EAI Endorsed Transactions
on Pervasive Health and Technology

Research Article EAI.EU

Prediction of Anemia using various Ensemble Learning and Boosting Techniques

Nalluri Shweta[1], Sagar Dhanraj Pande[1,*]

[1] School of Computer Science & Engineering (SCOPE), VIT-AP University, Amaravati, Andhra Pradesh, India

Abstract

INTRODUCTION: Anemia is a disease of great concern. It is mainly seen in people who are deficient in several vitamins like B12 and those who are deficient in iron. Neglecting the situation and leaving it untreated could lead to severe consequences in the future.Hence it is of great importance to predict Anemia in an individual and treat it in the optimum stage.
OBJECTIVES: In this paper, machine learning was used for the prediction of Anemia.
METHODS: The dataset used for this was formed by combining different datasets from Kaggle. The accuracy of various machine learning techniques was evaluated to find out the best one. Along with the supervised learning algorithms like Random Forest, SVM, Naive Bayes etc., Linear Discriminant Analysis, Quadratic Discriminant Analysis and ensemble learning methods were also performed.
RESULTS: Upon evaluation, among the best performers, the execution time was also taken into consideration to determine which classifier works well. Among all the algorithms used, XGboost worked thebest with an optimum execution time.
CONCLUSION: The conclusion is that for the data used in the work, XGboost results as the best model.

Keywords: Anemia, Prediction, Machine Learning, Random Forest, Ensemble learning, Boosting.

Received on 04 July 2023, accepted on 10 October 2023, published on 20 October 2023

doi: 10.4108/eetpht.9.4197

*Corresponding author. Email: sagarpande30@gmail.com

1. Introduction

With the increasing growth in technology and science, machine learning has risen to a very high level in the industry and is in good demand. For being so strong and versatile in nature, machine learning has emerged to be one of the most intriguing technologies [1]. The simulation of human intelligence by a machine is termed Artificial Intelligence (AI). Machine learning is a subfield of this domain. It mimics the way a human brain learns and gets trained to a specific environment. Making use of the vast amount of data present in the world, a machine learning model, on the application of various algorithms, trains itself in such a way that it can make predictions on the data in the same manner as a human does. Machine learning is leading the world. It has a wide range of applications in many industries like healthcare, agriculture, product recommendations, text analysis and review classifications. For more effective decision making with little human effort, it recognizes crucial and essential patterns. Algorithms used in machine learning succeeded in early detection and more precise prognosis of diseases such as Anemia, Diabetes, Skin Cancer, Alzheimer, Parkinson's disease, etc. with the view of saving our lives. Healthcare related data are utilized in medical research to forecast epidemics, identify various diseases, improve quality of life, and avert early deaths. This shows how essential machine learning is in health informatics.

One of the most critical diseases that must receive at most care and best curing treatment is Anemia. Anemia is a critical condition which results in an individual due to the drastic fall in the count of red blood cells. Red blood cells must always remain within the optimum range as they provide oxygen to blood tissues. Anemia defines the inability of the blood to carry oxygen which is due to insufficient amounts of red blood cells. The cause for Anemia is reduction in the amount of haemoglobin, a protein present in RBCs, which helps in carrying oxygen to all parts of our body. The symptoms shown by an anaemic person are being tired or feeling cold most of the time. It may be diagnosed in a human of any age and gender. The skin turns paler, and one may find it a little difficult to take breath. The feeling of dizziness, irregular heartbeat, headache, and chest pain are also the symptoms of Anemia. Anemia might be inborn or could be developed with growing age due to deficiency of iron or essential vitamins. It was proven that pregnant women are prone to Anemia and old aged people have a relatively higher chance to be diagnosed by Anemia. Anemia could be of multiple types like due to iron deficiency, Anemia of chronic disease, Sickle cell Anemia, Anemia due to vitamin B12 deficiency. The Complete Blood Cell reports and Peripheral blood smear reports are used to diagnose Anemia based on the count of red blood cells. When classified according to the MCV (Mean Corpuscular Volume) values, Anemia is of three types, namely, Normocytic Anemia, Microcytic Anemia, Macrocytic Anemia. Various types of Anemia are depicted in Figure 1.

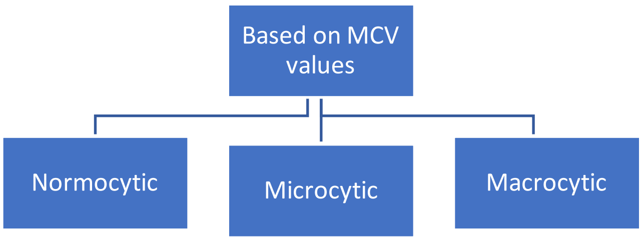

Figure 1. Classification of Anemia

It is very crucial to treat anemia in the beginning stages only at the earliest possible because insufficient supply of oxygen to our vital organs could lead to several other adverse effects like heart attacks. When left untreated, it may lead to death. Hence it is very important to predict the presence of Anemia from the reports when the above-mentioned symptoms are observed in a patient to cure them effectively.

