Mansoura Engineering Journal

Volume 49 | Issue 3 Article 4

2023

Machine-learning based stacked ensemble model for accurate multi classification of CBC Anemia

Asmaa Mohamed El-Boghdady
Teaching Assistant, Teaching Assistant at ALZarka Academy of Management Science and computer., asmaa.elboghdady16@gmail.com

Sherif Kishk
Electronics and Communications Engineering Department, Faculty of Engineering.

Mohamed M Ashour
Electronics and Communications Engineering Department, Faculty of Engineering.

Eman AbdElhalim
Electronics and Communications Engineering Department, Faculty of Engineering.

Recommended Citation
El-Boghdady, Asmaa Mohamed; Kishk, Sherif; Ashour, Mohamed M; and AbdElhalim, Eman (2023) "Machine-learning based stacked ensemble model for accurate multi classification of CBC Anemia," *Mansoura Engineering Journal*: Vol. 49 : Iss. 3 , Article 4.
Available at: https://doi.org/10.58491/2735-4202.3144

ORIGINAL STUDY

Machine-learning Based Stacked Ensemble Model for Accurate Multi Classification of Complete Blood Count Anemia

Asmaa M. El Boghdady*, Sherif Kishk, Mohamed M. Ashour, Eman AbdElhalim

Electronics and Communications Engineering Department, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

Abstract

Anemia disease (AD) is one of the most common blood problems in the world. According to the World Health Organization, it leads to a lack of oxygen and red blood cells. The problem of classifying anemia is crucial in prescribing an appropriate treatment plan. Unfortunately, it is difficult to accurately diagnose the type of anemia as early as possible in order to properly treat the patient. Such a diagnosis can be complicated by the exponential increase in patient numbers, hospital priority rules, and issues of access to adequate medical professionals. This paper proposes using a computer-based model to classify the type of possible anemia. The proposed model was created using machine learning techniques. Based on logistic regression, random forest (RF), RBF support vector, and stacking classification methods. Data has been containing complete blood test results from the Faculty of Medicine, Tokat Gaziosmanpaşa University, Turkey, including sex, white blood cell (WBC), red blood cell (RBC), hemoglobin (HGB), hematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), major histocompatibility complex (MHC), and symptoms. A dataset was taken from 15 300 patients, arranged in tabular form, including 24 features and five categories: 10 379 females and 4921 males. After applying the proposed model to classify data of anemic patients, we found that classifying anemia achieved very good results, for instance, the stacking technique reached an accuracy of 99.95 %. In addition, we achieved an accuracy of recall of about 99.95 % and an F1 score of about 99.95 %. These results were better when compared with results published in the previous literature.

Keywords: Anemia disease (AD), Logistic regression, Machine learning (ML), Random forest (RF), RBF support-vector, Stacking technique

1. Introduction

Anemia is one of the most common blood problem and a common health problem worldwide (Milman, 2011). The World Health Organization (also known as 'WHO') describes anemia as a condition in which the level of red blood cells falls below the safe limit, thereby reducing the oxygen-carrying capacity and making it insufficient to meet the needs of the body (Kassebaum et al., 2014). Anemia disease (AD), caused by iron deficiency (ID), may be one of the top 10 health risks contributing to mortality worldwide. Furthermore, AD affects nearly a quarter of the world's population. Anemia status and child characteristics, including age, nutritional status, and birth weight, are the most important factors affecting anemia in children aged 6–59 months (McLean et al., 2009). If this happens to children, it can have a detrimental effect on their ability to develop a healthy immune system and normal cognitive development (WH Organization, 2014; Vohra et al., 2022). AD in children is more common for many reasons, including genetic or nutritional factors, including iron, folate, vitamin A/B12, and copper deficiencies. AD is estimated to affect approximately 25 % of the world's population, with rates of 47. 5 % and 42 % among pregnant women and preschoolers, respectively. Therefore,

---

Received 18 September 2023; revised 16 November 2023; accepted 18 November 2023.
Available online 25 January 2024

* Corresponding author. Electronics and Communications Engineering Department, Faculty of Engineering, Mansoura University, Egypt.
E-mail address: asmaa.elboghdady16@gmail.com (A.M. El Boghdady).

https://doi.org/10.58491/2735-4202.3144

the prevalence of AD in society exceeds 40 %, making it a public health problem of great importance (Vohra et al., 2022). According to the World Health Organization. Anemia is common in the majority of countries in Africa and South Asia, as well as some countries in East Asia and the Pacific. Additionally, Asia has the highest rate of anemia in children, although Africa has the highest rate of anemia (Ewusie et al., 2014). A decrease in the concentration of red blood cell mass, hemoglobin, or hematocrit, are other definitions of anemia. It is important to know that age and gender affect normal hemoglobin and hematocrit values (Murphy, 2011) for humans. Accurate diagnosis is an important step towards correct treatment, because anemia, has serious consequences on quality of life, is a disease, and a frequent predictor of other serious diseases. In addition to medical expertise, physicians also use laboratory tests to evaluate various blood components to determine the pathologies that best explain a patient's symptoms and signs (Cuomo et al., 2018). Additionally, the patient's medical history plays an important role when gathering the data needed for diagnosis. In addition, laboratories perform tests to classify diseases and evaluate them to come up with a treatment plan. However, because clinical laboratories often present test results as discrete numerical values or categories, and because clinicians often focus on numbers that fall outside of certain reference ranges Therefore, laboratory test results are often overestimated (Gunčar et al., 2018). It is important to know that up to 70 % of medical decisions can be based on test results (Rashidi et al., 2019). Additionally, up to 17 % of hospital problems and about 10 % of patient deaths can be attributed to diagnostic errors (Cifra et al., 2021). To overcome these difficulties, many researchers and companies are trying to bring machine learning into diagnostic applications for these diseases. As a result, the diagnostic process is accelerated, becomes more accurate, produces results similar to the aforementioned diagnostic applications, and becomes more reliable (Sennaar, 2019). Recently, machine learning methods are effective in solving the problem of anemia prediction and related works have focused on the use of logistic regression (LR) (El-kenawy and SM, 2019), and support vector machine (SVM) (Korup and Stolle, 2014; Wang et al., 2013), Classification and regression trees (Youssef et al., 2016; Felicísimo et al., 2013), Naive Bayes (NB) (Tien Bui et al., 2012), artificial neural networks (ANN) (Tien Bui et al., 2012; Oh and Lee, 2017), etc. The ML algorithm takes anemia-related conditions as input to predict where anemia is likely to occur without being constrained by statistical assumptions. ML techniques such as SVM and ANN are the most popular anemia modeling methods. They have acquired promising predictive power for anemia modeling and are often used as a reference for testing new methods (El-kenawy and SM, 2019). NB and LR have also been widely used to treat anemia due to their ease of use. However, the simple or basic hypothesis space for learning algorithms has difficulty covering all situations where the type of data used differs, and thus debate remains about the best technique (Esteban-Medina et al., 2019). Modeling anemia remains problematic due to the complex nonlinear relationship between the number of traits and the type of anemia onset. Improving accuracy is the goal of anemia susceptibility assessment. An ensemble machine learning approach was recently shown to be the best option for modeling anemia susceptibility. By using ensemble methods, the hypothesis space of the fitness function can be expanded, resulting in improved prediction compared with individual algorithms (Rokach, 2009). The only algorithm commonly used to train an ensemble is the ‘base learner’ algorithm, which can be uniform or heterogeneous. Several anemic studies have investigated methods for building homogeneous learner bases using meta-learning. The researcher noted that the error rate in diagnosing the disease is very high. This situation needs to be improved by using machine learning algorithms dedicated to ensemble learning to reduce the error rate mentioned above. In this paper, we used KNN bagging, LR, decision trees, RBF SVM, and Random Forest (RF). As Table 1 shows the abbreviations referred to. After conducting many experiments, we found that the proposed model has the highest accuracy because the superposition technique is important in evaluating anemia sensitivity. This method is based on several machine-learning algorithms and was first used to diagnose anemia. With meta-learners at higher levels and heterogeneous base learners at lower levels, the technique has a two-level structure. It is important to emphasize that the benefits of the superposition method have been studied by many different professions. The document of the text is broken down as follows: Section 2 generally reviews work related to modern anemia detection using machine learning algorithms. Details of the methods and processing steps of the developed framework are described in detail in Section 3. Section 4 describes the performance and validation measures to evaluate the classification accuracy. Experimental results demonstrating the potential of the proposed pipeline are presented in Section 5, and Section 6

Table 1. Overview of abbreviation.

