	1
8	1	0	114	318	0	2	140	0	4.4	0	3	1	28.3	82.9	15.9	25.4	0
9	1	0	160	289	0	0	145	1	0.8	1	1	3	31.4	95.9	14.9	16	0
10	1	0	120	249	0	0	144	0	0.8	2	0	3	28.2	81.6	14.7	22	0
11	1	0	122	286	0	0	116	1	3.2	1	2	2	30.5	83.7	14.1	29.7	0
12	0	0	112	149	0	1	125	0	1.6	1	0	2	31.3	78.9	14.9	25.8	1
13	0	0	132	341	1	0	136	1	3	1	0	3	28.2	69.7	16.7	27.5	0
14	0	0	118	210	0	1	192	0	0.7	2	0	2	30.2	74.7	13.4	25.2	1
15	1	0	140	298	0	1	122	1	4.2	1	3	3	31	99.7	14.7	28.9	0
16	1	0	128	204	1	1	156	1	1	1	0	0	28.7	85.3	15.9	24.3	0
17	0	1	118	210	0	1	192	0	0.7	2	0	2	32	90.1	14	25.5	1
18	0	2	140	308	0	0	142	0	1.5	2	1	2	27.9	74.7	15.9	24	1
19	1	0	124	266	0	0	109	1	2.2	1	1	3	30.1	94.2	15.4	24.6	0
20	0	1	120	244	0	1	162	0	1.1	2	0	2	31.4	85.1	13.6	19.3	1
21	1	2	140	211	1	0	165	0	0	2	0	2	29.2	93.2	16.5	19.3	1

Fig 1.5 - Dataset

The above Dataset consists of columns contains symptoms and parameters which are required to predict the disease.it is made from kaggle and combining several records for heart disease.

Data Pre-processing:

The machine learning (ML) model's error rate is as near to the dataset's real error rate as is reasonably possible, according to validation processes, which are used to determine it.Utilising duplicate values and a description of the data type, such as an integer or float variable, the missing value may be located.When the model design includes skill from the validation dataset, the evaluation gets more and more skewed. Although this is routinely done, To assess a certain model, one uses the validation collection.

Data visualization:

In applied statistics, Actually, the majority of statistics' attention is given to numerical estimates and data descriptions. This might be useful for examining and learning about a dataset to seek for trends, falsified data, outliers, and more. Instead of using measures of association or relevance, visualizations may be used to illustrate and convey key linkages in plots and charts that are more immediate and relevant to stakeholders.

Comparing Algorithm

A requirement that each algorithm be tested using a consistent test harness can ensure that each method is evaluated uniformly on the same data, which is necessary for performing a fair comparison of machine learning algorithms.a computer learning method that can classify different types of anaemia.Machine learning techniques have shown promise in the diagnosis and classification of anaemia and routinely deliver trustworthy findings.

IV. EXPERIMENTAL ANALYSIS AND RESULTS

Logistic Regression

When predicting one of two potential outcomes in binary classification, which is a statistical technique, logistic regression is utilised. Regression analysis of this kind analyses the likelihood that an event will occur.The logistic regression model simulates the likelihood of the binary result using a logistic function, commonly referred to as a sigmoid function. Any real-valued integer may be converted using the sigmoid function. By identifying the parameter values that maximise the chance of actually witnessing the binary outcomes given the predictors, a procedure known as maximum likelihood estimation is used to estimate the model parameters, also known as coefficients.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

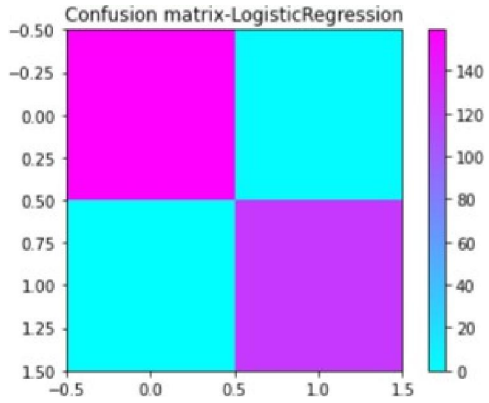

Fig 1.8:Confusion matrix

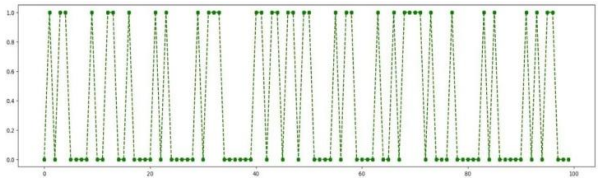

Fig 1.9:Result graph

Gender	Precision	Recall	F1-Score
0	1.00	0.99	0.99
1	0.98	1.00	0.99

Multi Level Perceptron Classifier:

Multi-layer Perceptron (MLP) classifiers, also known as artificial neural networks, type of supervised machine learning model used for classification tasks. MLP classifiers are a type of artificial neural network that consists of multiple interconnected layers of nodes, or neurons, organized in a sequential manner.