Several machine learning algorithms can be used to predict anemia from the CBC reports. In this work, several machine learning algorithms like Logistic Regression, Support Vector Machine (SVM), Decision Trees, Artificial Neural Networks, KNN, Naive Bayes, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Ensemble learning were used for the prediction of anemia in the given dataset. Several related works have also been done in the past.

2. Literature Review

A good amount of research has been performed by various people for the prediction of Anemia. In the study [2] done by Prakriti Dhakal, Santosh Khanal and Rabindra Bista, they have performed various machine learning algorithms on the data they have collected from Kanti Children Hospital to predict Anemia in children of age less than 5. According to their study, an accuracy of 98.4% was shown by Random Forest which has performed the best on their dataset. In [3] efforts were made to compare various supervised machine learning algorithms and as a result they have also found that Random Forest has shown the best accuracy of 53% followed by SVM with an accuracy of 43%. It was observed that in the work [4] they have put in efforts to find the accuracy for Naive Bayes, Random Forest and Decision tree and Naive Bayes got the highest accuracy of 96.09% while Random Forest got ac accuracy of 95.32% and Decision tree got an accuracy of 95.46%. When Mean Absolute Error was used, Decision Tree (C4.5) showed a higher value of 0.0347. In [5] they have used non-invasive methods to diagnose anemia from the tweets. They have tried to identify the different emotions which are connected to anemia. Their evaluation results showed that SMO classifier has given them a higher accuracy of 98.96%, Bagging showed an accuracy of 63.44% and Random Forest showed a lower accuracy of 41.82% which was contrary to the above studies. In the work [6], they have built a computer aided illness diagnosis which will properly detect anemia in a person.

The system takes the blood cells levels entered by the user and then evaluates the models based on the input given by the user. To perform this task, Lasso and Ridge regressors were used out of which Ridge regression showed better accuracy. All these works have made use of the data obtained from blood tests and performed their models on that. On the data collected from the demographic health survey of Ethiopia, the study [7] was conducted to predict Anemia using ensemble learning ml algorithms in pregnant women. factors like region, educational level, pregnancy duration, history of place of delivery, Vitamin A in the last 6 months, etc. were taken into consideration for building the model. Their results show that catboost has the highest accuracy of 97.44% while random forest and extreme gradient boost have an accuracy of 94.4% and 95.21% respectively. Machine learning, as discussed earlier, has become a leading domain in healthcare. The study [8] has shown how Naive Bayes and Decision Tree have been used to generate a system that can predict disease and recommend a treatment. In the research work [9] they have worked towards prevention of anemia among women who are pregnant. In work [10] models were developed to predict if a child is anemic or not in Afghanistan using the data collected from Kunduz province's hospitals. Their study showed that Random Forest worked best with accuracy of 86.4%. In the research work [11], ML algorithms were used to determine if conjunctiva of eyes, palpable palm or color of fingernails is more accurate to detect anemia among children. Their study showed that CNN has worked better, being 99.12% accurate. Efforts were put in work

[12] to diagnose anemia using the images of conjunctiva of eye of patients belonging to India and Italy. Comparative analysis of existing techniques is represented in Table 1.

Table 1. Accuracies obtained in existing works.

Study	Accuracy and Model
Ref [2]	98.4% - Random Forest
Ref [3]	53% - Random Forest
Ref [4]	96.09% - Naive Bayes
Ref [5]	98.96% - SMO
Ref [7]	97.44% - Random Forest
Ref [10]	86.4%- Random Forest

3. Methodology

Anemia prediction with various machine learning algorithms have been incorporated. As a part of this, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Artificial Neural Network, hybridizing two algorithms and performing ensemble machine learning algorithms were used. Figure 2 depicts the flow of the proposed methodology.