Abbreviation symbol	Abbreviation name
AD	Anemia Disease
ML	Machine learning
LR	Logistic Regression
SVM	Support Vector Machine
RBF-SVM	RBF Support-Vector
RF	Random Forest
KNN	K-Nearest Neighbor
ANN	Artificial Neural Networks
ELM	Extreme Learning Machine
RELM	Regularized Extreme Learning Machine
NB	Naive Bayes
DT	Decision Tree
BDT	Bagged Decision Trees
BT	Boosted Trees
DHS	Dynamic Harmony Search
ANFIS	Adaptive Neuro-Fuzzy Inference System
MLR	Multivariate Linear Regression
CBC	Complete Blood Count
SMOTE	synthetic minority oversampling technique
ROC	receiver operating characteristic
AUC	Area under the ROC Curve

presents discussions and results. Finally, section 7 provides conclusions for this research.

2. Literature review

In recent decades, studies have used the CBC dataset. These studies concentrated on creating and analyzing a variety of machine learning techniques, as shown in Table 2. Beck et al., who developed a computer-supported system for medical education research, conducted one of the earliest investigations on computer-aided anemia diagnosis. In 1989 (Esteban-Medina et al., 2019), they introduced the PlanAlyzer for the diagnosis of anemia. In that article, a system is proposed to make it possible to identify anemia in clinical practice settings. An artificial learning model was created for this system utilizing four different methods. As classification algorithms, Artificial Neural Networks, Support Vector Machines, Naive Bayes, and Ensemble Decision Tree techniques are employed. 12 separate anemia types were identified by comparing the models to a dataset of 1663 samples. They included 25 factors total, which included hemogram information as well as demographics including age, sex, chronic illnesses, and symptoms. Information is acquired at a university hospital in Turkey by checking through patient records. The model also made use of eight other datasets that were created using particular feature selection techniques in addition to the information the doctors used. Students and medical experts will both benefit from the interface's decision-support capabilities. Each of the four distinct methods of categorizing data has a good success rate. Each model is verified using the following metrics: classification error, area under curve, precision, recall, F-score. The accuracy of bagged decision trees was highest (85.6 %) (Karagül Yõldõz et al., 2021). There are numbers of advantages such as using four different artificial learning methods and 12 different anemia types, disadvantages of this research are numbers of data are small and Low Accuracy (Karagül Yõldõz et al., 2021). Unlike in our work, we achieved high accuracy near to 99.95 %.

Medical professionals use healthcare data to identify diseases, predict epidemics, improve quality of life, and prevent avoidable deaths (Jaiswal et al., 2019). Using WEKA, a comparison of various categorization techniques for hematological data (Oguntimilehin et al., 2014; Gomathy et al., 2022). The J48 classifier is the most effective approach based on the hematological data, with an accuracy of 97.16 % and a build time of 0.03 s (Geetha et al., 2022).

Have the following advantages: Computer-aided disease diagnosis is less expensive, saves time, is more accurate, and eliminates the need for additional personnel for medical decision-making. Disadvantages: the dataset used is not large enough to give high accuracy and the best results. Binary (anemic or not anemic).

Thalassemia screening and iron deficiency anemia (IDA) prediction are two medical treatments that use anemia classification. This study successfully tested a novel method for IDA and β-thalassemia trait ((β-TT) discrimination. The approach is built on the Dynamic Harmony Search (DHS). A complete blood count (CBC), a simple and low-cost laboratory technique, serves as the system's input. The artificial neural network (ANN), the adaptive neuro-fuzzy inference system (ANFIS), the support vector machine (SVM), the k-nearest neighbor (KNN), the k-nearest neighbor (KNN), the k-nearest neighbor (KNN), the structured representation on genetic algorithm in nonlinier function fitting (STROGANOFF), The results show that the proposed method has an accuracy of about 98 % (Qasem and Mosavi, 2020).

The advantages of this research are High Accuracy and using different methods to improve the accuracy, on the other hand the disadvantages such as binary classification and number of data is small.

Saputra et al., 2023 were developed using Extreme Learning Machine (ELM) technique. Performance was then evaluated using a confusion matrix and

Table 2. Overview of previous machine learning techniques for intrusion detection.

Ref. (year)	Author	Methodology	Dataset	Results	Advantage	Disadvantage	Classes
Saldivar-Espinoza et al., 2019	Saldivar-Espinoza, Bryan et al.	7-layer CNN	300 samples from young children Conjunctiva	(ACC = 42.96 %, SN = 77.58 % and SP = 36.03 %)	-Using 7-layer CNN - number of data is appropriate	low accuracy Binary classification	2
Kasiviswanathan et al., 2020	Kasiviswanathan, Sivachandar et al.	(CNN)	The model uses 135 eye images	(f1 score = 94 %)	-performed conjunctival region segmentation for non-invasive anemia detection applications using deep learning.	- they used binary classification to know anemia or not anemia	2
Çil et al., 2020	Çil, Betül et al.	(LR)- (KNN)-(SVM)-(ELM) − (RELM)	Lab tests revealed that adult Thais have HMA.	Co-evaluation of male and female patients yielded accuracy results of 96.30 % for female, 94.37 % for male, and 95.59 %; f1 score = 0.98 %.	-Using different methos -High Accuracy	-No of Patients is very low (152) -Binary classification	2
Qasem and Mosavi, 2020	Qasem et al.	(DHS)-ANN)-(ANFIS)- KNN)	132 CBC samples	DHS(ACC = 98 %)	-High Accuracy -using different methods to improve the accuracy	-Binary classification -size of data is small	2
Karagül Yõldõz et al., 2021	Yõldõz et al.	ANN, and BDT-BT	1663 samples, at a university hospital in Turkey by checking patient records.	boosted trees (ACC = 83.0 %) ANN (ACC = 79.6 %). Bagged decision trees (ACC = 85.6 %).	-using four different artificial learning methods -12 different anemia types	-No of data is small -Low Accuracy	12
Geetha et al., 2022	Geetha, V. et al.	J48 classifier- Naive Bayes	CBC dataset	J48 (ACC = 97.16 %) Naïve Bayes (Acc = 29.71 %)	- Computer-aided disease diagnosis is less expensive, saves time, is more accurate, and eliminates the need for additional personnel for medical decision-making.	-the dataset used is not large enough to give high accuracy and the best results. -Binary (anemic or not anemic)	2

Sarsam et al., 2022	Sarsam, Samer Muthana et al.	SMO, Bagging, OneR, RF	1 738 759 English tweets	SMO (ACC = 98.96 % and PER = 96 %)	they used k-means algorithm and Latent Dirichlet Allocation (LDA) algorithm to cluster similar tweets and identify hidden disease topics and high accuracy almost up to 98.96 %.	-Binary classification	2
Jaiswal et al., 2019	Saputra et al.	(ELM)	(CBC)	(ACC = 99.21 %, SN = 98.44 %, PER = 9.30 % and f1-score = 99.84 %)	-Using data mining and supervised Ml (NB-RF-DT) Multi-classification -multi-classification	Accuracy almost up to 99.21 % only	7
Dimauro et al., 2023	DiMauro, Giovanni et al.	SVM- ANN - linear regression - DT - k-NN - LS-SVM (least squares support vector machine-MLR-kM (k-means).	There are 218 pictures of eyes on eyes-defy-anemia.	(ACC = 83 %, SN = 69 % and SP = 87 %)	-Using different algorithms such as (svm-ANN-DT-KNN-LS-)	-Size of data is small - Low Accuracy - Binary classification	2

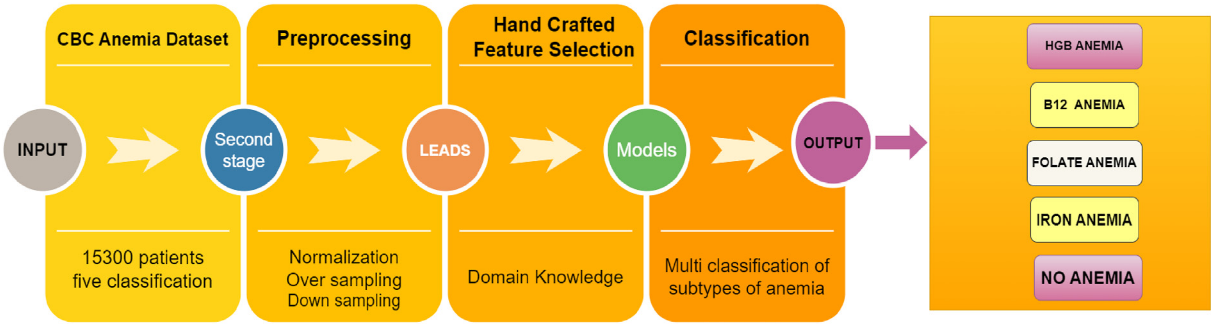


Fig. 1. Overall pipeline for Methods.