The basic building blocks of an MLP classifier are neurons, which receive input values, apply an activation function to produce an output, and pass it on to the next layer. An MLP classifier has an input,output layer.In between them a hidden layer is present. Each layer, except the output layer, is associated with a weight matrix that represents the strength of connections between neurons in adjacent layers. Iterative weight adjustments are used throughout training to reduce the discrepancy between expected and actual results.

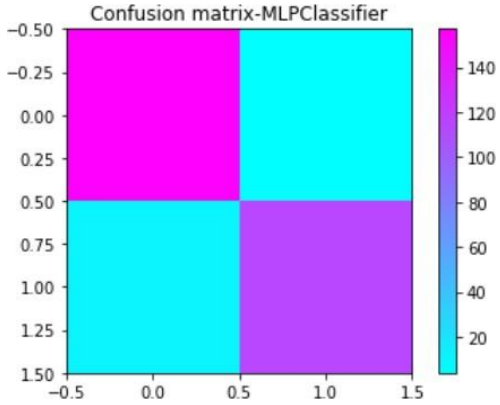

Fig 2.0:Confusion matrix

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 8, April 2024

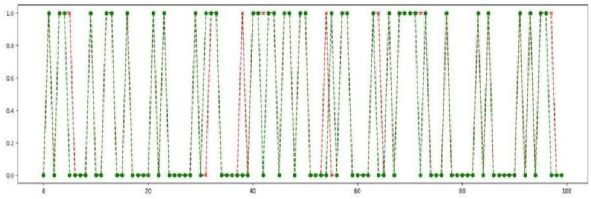

Fig 2.1:Result graph

Performance Analysis:

Gender	Precision	Recall	F1-score
0	0.94	0.98	0.96
1	0.97	0.92	0.94

Algorithm	logistic regression	Mlp classifier
Accuracy	97%	94%

V. CONCLUSION

In conclusion, using statistical data to effectively detect anemia has significant potential for machine learning algorithms. Different learning models, including logistic regression, decision trees, and mlp classifiers, have successfully classified anaemia with high precision and recall by employing features like haemoglobin level. However, additional study is required to examine the performance of these models in actual clinical situations and to enhance the models' interpretability.

REFERENCES

[1]. M. Gandhi, "Predictions in heart disease using techniques of data mining," in Proceedings of the International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520–525, Noida, India, February 2015.

[2]. S. Abdullah, "A Data mining Model for predicting the Coronary Heart Disease using Random Forest Classifier," International Journal of Computer Application, vol. 22, 2012.

[3]. S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine learning improve cardiovascular risk prediction using routine clinical data?" PLoS One, vol. 12, no. 4, Article ID e0174944, 2017.

[4]. V. V. Ramalingam, A. Dandapath, and M. K. Raja, "''Heart disease prediction using machine learning techniques: a survey," International Journal of Engineering & Technology, vol. 7, no. 2.8, pp. 684–687, 2018.

[5]. L. Baccour, "Amende d fuse d TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets R," Expert Systems With Applications, vol. 99, pp. 115–125, 2018.

[6]. R. Das, I. Turkoglu, and A. Sengur, "Expert Systems with Applications Effective diagnosis of heart disease through neural networks ensembles," Expert Systems with Applications, vol. 36, no. 4, pp. 7675–7680, 2009.

[7]. Cheng and H. Chiu, "An artificial neural network model for the evaluation of carotid artery stenting prognosis using a nationalwide database," in Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2566–2569, Jeju Island, Republic of Korea, July.

[8]. J. Nahar, T. Imam, and K. S. Tickle, "Expert Systems with Applications Association rule mining to detect factors which contribute to heart disease in males and females," Expert Systems with Applications, vol. 40, no. 4, pp. 1086–1093, 2013.

[9]. S. Zaman and R. Toufiq, "Codon based back propagation neural network approach to classify hypertension gene sequences," in Proceedings of the 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 443–446, Cox's Bazar, Bangladesh, February 2017.