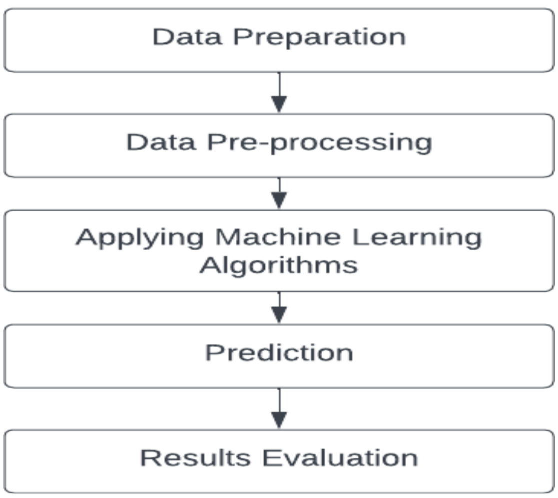

Figure 2. Methodology

3.1. Data Collection

For this data, the Anemia datasets that are available on Kaggle were combined and then utilized for the implementation purpose. For combining the dataset necessary preprocessing techniques were performed. The necessary values which were missing upon combining were added manually into the dataset. These two datasets were the hospital results of CBC results and have the attributes Gender, Hemoglobin, MCV, MCH, MCHC and the target attribute which shows if the person is Anemic or not. It was ensured that there are no missing values or null values in the dataset.

3.2. Data Pre-processing

The data collected has a total of 16721 instances and 6 columns, where each data shows if that person is suffering from Anemia or not. Upon performing data visualization, it could see that out of 16721 data instances, 10126 were shown to have no Anemia and 6595 had. Amongst all the people, it was found that 72.3% had Normocytic Anemia, 26.2% had Microcytic Anemia while the remaining 1.5% of the people had Macrocytic Anemia. The data is said to be not biased as the ratio of the target variable is 60:40, not being completely biased only to one class. Figure 3 represents the data distribution whereas Figure 4 represents the class distribution of the target class.

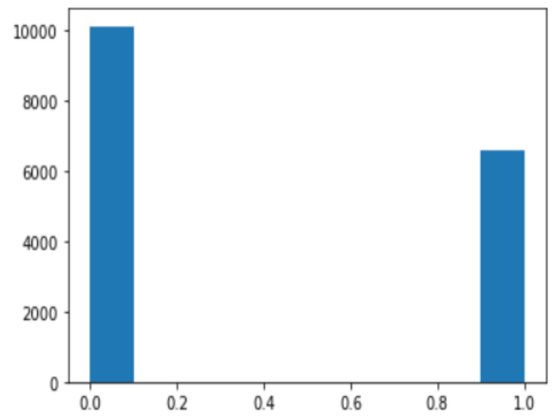

Figure 3. Bar plot for the data distribution of target class

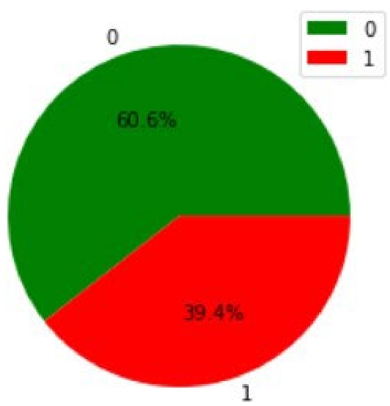

Figure 4. Pie chart for the visualization of the data of target class

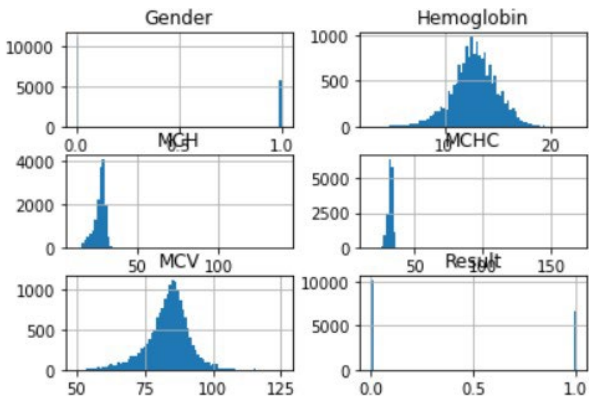

Figure 5. Histogram to visualize the frequency of various datapoints of each variable.

Figure 5 represents the frequency values of various variable in the dataset. As a part of the data pre-processing to get more accurate results, all the numerical values were scaled down to values between 0 and 1 using standard scaling method. The resulted data of scaling of the dataset is represented in Figure 6.

	Gender	Hemoglobin	MCH	MCHC	MCV	Result	Type
0	1	0.612025	0.080269	0.073568	0.448056	0	Normocytic
1	0	0.663412	0.101621	0.068232	0.292822	0	Microcytic
2	0	0.308839	0.070779	0.076903	0.282208	1	Microcytic
3	0	0.612025	0.027284	0.088908	0.498474	0	Normocytic
4	1	0.601747	0.074733	0.067565	0.657689	0	Normocytic
...	...	...	...	...	...	...	...
16716	0	0.463001	0.165678	0.106917	0.641767	0	Normocytic
16717	0	0.519527	0.118229	0.098246	0.446729	0	Normocytic
16718	0	0.550360	0.119810	0.101581	0.441422	0	Normocytic
16719	0	0.468140	0.114274	0.087574	0.486533	0	Normocytic
16720	0	0.555498	0.129300	0.108917	0.450710	0	Normocytic

Figure 6. Data after performing scaling.