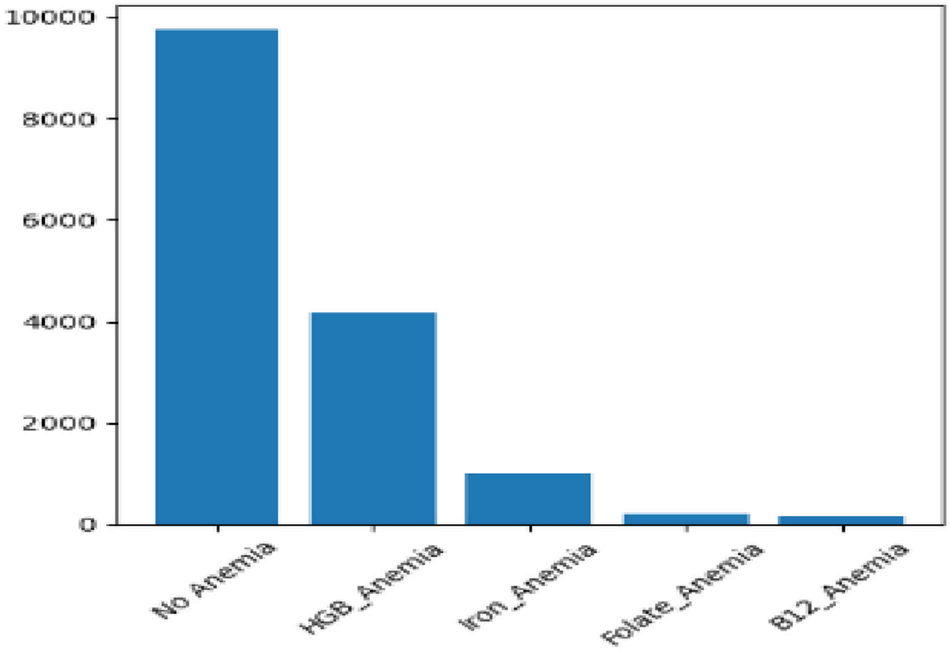


Fig. 2. Distribution of unbalanced datasets.

190 data points representing the four classes. The results show that the F1 score is 98.84 %, the accuracy is 99.21 %, the sensitivity is 98.44 %, and the precision is 99.30 % (Çil et al., 2020). To distinguish between IDA and -Thalassemia, a decision-assistance system has been developed. The classification methods Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Extreme Learning Machine, and Regularized Extreme Learning Machine were used in the suggested system. The accuracy, sensitivity, f-measure, and specificity parameters were utilized to test the classification performance utilizing 342 patients' collected data on hemoglobin, RBC, hematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), and RDW features. When male and female patients were evaluated together, the accuracy was 95.59 %, compared with 96.30 % for female patients.

One of the advantages of this research is that it uses a different methods such as (ELM, RELM, SVM, DT, KNN) and High Accuracy, but the disadvantages are No of Patients is very low (152) and

Table 3. Data split showing types, the number of patients, males, females, features and classes for the dataset.

Type	# Patients	Percentage	#Male	#Female	#Features	#Classes
No Anemia	9747	64 %	4921	10 379	24	5
HGB Anemia	1019	7 %				
Iron Anemia	4182	27 %				
B12 Anemia	199	1 %				
Folate Anemia	153	1 %				

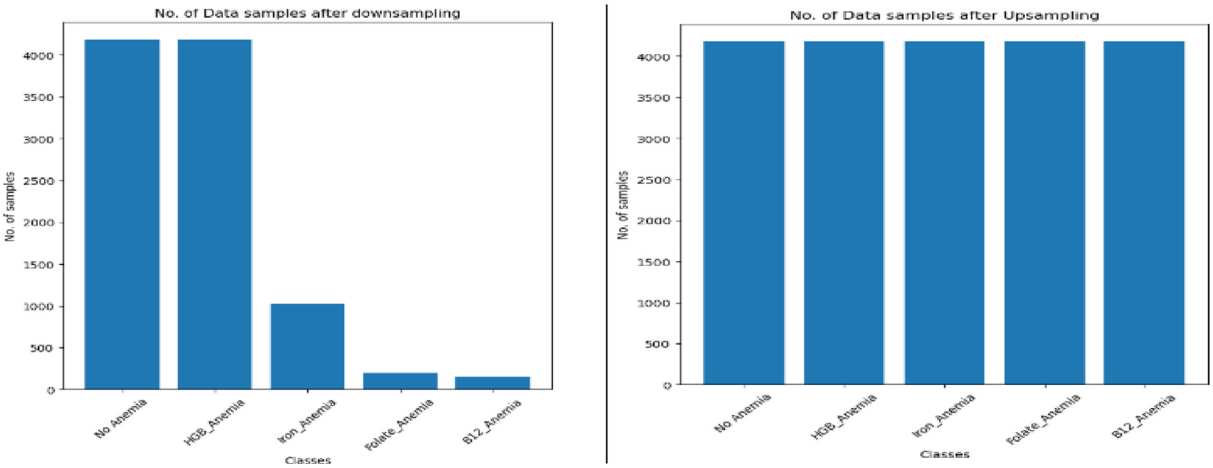


Fig. 3. Shows a comparison between data after down sampling and after up sampling.

Binary classification. Unlike in our research, we have achieved higher accuracy than them, multi classification and we have used Stacking technique.

The Eyes-defy-anemia dataset (Dimauro et al., 2023) contains 218 images of eyes, specifically conjunctivas, for research on the diagnosis or assessment of anemia based on the pallor of the conjunctiva. Despite the 19.9 % class imbalance in the dataset, the system performed well on the palpebral conjunctiva, achieving an accuracy of 83 %, a sensitivity of 69 %, and a specificity of 87 %. These results suggest that the automated anemia diagnosis is impacted by the class divide.

One of the advantages of this research is that it uses a different number of algorithms such as (svm-ANN-DT-KNN-LS), but there are some disadvantages such as Size of data is small, Low Accuracy, and Binary classification. But in our work, we used a multiple classification consisting of five classes We achieved high accuracy up to 95.95 %

During the 6-month period from December 1, 2019 to May 31, 2020, 1 738 759 English tweets were collected and analyzed. In implementing Twitter's free streaming application programming interface (API), the terms 'I have anemia', 'anemia', and 'causes of anemia' were used to explore the relevant tweets. Next, we perform some data preparation tasks to prepare the data for the analysis phase. In order to test the proposed method, several classifiers were used. Forecasting accuracy was increased to 98.96 % using Sequential Minimum Optimization (SMO) (Sarsam et al., 2022). Advantages: they used k-means algorithm and Latent Dirichlet Allocation (LDA) algorithm to cluster similar tweets and identify hidden disease topics and high accuracy almost up to 98.96 %. Disadvantage: Binary classification.

Three state health centers in Lima gathered the information with the permission of each facility by receiving over 300 initial samples from newborn infants. The best results were obtained from a threshold of Hb = 11, giving a sensitivity of 77.58 % (Saldivar-Espinoza et al., 2019).

One of the advantages of this research is that it uses 7-layer CNN of 300 samples from young children's Conjunctiva and the number of data is appropriate, On the other hand, there are

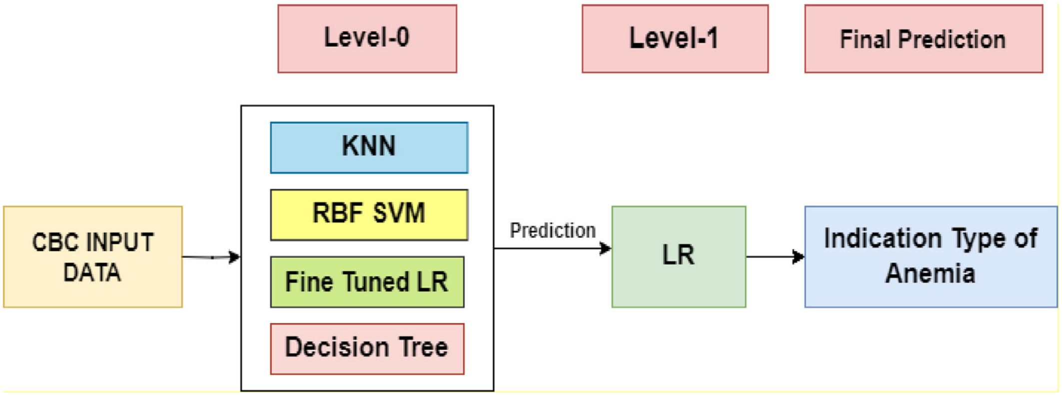
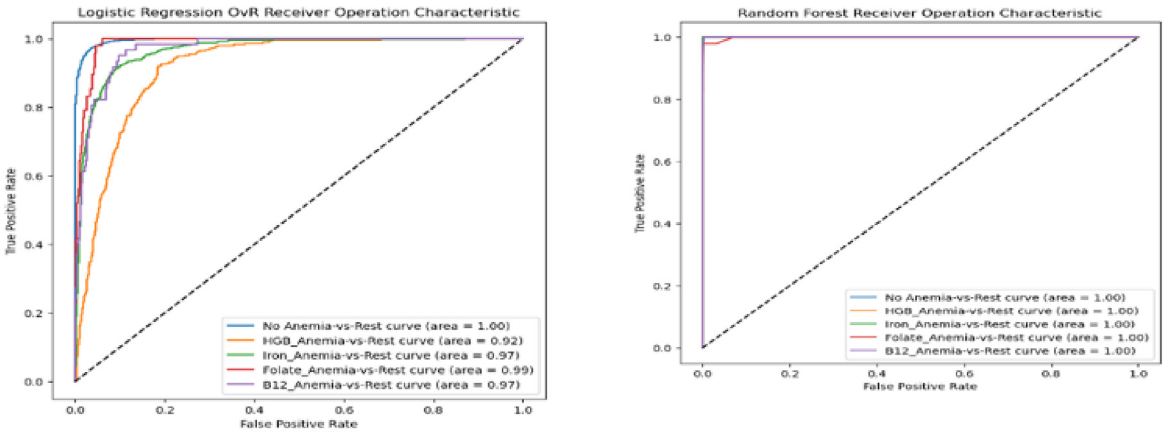
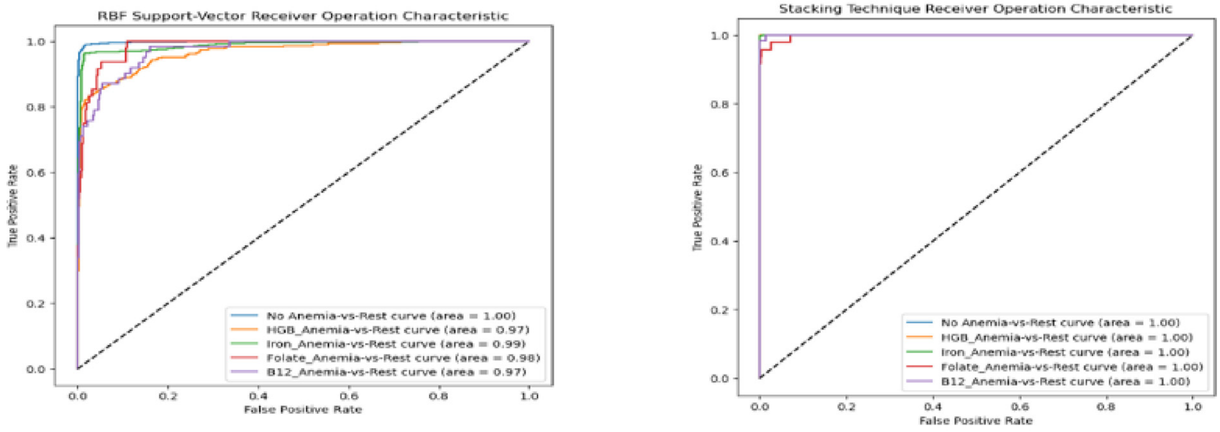


Fig. 4. Stacking-based ensemble framework for anemia susceptibility modeling. RBF support vector machine random forest, logistic regression.



(a) ROC of LR

(b) ROC of RF



(c) ROC of RBF SVM

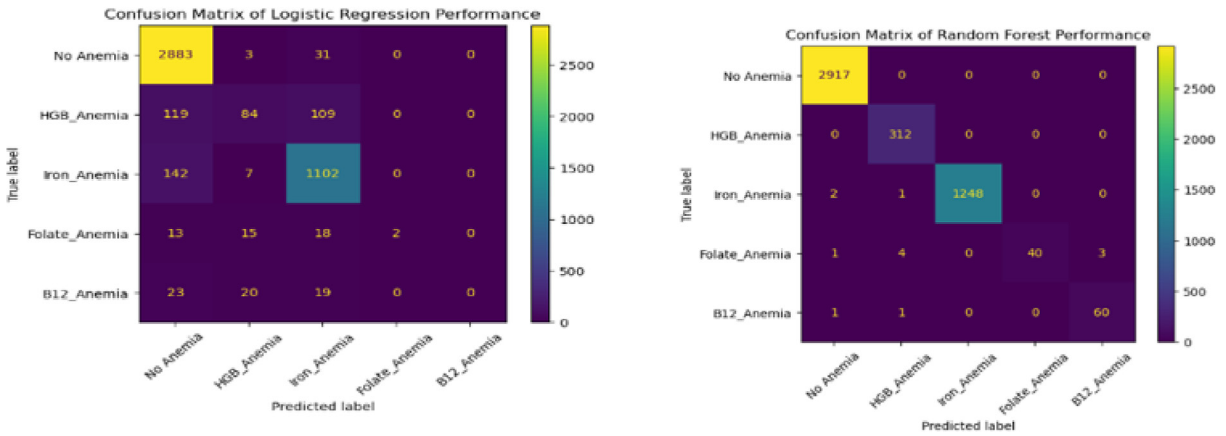
(d) ROC of Proposed Model

Fig. 5. Receiver operating characteristics for different models.

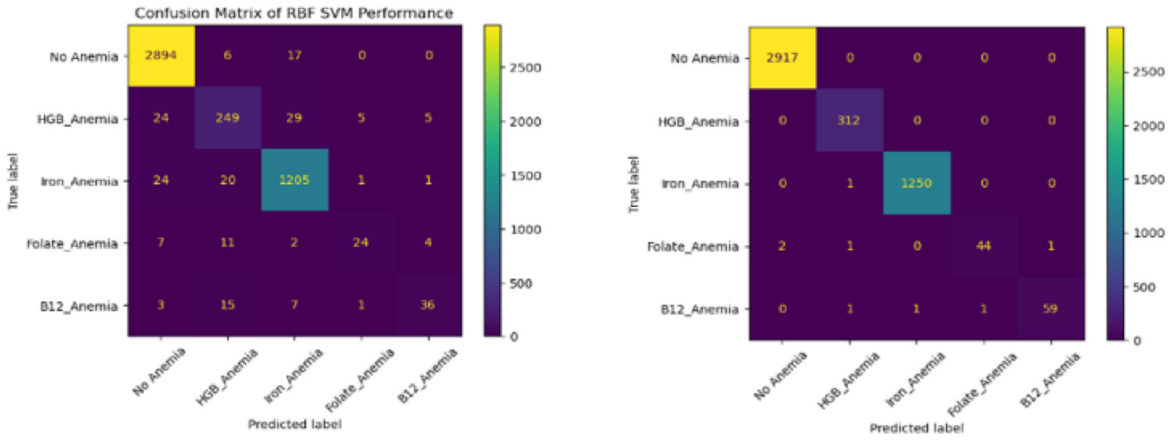
disadvantages: the data set used is not large enough to provide high accuracy and best results almost up to) 42.96 %) and binary classification.

The model was applied to 135 images of eyes with clear lower eyelids for training and validation. Deep learning algorithms were applied because it was challenging to semantically partition the conjunctiva due to fluid, nerves, wrinkles, light reflection and shadows in the eye area The UNBCSM architecture uses RGB images in JPEG format fed to the 'input image' block and manually segmented masks in NumPy array format as the training platform (Kasiviswanathan et al., 2020) have the following advantages: they perform conjunctival region segmentation for non-invasive anemia detection applications using deep learning. The remainder of the document is divided as follows. The proposed U-Net-based conjunctival segmentation model (UNBCSM) uses a fine-tuned U-Net architecture for efficient semantic conjunctival segmentation from captured digital eye images by consumer cameras in non-environments. Disadvantages: they used binary classification to know anemia or not anemia. The proposed method is discussed in detail in part 2, part 3 focuses on evaluating and discussing the experimental results Which consists of 5 classes using stacking technique. The accuracy reaches up to 99.95 %. To sum up the results of the study, please refer to section 4. We combined the algorithms in our model as described above due to problems with each method used separately, such as inefficiency, low precision, rec all, f1 score, etc and the inability to clearly detect the disease early. As a result, our proposed model performed better than previous models in terms of efficiency, allowing for easy disease identification and early disease prediction (Table 2).