To ensure that the data is not biased to a specific class, the data was balanced using SMOTE i.e., Synthetic minority Oversampling Technique which oversamples the minority class. Upon performing this, the total number Ξ of instances increased to 20252, with each class having 10126 instances. Hence it was balanced. Data distribution after post scaling is depicted in Figure 7.

The Pearson correlation coefficient of the different independent variables was visualized to analyze the strength between the continuous independent variables is represented in Figure 8. From the plot it was inferred that there is a high dependency between the attributes MCH and MCHC.

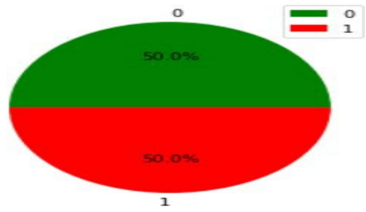

Figure 7. Data visualization Scaled target class.

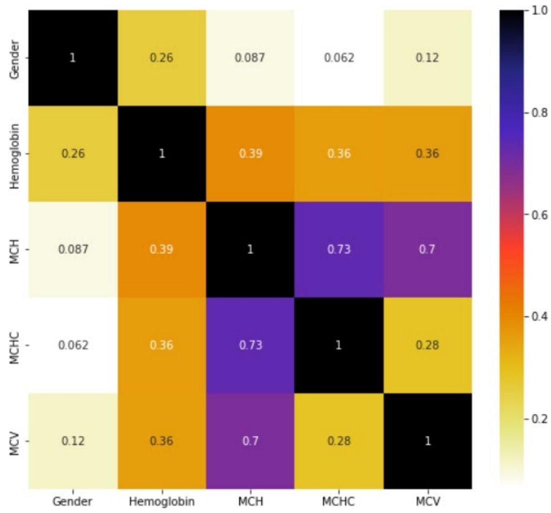

Figure 8. Pearson Correlation Coefficient Matrix

3.3. Machine Learning Algorithms

For the prediction of a person being anemic or not Logistic Regression, Support Vector Machine (SVM), Decision Trees, Artificial Neural Networks, KNN, Naive Bayes, Linear Discriminant Analysis, Quadratic Discriminant Analysis were implemented. Later ensemble learning methods were also used to check if they would improve the accuracy of the weak learners. Stratified K-fold cross validation was also performed for weak learners. For performing all these algorithms, data was divided into two samples, training, and testing in the ratio 7:3 to train on more instances and test better. The outcome for the prediction is binary i.e., 0 indicating the absence of Anemia and 1 indicating its presence.

Logistic Regression

It is a simple method for binary classification widely used in medical diagnosis. It finds the relationship between input features and target variable where the output is modeled as the probability of belonging to one of the classes. The accuracy obtained on training and testing the data with this algorithm is 98.40% and 98.15%

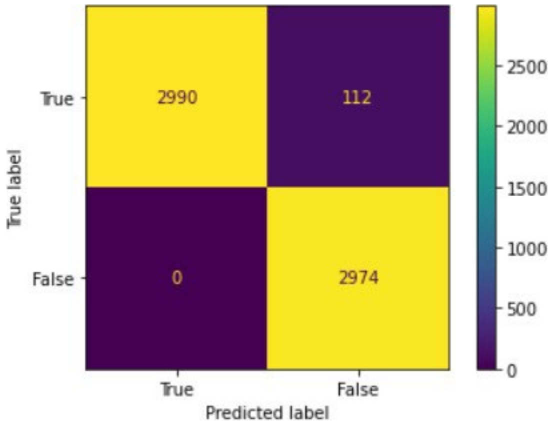

Figure 9. Confusion Matrix for Logistic Regression

respectively. The precision was 0.959 and the recall was 0.996 and f1 score was 0.9774.
The mean obtained upon performing Stratified K-fold Cross Validation with 7 splits was 96.4% and it increased to 97.87% when implemented using 15 splits. Figure 9 represents the confusion matrix for Logistic Regression.