We used KNN bagging, LR, Decision Trees, RBF SVM, and Random Forest. After conducting many



(a) Confusion Matrix of LR (b) Confusion Matrix of RF



(c) Confusion Matrix of RBF SVM (d) Confusion Matrix of Proposed Model

Fig. 6. Confusion metrics used by various models.

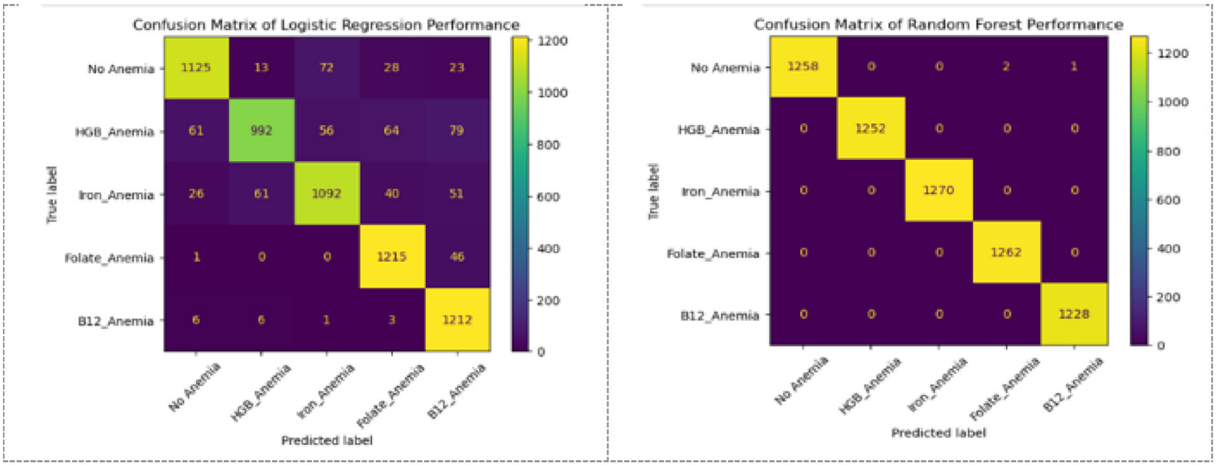
experiments, we found that the proposed model has the highest accuracy because the superposition technique is important in evaluating anemia sensitivity. This method is based on several machine learning algorithms and was first used to diagnose anemia. With meta-learners at higher levels and heterogeneous base learners at lower levels, the technique has a two-level structure. It is important to emphasize that the benefits of the superposition method have been studied by many different professions.

3. Materials and methods

The proposed analysis pipeline, that includes preprocessing, hand-crafted feature selection, classification, and parameter setup among its several analysis stages. The suggested model's structure is presented as a block diagram, as shown in Fig. 1, and is discussed in the subsections that are below.

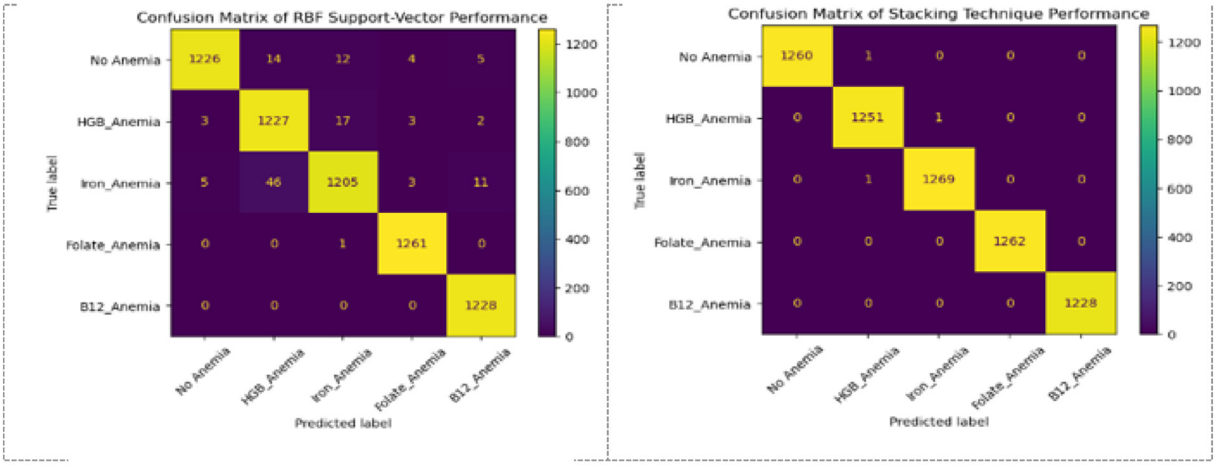
3.1. Datasets

The anemia dataset used in this paper were collected from the Tokat Gaziosmanpaşa University in Turkey's Faculty of Medicine (Kilicarslan et al., 2021) The statistics include the complete blood count test results from 15 300 patients. From the five-year span between 2013 and 2018. Children, cancer patients, and women who were pregnant were not included in the study. The dataset had 24 features and 5 classes that were used in the study. The data is imbalanced as shown in Fig. 2. Anemia is related to (hemoglobin (HGB) - Folate - TSD - Ferrite - B12) so we will take them into account, with highly correlated features to them Given that the dataset's parameter differences were relatively large, This dataset contains 10 379 of whom were women and 4921 of them were men. The dataset includes 9747 (64 %) patients without anaemia, 1019 (7 % of patients) have HGB-anemia, 4182 (27 %)



(a) Confusion Matrix of LR

(b) Confusion Matrix of RF



(c) Confusion Matrix of RBF SVM

(d) Confusion Matrix of Proposed Model

Fig. 7. Confusion metrics used by various models.

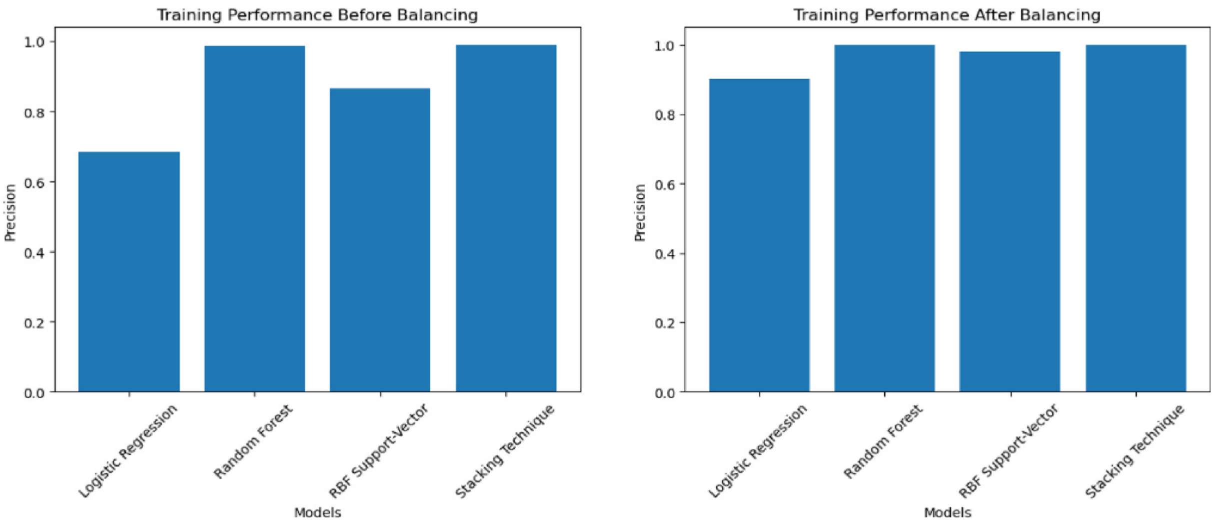


Fig. 8. Comparison between Precision before and after balancing.

Table 4. Comparison lesson performance metrics for our proposed model with different models in imbalanced data.

Model	ACC	PER	REC	F1-score
Logistic Regression	80.86 %	68.38 %	43.60 %	45.55 %
Random Forest	99.71 %	98.64 %	95.97 %	97.15 %
RBF Support-vector machine	96.03 %	86.41 %	76.68 %	80.65 %
Proposed model	99.82 %	99 %	97.34 %	98.15 %

Table 5. Comparison area under the curve (AUC) for our proposed model with different models in imbalanced data.