Logistic Regression is applied using the formula [13]:

$$y = \frac{e^{(m+nx)}}{1 + e^{(m+nx}}$$

Where,
y: Dependent variable
x: Independent variable
m: Intercept
n: Slope of the line

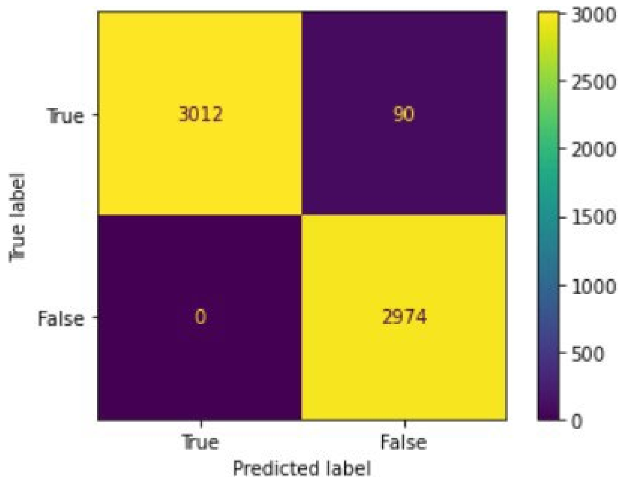

Figure 10. Confusion Matrix for SVM

Support Vector Machine

Support Vector Machine is an algorithm that uses classification methods to address two group classification issues.
When trained using this model, the test accuracy obtained was 98.51% and the training accuracy was 98.57%. This accuracy was obtained using the RBF kernel with C parameter. The stratified k-fold result was 97.94% upon 15 folds. Figure 10 represents the SVM algorithm's confusion matrix.

When ν- parameter was used instead of C parameter with RBF kernel the accuracy was relatively less. It fell to 89.99% and the time taken was more than that for C parameter.
Later, to create a pipeline between Logistic Regression and SVM where Logistic Regression works for feature selection and uses L1 Regularization to prevent overfitting were considered. The highest accuracy of 99.85% was obtained when C parameter was 11 and γ value was 0.3.
The RBF for an SVM is formulated as [14]:

$$K(A,B) = e^{(-\frac{||A-B||^2}{2\sigma^2})}$$

where σ is our hyperparameter(variance) and $||A-B||^2$ is the distance between the points A and B.

Random Forest

Random forest is also a supervised learning algorithm which is ensemble learning. It is used in classification as well as regression problems. This algorithm consists of many decision trees. It determines the output from the predictions made by the different trees. It is usually preferred over other algorithms as it overcomes the problem of overfitting and diminishes the effort of hyper-parameter tuning. Upon using this algorithm, the accuracy reached was highest so far, which is, 99.96%. Figure 11 represents the Random Forest algorithm's confusion matrix.

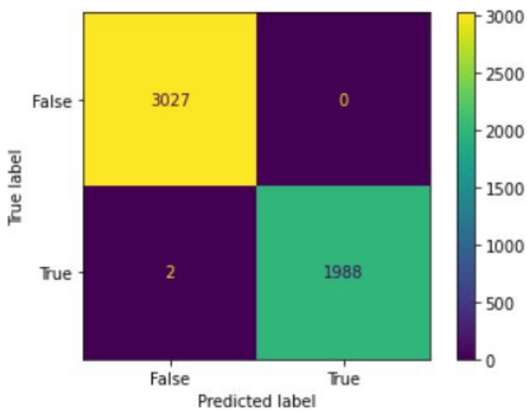

Figure 11. Confusion Matrix of Random Forest

KNN

KNN, called as, K-Nearest neighbor algorithm, is also a classification algorithm widely used in machine learning. Upon using this the training accuracy achieved was 98.49% and testing accuracy as 96.09% with 3 as the number of nearest neighbors. Figure 12 represents the KNN algorithm's confusion matrix.

When using the KNeighborsClassifier in python, it by default calculates the Euclidean distance which is given by [15]:

$$d(U,V) = \sqrt{(u1 - v1)^2 + (u2 - v2)^2}$$

where u and v are the data points whose distance were figured out.

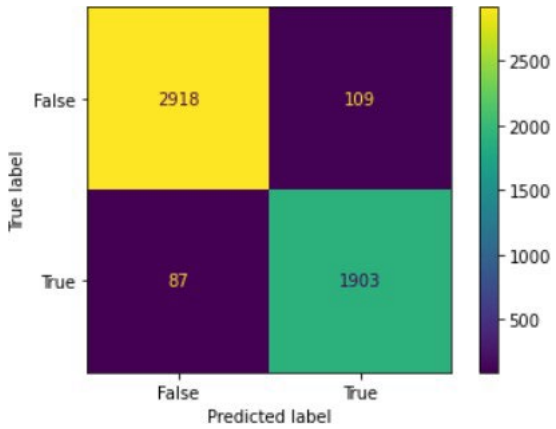

Figure 12. Confusion Matrix of KNN

Naïve Bayes

Naive Bayes is used to classify the instances based on its previous learning. It assumes the variables to be independent of each other. The accuracy achieved was 83.39%. Figure 13 represents the Naïve Bayes algorithm's confusion matrix.

The calculation of probability for a target class is done using the following formula [16]:

$$P(bi|a1, a2..an) = P(a1|bi) * P(a2|bi) \ldots P(an|bi) * P(bi)$$

Where,

P(ai|bi) is the individual conditional probability of all the independent variables with the target class.

P(bi) is the probability of the target class bi.

Figure 13. Confusion Matrix of Naïve Bayes

Artificial Neural Network (ANN)

In this work, ANN made use of 4 layers, where one layer is the input layer, 2 hidden layers and one output layer. ReLu function was the activation function for the hidden layers and the sigmoid function was used as the activation function for the output layer. Upon performing 50 epochs with adam as the optimizer and binary cross entropy as the loss function, an accuracy of 96.39% was achieved with 0.124 as loss. Figure 14 represents the ANN algorithm's confusion matrix. Figure 15 depicts the ANN architecture.

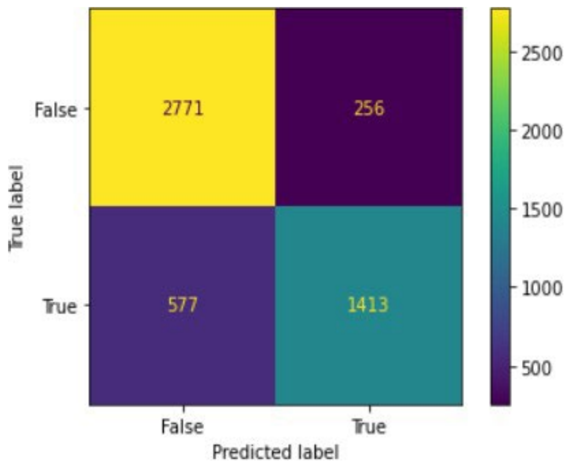

Figure 13. Confusion Matrix of Naïve Bayes

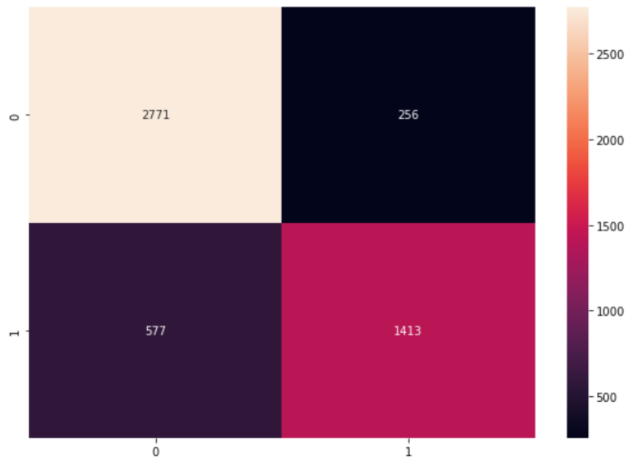

Figure 14. Confusion Matrix for ANN

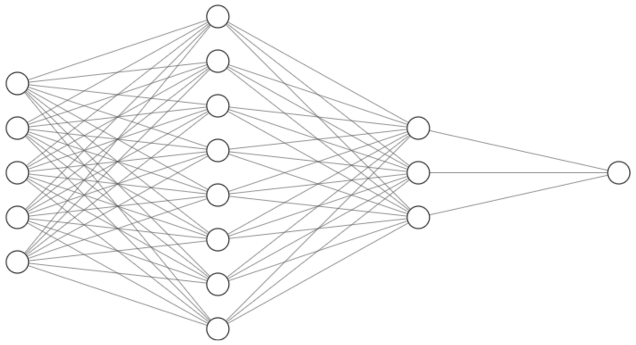

Figure 15. ANN Architecture

When tried to increase the accuracy by hybridizing with SVM, the highest accuracy obtained was 98.25%.

Linear Discriminant Analysis and Quadratic Discriminant Analysis

These are classification algorithms that build a linear and non-linear boundary between the classifiers. The accuracy obtained was 95.15% and 91.80% respectively. Table 2 depicts Comparative analysis of all accuracies obtained by using various machine Learning model the models.

Table 2. Accuracy Comparison

Model	Accuracy obtained
Random Forest	99.96%
SVM	98.51%
KNN	98.49%
Logistic Regression	98.15%
ANN	96.39%
LDA	95.15%
QDA	91.80%
Naive Bayes	83.39%

From Table 2, it can be observed that Random Forest gave the highest accuracy of 99.96% and the least was given by Naive Bayes which is 83.39%. It was also observed that Logistic Regression performed better on this data when compared to LDA and QDA.