ALL classes	AUC (Logistic Regression)	AUC (Random Forest)	AUC (svm)	AUC (Stacking Technique)
No anemia	1	1	1	1
HGB anemia	0.92	1	0.97	1
Iron anemia	0.97	1	0.99	1
Folate anemia	0.99	1	0.98	1
B_{12} anemia	0.97	1	0.97	1

Table 6. Comparison lesson performance metrics for our proposed model with different models in balanced data.

Models	Accuracy	PER	REC	F1-score
LR	89.84 %	90.00 %	89.8 %	89.73 %
RF	99.95 %	99.95 %	99.95 %	99.95 %
RBF Support-vector machine	97.99 %	98.00 %	98.00 %	97.99 %
Proposed model	99.95 %	99.95 %	99.95 %	99.95 %

Table 7. Comparison area under the curve (AUC) for our proposed model with different models in balanced data.

ALL classes	AUC (Logistic Regression)	AUC (Random Forest)	AUC (svm)	AUC (Proposed model)
No anemia	1	1	1	1
HGB anemia	0.75	1	0.99	1
Iron anemia	0.96	1	0.99	1
Folate anemia	0.99	1	1	1
B_{12} anemia	0.96	1	1	1

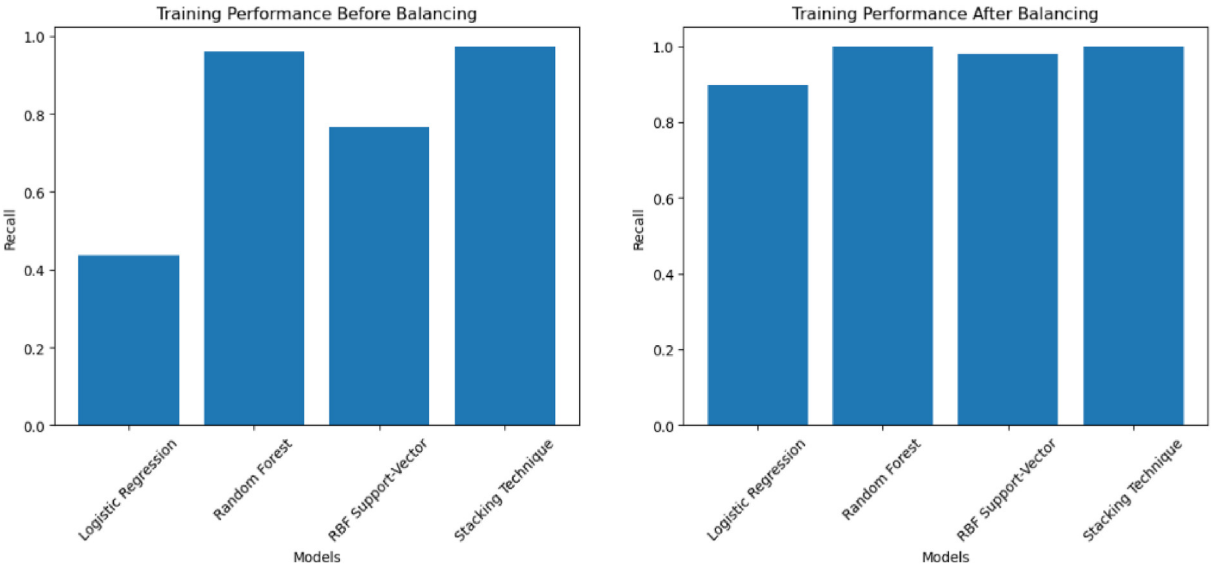


Fig. 9. Comparison between recall before and after balancing.

Table 8. Comparison of the suggested pipeline's efficiency with previous literature methodologies.

Method	Methodology	#Classes	ACC	REC	PER	AUC	F1-score
Saldivar-Espinoza et al., 2019	CNN	2	42.96 %	–	–	–	–
Geetha et al., 2022	CBC	2	97.16 %	–	–	–	–
Sarsam et al., 2022	1 738 759 English tweets	2	98.96 %	–	96 %	–	–
Saputra et al., 2023	CBC	7	99.21 %		9.30 %	–	99.84 %
Dimauro et al., 2023	218 pictures of eyes on anemia	2	83 %				–
Suggested model	CBC	5	99.95 %	99.95 %.	99.95 %	1	99.95 %.

have iron deficiency, and 199 (1 % of patients) have B12 deficiency, and 153 (1 %) patients with folate deficit as shown in Table 3.

3.2. Preprocessing

Preprocessing is the data cleaning step helps improve input data to work in the field of machine learning. It also shows the importance of the problem. First of all, it ensures the applicability of a particular model, especially when tested on datasets separate from those used for training. Preprocessing also speeds up and improves network performance by lowering data noise and/or errors. In our work, we found a problem, The data is imbalanced which means it will effect the performance of learning for each class, especially 'Folate Anemia' and 'B12 Anemia' Classes, So we will give use some of technique such as L2 Normalization.

$$d(p,q)=\sqrt{\sum_{i=1}^{n}(q_i-p_i)^2} \tag{1}$$

where,

P, q = two points in Euclidean space n.

qi, pi = Euclidean vector, starting from the origin of space (initial point).

n = n-space.

It two Shots as shown in Fig. 3: First Shot: try to train a fine-tuned models to learn the imbalanced data by using 'Logistic Regression' (one-vs-rest), fine-tuned 'Random Forest' fine tuned LR, 'Decision Tree', 'RBF Support-Vector machine' and proposed model. Second Shot: try to balance the dataset and go for training, the data will be balanced using 'Down-Sampling' so,'No Anemia Class' has many data samples compared with the others, then we use 'SMOTE UP-Sampling' to increase the data samples for each Class. Note that this resampling method does not have an absolute advantage over other sampling methods. The application of these three methods depends on the use case to which they apply and the data set itself. A combination of oversampling and under sampling is also often effective.

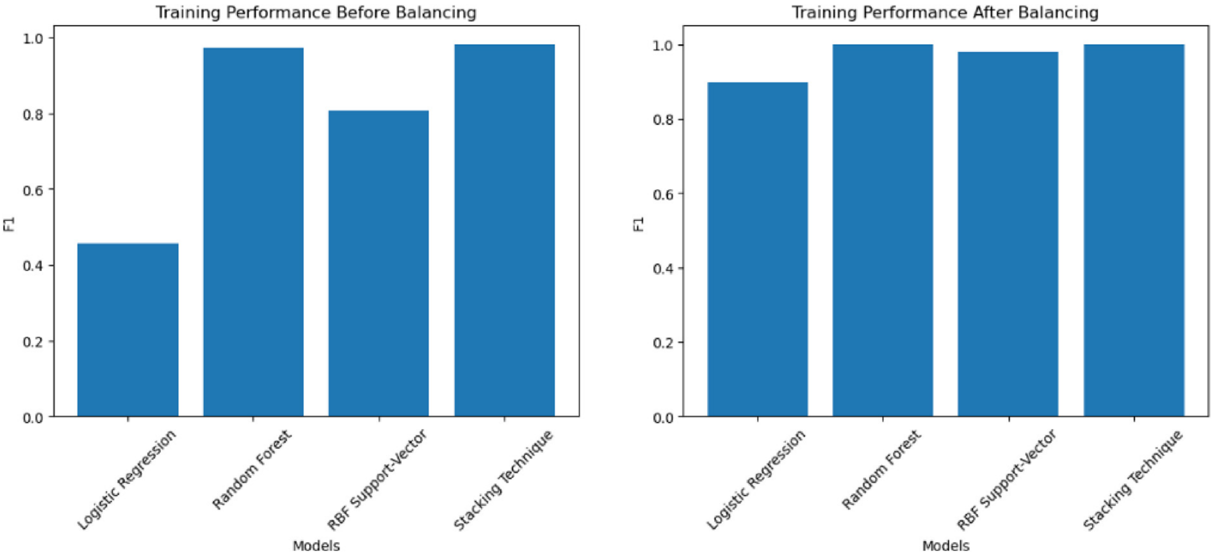


Fig. 10. Comparison between F1-Score before and after balancing.