Ensemble Learning

(i)Voting

It is a machine learning technique that takes the predictions of multiple weak models and combines them to give the result. It is a method where multiple models are trained and combined in some way to increase the complete predictive accuracy and robustness of the model [17][18]. In this paper, several hard voting and soft voting where used. Also, it combines the predictions made by Logistic Regression, Random Forest, SVM, KNN, Naive Bayes and ANN. Hyperparameter tuning of the weak algorithms was done and on performing hard voting an accuracy of 99.88% was obtained with 2,62 seconds of execution. When soft voting was applied, 99.92% accuracy was obtained for 6.719 seconds. Figure 16 depicts the voting ensemble learning architecture.

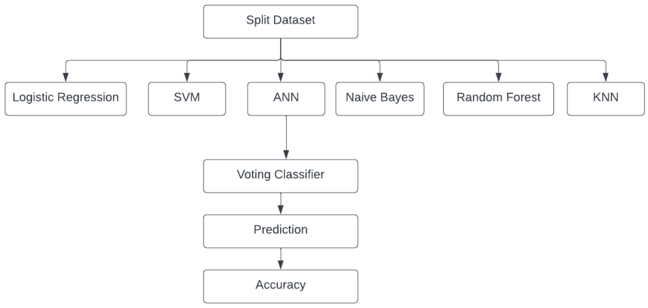

Figure 16. Architecture of Voting Ensemble Learning

(ii)Bagging

Bagging is also another ensemble learning method used to reduce variance and improve model generalization. In this work, the bagging algorithm to the 4 base estimators and have taken bags of size 10,15, 20, 25, 30, 35 and 40 were implemented. The sample size I took is 70% of the data. Table 3 depicts weak learner approaches result comparison.

Table 3. Comparison of Bagging on various weak learners

Size	RF	SVM	ANN	NB	KNN
10	83.23%	72.05%	91.60%	77.81%	85.29%
15	93.82%	75.54%	90.93%	78.77%	91.40%
20	90.06%	72.65%	92.50%	72.27%	84.53%
25	93.32%	71.93%	92.10%	77.57%	91.36%
30	95.57%	71.09%	83.11%	81.46%	92.42%
35	90.49%	74.32%	91.38%	77.89%	90.17%
40	94.19%	75.40%	92.70%	82.93%	88.75%

Bagging Algorithm has shown the above accuracies for different numbers of samples taken. It must be noted that the execution time for Bagging was exceptionally high. Among all the accuracies found so far using Bagging, Random Forest has shown more accuracy of 95.57%. Figure 17 represents the architecture of Bagging approach.

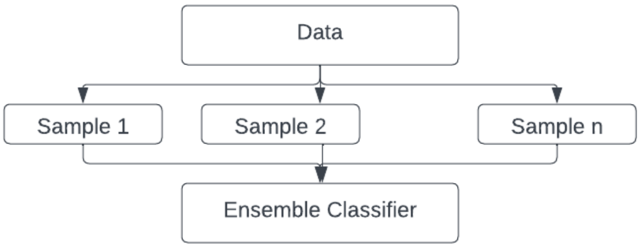

Figure 17. Architecture for Bagging

(iii)Boosting

This machine learning algorithm also combines the results of the weak learners but in a sequential manner [19][20]. In this work, the AdaBoost classifier and Xgb classifier were used. In Adaboost classifier was used, it worked well when used with decision tree as the base estimator. The accuracy was low when logistic regression was used as the base estimator. But when used without Logistic Regression as base estimator, it gave an accuracy of 100% with the timetaken for execution as 0.449 seconds. But when XGboost was performed, it showed an accuracy of 100% with time of execution as 0.2404 seconds.

4. Conclusion

Various machine learning algorithms were implemented, it was found that the best working algorithms were Adaboost and XGboost with 100% accuracy. But to evaluate which one performs better, the execution time parameter is taken into consideration. In table 4, time comparison of outperformed algorithm is depicted.

Table 4. Execution times for the best algorithms

Algorithm	**Execution time (in sec)**
XGboost	0.2404 seconds
Adaboost	0.449 seconds

Though the execution time difference is very less among them, since XGboost has worked fast, it can be observed that the XGboost has outperformed among all the approaches**.** In future, the work can be extended using optimized machine learning algorithms along with nature inspired algorithms.