3.3. Proposed model

Ensemble learning is a combinatorial learning method in which different analyzes intelligently interact to produce better, more accurate, and more stable results than could be achieved by a single analysis. The integration of data fusion, data modeling, and data mining into a single system is the goal of ensemble learning. These learning features were previously employed to produce subpar predicting outcomes when using several learning algorithms. Ensemble learning originates from a sequence of features with a choice of initial transformations. Additionally, ensemble learning combines data from these findings to provide adaptive knowledge discovery and prediction through voting systems. The stacking ensemble method was first introduced by Wolpert (1992). Unlike the majority of ensemble learning methods now in use, stacking makes use of a variety of algorithms with meta-learning capabilities (Zhou, 2012). Level 0 and level 1 are the two levels of the stacking structure, and (level-1) meta-learner level aggregates (level-0) outputs from numerous base learners. A rough drawing of the stacking structure created for this experiment is shown in Fig. 4. The needs of the base learners were met by SVM, kNN, LR, and Decision Tree These machine learning methods are commonly used to model anemia susceptibility (Ting and Witten, 1999).

Once all the basic learning algorithms have been prepared, they will be integrated into a complete framework using the superposition method. Suppose the original data set D consists of examples di=(xi, Yi), where xi indicates number of features, and Yi indicates the types of anemia). I∊ [1, K], where N represents the total number of model datasets. Basic learning algorithms such as RF, LR, RBF SVM are denoted as Lt (t = 1, 2, 3, 4). First, the data set D is divided multiple times into two disjoint subsets; one is used to train the base learning algorithms to produce level 0 classifiers, denoted Ht:

$$H_t^i = I_t\,(d - d_i)\gamma \mathrm{i} = 1, 2, \ldots\ldots., \mathrm{n};\, \gamma \mathrm{t} = 1, 2, 3, 4 \tag{2}$$

where,

di=(xi, Yi)

xi indicate number of features.

Yi indicates types of anemia.

N represents the total number of model datasets.

The remaining examples are used to make predictions (z_{it}) using trained classifiers:

$$Z_{it} = H_t^i(x_i) \tag{3}$$

These results from level 0 classifiers as well as their actual classification form a new basis dataset D0=((zit, zit, …. _, zit), Yi), and then sent to level 1 for meta-learner (LR) training. Therefore, LR can aggregate the classification results from the base learners to make final predictions for new instances.

$$YX = IR(h1(x), h2(x), h3(x), h4(x)) \tag{4}$$

Then each predictor is defined as a function hi that computes the predictions of the given input data at iteration h, where (h = 1; k): which LR: Logistic Regression.

Based on four unique candidate algorithms and a stacking technique, an initial ensemble model was constructed, called LR-RF-RBF SVM.

4. Results and discussion

4.1. Evaluation metrics

Generally speaking, the following metrics may be used to assess the performance of a categorization model: classifier precision F1 score for classification precision. All of these measures depend on four expected outcomes: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). In our specific model, TP represents anemic patients correctly identified by our system, FP represents healthy subjects identified as patients, TN represents Healthy subjects were correctly identified and FNs were anemic patients classified as healthy subjects. Based on these results, various evaluation parameters can be measured. For example, classification accuracy represents the number of correctly classified labels divided by the total number of classified labels. Recall (specificity) is another metric that calculates the proportion of positive (negative) samples that are correctly classified. Correctly predicted positive models from the total number of predicted models in the positive class are estimated using the accuracy index. Finally, the degree of agreement between REC and PER values is described by the F1 score. These measures are respectively defined mathematically by as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where,

ACC= Accuracy.

TP = True positive.

TN = True negative.

FP= False positive.

FN= False negative.

Which is the common assessment metric in use, This is the proportion of correctly predicted outcomes in each instance.

$$PER = \frac{TP}{TP + FP} \tag{6}$$

One indicator of model performance is the accuracy or quality of successful predictions. It shows how often an ML model is correct when predicting the target class, it is used to calculate the ratio of the total number of positive scores predicted by the classification algorithm in a positive class divided by the number of correct positive scores.

$$REC = \frac{TP}{TP + FN} \tag{7}$$

It is a model's ability to find all relevant instances in a data source used to calculate the percentage of positive patterns that are classified properly.

$$F1{-}score = \frac{2*PER*REC}{PER + REC} \tag{8}$$

The harmonic mean of the F1 score combines memory and precision. To maximize the F1 score, recall and precision must both be maximized concurrently. To successfully handle the distribution issue, F1-score is applied. When the dataset contains classes with imbalances, it is helpful. It represents the cumulative mean between recall and precision values. In addition to accuracy measurements, receiver operating characteristic (ROC) curves were also used to validate and support the robustness and accuracy of our deep learning model. ROC describes the classification results of a model based on the ratio of true positives and false positives at different classification thresholds. Quantitatively, an additional evaluation metric is often used in classification systems using the area under the curve (AUC) of the ROC. The AUC value illustrates the model's performance in the discriminant classes. The AUC function can be expressed as in Equation. (9):

$$AUC = \frac{Sp - Np(Nn+1)/2}{NpNn} \tag{9}$$

where,

AUC = area under the curve.

PS = the sum of all the positive examples ranked.

P_{n}, N_{n} signifies the number of positive and negative examples.

4.2. Simulation setup

Thus, we utilized Kaggle and the open-source software Python 3.0 to carry out the classifier's training and validation, and the run time was 1524 s. Python programming was used to developed the model. In particular, we used Scikit-learn, a Python library for machine learning. They are simple and effective tools for data mining and data analysis. This framework is built on several popular Python packages, specifically NumPy, SciPy, and matplotlib. Five categories were created from the entire sample of scans: training (70 % of scans), testing (30 % of scans). Proposed model was employed and the result of the lesson performance metrics compared with other models are given in Table 6. As is clear. Additionally (Figs. 6–8), show the accuracy, confusion matrices, ROC curves for several models. The proposed model in imbalanced data. Achieved ACC, PER, REC, and F1-score of 99.82,99, 97.34, and 98.15 %, respectively illustrated in Table 4. This table shows that our proposed method achieves maximum efficiency. In addition to accuracy measurements, ROC were also used to verify and validate the robustness and accuracy of our system. The ROC describes the classification results of the model based on the true positive and false positive rates, and the AUC of the ROC is used as an additional evaluation metric, as shown in Fig. 5. Table 4 shows the AUC of all against the type of classes. Table 5 shows the comparison of AUC for our proposed model with different models in imbalanced data. As we can see in this table to highlight the advantage of the proposed model, we compared its performance with the base model before balancing in Table 4, Fig. 5 shows that the proposed model reaches the highest-level ACC, PER, REC, and F1-score after balancing (Table 5).

As described in Section 3., additional evaluation is performed using various model methodologies in order to select the ideal parameter value. Table 6 displays the precision of the test findings. Depending on the model, different performance criteria are suggested for classifying anemia. Fig. 8 shows the before-and-after balance curves for the REC, PER, F1-Score, and Training Performance for the different models, with the proposed model achieving the highest accuracy (99.95 %).

To achieve the best level of accuracy, our pipeline is trained using multiple ML algorithms, including LR, RBF SVM, and RF Table 7. The most accurate model, stacking, had a high accuracy of 99.95 % and employed the same hyperparameters. The confusion matrices and PER, REC, and F1-SCORE performance before and after balancing for the various pretrained models and Fig. 10 shows the comparison between F1-score before and after balancing are shown in Figs. 7–10, respectively. We compared our approach with several Classes using the CBC dataset and DL-based methods as well. Table 7 displays such a comparison of AUC using several models (LR-RF-RBF SVM) (Table 8).

The proposed pipeline showed 99.95 % accuracy, 99.95 % recall, 99.95 % precision, and 100 % ROC, respectively. A data collection of 15 300 patients was used for testing and validation. By comparing our model's performance to that of well-known machine learning, we were able to confirm higher multi-class classification. In our upcoming study, we will mix several datasets using advanced methods to improve the efficiency and performance.

5. Discussion

In this paper, we faced many problems while trying to diagnose the type of anemia. The first problem was the data imbalance which affected the performance of learning for each type, especially 'Folate Anemia' and 'B12 Anemia' Classes. Therefore, we used several techniques such as L2 Normalization to solve this problem. Our solution was to make two shots the first one: try to train a fine-tuned models to learn the imbalanced data by using 'LR' (one-vs-rest), fine-tuned 'RF' fine-tuned LR, 'DT','RBF-SVM' and Proposed model. Second Shot: try to balance the dataset and go for training. The data was balanced using 'Down-Sampling' so that, 'No Anemia Class' has many data samples compared with the others. We applied 'SMOTE UP-Sampling' to increase data samples for each Class. We noted that this resampling method does not have an absolute advantage over other sampling methods. The application of these methods depends mainly on the data set. A combination of oversampling for some classes and undersampling for other classes were very effective solution. We illustrated through previous figures. Comparisons between results of our suggested model after and before balancing. Results shown in Fig. 8 proved that our suggested model demonstrated the highest accuracy. Other ways to evaluate the performance of our model are confusion matrices that shown in Figs. 6 and 7 and ROC curves shown in Fig. 5. Moreover, we compared results from our model after Balancing that Stacking the best result shown in Table 6.

6. Conclusion

This paper has presented an AD system that can differentiate five types of anemia by applying machine learning-based pipeline schemes. We proved that these schemes can accurately detect multiple classes of anemia. We concentrated on LR, RF, RBF support-vector machine and Stacking schemes. The proposed pipeline methodology showed 99.95 % accuracy, 99.95 % recall, 99.95 % precision, and 100 % ROC, respectively. A data collection of 15 300 patients was used for testing and validating. By comparing the results from our models' performance to that of well-known machine learning techniques, we found that the proposed model was able to confirm higher multi-class classification. In the upcoming studies, we'll try to mix several datasets using advanced methods to improve the efficiency and performance of the proposed model (Fig. 10).

Author credit statement

Authored by Eng. A M E-B, teaching assistant at AL Zarka Academy of Management Science and Computer. Supervised by: Prof. S K, Professor, Dept. of Electronics and Communications Engineering, Faculty of Engineering, Mansoura University. Assoc. Prof. M M A, Associate professor, Dept. of Electronics and Communications Engineering, Faculty of Engineering, Mansoura University. Asst. Prof. E A: Assistant Professor, Dept. of Electronics and Communications Engineering, Faculty of Engineering, Mansoura University.

Authors contribution

Conception or design of the work: S K, M M A, E A, A M. E-B (20 %/30 %/20 %/30 %). Data collection and tools: S K, M M A, E A, A M. E-B (20 %/20 %/20 %/40 %). Data analysis and interpretation: S K, M M A, E A, A M. E-B (20 %/25 %/25 %/30 %). Methodology: S K, M M A, E A, A M. E-B (40 %/15 %/15 %/30 %). Project administration: S K, M M A, E A, A M. E-B (30 %/20 %/20 %/30 %). Software: S K, M M A, E A, A M. E-B (20 %/25 %/25 %/30 %). Drafting the article: S K, M M A, E A, A M. E-B (25 %/20 %/20 %/35 %). Critical revision of the article: S K, M M A, E A, A M. E-B (20 %/25 %/25 %/30 %).

Conflicts of interest

Authors have no conflict of interest to declare.

References

Çil, B., Ayyõldõz, H., Tuncer, T., 2020. Discrimination of β-thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system. Med. Hypotheses 138, 109611.

Cifra, C.L., Custer, J.W., Singh, H., Fackler, J.C., 2021. Diagnostic errors in pediatric critical care: a systematic review. Pediatr. Crit. Care Med. 22, 701.

Cuomo, R., Cargiolli, M., Cassarano, S., Carabotti, M., Annibale, B., 2018. Treatment of diverticular disease, targeting symptoms or underlying mechanisms. Curr. Opin. Pharmacol. 43, 124−131.

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. Artif. Intell. Med. 136, 102477.

El-kenawy, E., SM, E., 2019. A machine learning model for hemoglobin estimation and anemia classification. Int. J. Comput. Sci. Inf. Secur. 17, 100−108.

Esteban-Medina, M., Peña-Chilet, M., Loucera, C., Dopazo, J., 2019. Exploring the druggable space around the Fanconi anemia pathway using machine learning and mechanistic models. BMC Bioinf. 20, 1−15.

Ewusie, J.E., Ahiadeke, C., Beyene, J., Hamid, J.S., 2014. Prevalence of anemia among under-5 children in the Ghanaian population: estimates from the Ghana demographic and health survey. BMC Publ. Health 14, 1−9.

Felicísimo, Á.M., Cuartero, A., Remondo, J., Quirós, E., 2013. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10, 175−189.

Geetha, V., Gomathy, C.K., Keerthi, K., Pavithra, N., 2022. Diagnostic approach to anemia in adults using machine learning. J. Pharm. Negat. Results 13, 3713−3717.

Gomathy, C.K., Geetha, V., Praneetha, G., 2022. Medicine identification using open Cv. J. Pharm. Negat. Results 53, 4046−4072.

Gunčar, G., et al., 2018. An application of machine learning to haematological diagnosis. Sci. Rep. 8, 411.

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. Springer, pp. 463−469.

Karagül Yõldõz, T., Yurtay, N., Öneç, B., 2021. Classifying anemia types using artificial learning methods. Eng. Sci. Technol. Int. J. 24, 50−70.

Kasiviswanathan, S., Bai Vijayan, T., Simone, L., Dimauro, G., 2020. Semantic segmentation of conjunctiva region for non-invasive anemia detection applications. Electronics (Basel) 9, 1309.

Kassebaum, N.J., et al., 2014. A systematic analysis of global anemia burden from 1990 to 2010. Blood J. Am. Soc. Hematol. 123, 615−624.

Kilicarslan, S., Celik, M., Sahin, Ş., 2021. Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification. Biomed. Signal Proc. Contr. 63, 102231, 2022, [Online]. Available: https://data.mendeley.com/datasets/dt89jydgnv/1 mia disease classification.

Korup, O., Stolle, A., 2014. Landslide prediction from machine learning. Geol. Today 30, 26−33.

McLean, E., Cogswell, M., Egli, I., Wojdyla, D., De Benoist, B., 2009. Worldwide prevalence of anaemia, WHO vitamin and mineral nutrition information system, 1993−2005. Public Health Nutr. 12, 444−454.

Milman, N., 2011. Anemia—still a major health problem in many parts of the world. Ann. Hematol. 90, 369−377.

Murphy, J.F., 2011. Haemoglobin Concentrations for the Diagnosis of Anaemia and Assessment of Severity. Vitamin and Mineral Nutrition Information System. World Health Organization, Geneva, p. 2002.

Oguntimilehin, A., Adetunmbi, A.O., Olatunji, K.A., 2014. A machine learning based clinical decision support system for diagnosis and treatment of typhoid fever. In: A Machine Learning Based Clinical Decision Support System for Diagnosis and Treatment of Typhoid Fever, 4, pp. 1−9.

Oh, H.-J., Lee, S., 2017. Shallow landslide susceptibility modeling using the data mining models artificial neural network and boosted tree. Appl. Sci. 7, 1000.

Qasem, S.N., Mosavi, A., 2020. Novel Meta-Heuristic Model for Discrimination between Iron Deficiency Anemia and B-Thalassemia with CBC Indices Based on Dynamic Harmony Search. DHS.

Rashidi, H.H., Tran, N.K., Betts, E.V., Howell, L.P., Green, R., 2019. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. Acad. Pathol. 6, 2374289519873088.

Rokach, L., 2009. Taxonomy for characterizing ensemble methods in classification tasks: a review and annotated bibliography. Comput. Stat. Data Anal. 53, 4046−4072.

Saldivar-Espinoza, B., Núñez-Fernández, D., Porras-Barrientos, F., Alva-Mantari, A., Leslie, L.S., Zimic, M., 2019. Portable System for the Prediction of Anemia Based on the Ocular Conjunctiva Using Artificial Intelligence. arXiv preprint arXiv:1910.12399.

Saputra, D.C.E., Sunat, K., Ratnaningsih, T., 2023. A new artificial intelligence approach using extreme learning machine as the potentially effective model to predict and analyze the diagnosis of anemia. In: Healthcare. MDPI, pp. 1−9.

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. Artif. Intell. Med. 134, 102428.

Sennaar, K., 2019. Machine Learning for Medical Diagnostics-4 Current Applications. URL: https://emerj.com/ai-sector-overviews/machine-learningmedical-diagnostics-4-current-applications (дата обращения 22.10. 2019).

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naive Bayes Models. Math. Probl Eng. 2012, 175−189.

Ting, K.M., Witten, I.H., 1999. Issues in stacked generalization. J. Artif. Intell. Res. 10, 271−289.

Vohra, R., Hussain, A., Dudyala, A.K., Pahareeya, J., Khan, W., 2022. Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting. PLoS One 17, 7.

Wang, L.-J., Sawada, K., Moriguchi, S., 2013. Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. Comput. Geosci. 57, 81−92.

WH Organization, 2014. Serum Transferrin Receptor Levels for the Assessment of Iron Status and Iron Deficiency in Populations. World Health Organization.

Wolpert, D.H., 1992. Stacked generalization. Neural Netw. 5, 241−259.

Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13, 839−856.

Zhou, Z.-H., 2012. Ensemble Methods: Foundations and Algorithms. CRC Press.