References

[1] Kumar K, Kumar P, Deb D, Unguresan ML, Muresan V. Artificial Intelligence and Machine Learning Based Intervention in Medical Infrastructure: A Review and Future Trends. Healthcare (Basel). 2023 Jan 10;11(2):207

[2] Dhakal, Prakriti & Khanal, Santosh & Bista, Rabindra. (2023). Prediction of Anemia using Machine Learning Algorithms. International Journal of Computer Science and Information Technology. 15. 15-30. 10.5121/ijcsit.2023.15102.

[3] Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, *19*(1), 1-16.

[4] Jaiswal, M., Srivastava, A., & Siddiqui, T. J. (2019). Machine learning algorithms for anemia disease prediction. In *Recent Trends in Communication, Computing, and Electronics: Select Proceedings of IC3E 2018* (pp. 463-469). Springer Singapore.

[5] Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., & Shibghatullah, A. S. (2022). A non-invasive machine learning mechanism for early disease recognition on Twitter: The case of anemia. *Artificial Intelligence in Medicine*, *134*, 102428.

[6] Geetha, V., Gomathy, C. K., Keerthi, K., & Pavithra, N. (2022). Diagnostic Approach To Anemia In Adults Using Machine Learning. *Journal of Pharmaceutical Negative Results*, 3713-3717.

[7] Dejene, B. E., Abuhay, T. M., & Bogale, D. S. (2022). Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm. *BMC Medical Informatics and Decision Making*, *22*(1), 1-11.

[8] Gupta, J. P., Singh, A., & Kumar, R. K. (2021). A computer-based disease prediction and medicine recommendation system using machine learning approach. *Int J Adv Res Eng Technol (IJARET)*, *12*(3), 673-683.

[9] Abd Rahman, R., Idris, I. B., Md Isa, Z., & Abd Rahman, R. (2022). The effectiveness of a theory-based intervention program for pregnant women with anemia: A randomized control trial. *Plos one*, *17*(12), e0278192.

[10] Zahirzada, A., Zaheer, N., & Shahpoor, M. A. (2023). Machine Learning Algorithms to Predict Anemia in Children Under the Age of Five Years in Afghanistan: A Case of Kunduz Province. *Journal of Survey in Fisheries Sciences*, *10*(4S), 752-762.

[11] Asare, Justice & Appiahene, Peter & Donkoh, Emmanuel & Dimauro, Giovanni. (2023). Iron Deficiency Anemia Detection using Machine Learning Models: A Comparative Study of Fingernails, Palm and Conjunctiva of the Eye Images. 10.22541/au.167570558.82410707/v1.

[12] Nancy Noella, R. S., & Priyadarshini, J. (2023). Machine learning algorithms for the diagnosis of Alzheimer and Parkinson disease. *Journal of Medical Engineering & Technology*, *47*(1), 35-43.

[13] Dimauro, G., Griseta, M. E., Camporeale, M. G., Clemente, F., Guarini, A., & Maglietta, R. (2023). An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset. *Artificial Intelligence in Medicine*, *136*, 102477.

[14] Sarkodie, K., Fergusson-Rees, A., Abdulkadir, M., & Asiedu, N. Y. (2023). Gas-liquid flow regime identification via a non-intrusive optical sensor combined with polynomial regression and linear discriminant analysis. *Annals of Nuclear Energy*, *180*, 109424.

[15] Sharma, S., & Jain, A. (2023). Hybrid ensemble learning with feature selection for sentiment classification in social media. In *Research Anthology on Applying Social Networking Strategies to Classrooms and Libraries* (pp. 1183-1203). IGI Global.

[16] Asare, J. W., Appiahene, P., & Donkoh, E. T. (2023). Detection of anaemia using medical images: A comparative study of machine learning algorithms–A systematic literature review. *Informatics in Medicine Unlocked*, 101283.

[17] Zhang, A., Lou, J., Pan, Z., Luo, J., Zhang, X., Zhang, H., ... & Chen, L. (2022). Prediction of anemia using facial images and deep learning technology in the emergency department. *Frontiers in Public Health*, *10*, 3917.
[18] Kilicarslan, S., Celik, M., & Sahin, Ş. (2021). Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification. *Biomedical Signal Processing and Control*, *63*, 102231.
[19] Yeruva, S., Varalakshmi, M. S., Gowtham, B. P., Chandana, Y. H., & Prasad, P. K. (2021). Identification of sickle cell anemia using deep neural networks. *Emerging Science Journal*, *5*(2), 200-210.
[20] Balaji, E., Brindha, D., Elumalai, V. K., & Vikrama, R. (2021). Automatic and non-invasive Parkinson's disease diagnosis and severity rating using LSTM network. *Applied Soft Computing*, *108*, 107463.