I. J. Education and Management Engineering, **2024, 1, 1-15**
Published Online on February 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijeme.2024.01.01

Machine Learning Algorithms for Iron Deficiency Anemia Detection in Children Using Palm Images

Stephen Afrifa*
Department of Information and Communication Engineering, Tianjin University, Tianjin 300072, China
Department of Information Technology and Decision Sciences, University of Energy and Natural Resources, Sunyani 00233, Ghana
Email: afrifastephen@tju.edu.cn
ORCID iD: https://orcid.org/0000-0002-9308-3137
*Corresponding Author

Peter Appiahene
Department of Information Technology and Decision Sciences, University of Energy and Natural Resources, Sunyani 00233, Ghana
Email: peter.appiahene@uenr.edu.gh
ORCID iD: https://orcid.org/0000-0002-6098-4537

Tao Zhang
Department of Information and Communication Engineering, Tianjin University, Tianjin 300072, China
Email: zhangtao@tju.edu.cn
ORCID iD: https://orcid.org/0000-0003-2317-644X

Vijayakumar Varadarajan
School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia
International Divisions, Ajeenkya D. Y. Patil University, Pune 412105, India
School of Information Technology, Swiss School of Business Management, 1213 Geneva, Switzerland
Email: v.varadarajan@unsw.edu.au
ORCID iD: https://orcid.org/0000-0003-3752-7220

Received: 29 May, 2023; Revised: 01 August, 2023; Accepted: 28 August, 2023; Published: 08 February, 2024

Abstract: Anemia is a common condition among adults, particularly in children and pregnant women. Anemia is defined as a lack of healthy red blood cells or hemoglobin. Early identification of anemia is critical for excellent health and well-being, which contributes to the sustainable development goals (SDGs), notably SDG 3. The intrusive way to detecting anemia has several hurdles, including anxiety and cost, which impedes health development. With the advent of technology, it is critical to create non-invasive techniques to diagnose anemia that can minimize costs while also improving detection efficacy. A distinct non-invasive technique is developed in this study employing machine learning (ML) models. This study's dataset contains 4260 observations of non-anemic (0) and anemic (1) children. To train the dataset, six (6) different ML models were employed: k-Nearest Neighbor (KNN), decision tree (DT), logistic regression (LR), nave bayes (NB), random forest (RF), and kernel-support vector machine (KSVM). The DT and RF models obtained the highest accuracy of 99.92%, followed by the KNN at 98.98%. The ML models used in this study produced substantial results. The models also received high marks on performance evaluation metrics such as accuracy, recall, F1-score, and Area Under the Curve-Receiver Operating Characteristics (AUC-ROC). When compared to the other ML models, the DT and RF had the best precision (1.000), recall (0.9987), F1-score (0.9994), and AUC-ROC (0.9994) ratings. According to the findings, ML models are crucial in the detection of anemia using a non-invasive technique, which is critical for health facilities to boost efficiency and quality healthcare. Various machine learning models were used in this study to detect anemia in children using palm images. Finally, the findings confirm earlier studies on the effectiveness of ML algorithms as a non-invasive means of detecting iron deficiency anemia.

Index Terms: Anemia, machine learning, health, artificial intelligence, sustainable development, deep learning

1. Introduction

Anemia is a condition in which the number of red blood cells declines and the blood's oxygen-carrying capacity becomes insufficient to fulfill physiological demands [1, 2]. This condition can also be affected by altitude, smoking, age, gender, and other health factors such as pregnancy. This ailment causes anemia, or a lack of hemoglobin levels in the human body. The most prevalent cause of anemia is a lack of iron [3]. Iron is required by your body to produce hemoglobin [4]. Hemoglobin is an iron-rich protein that gives blood its red hue. Hemoglobin is responsible for transporting oxygen from the lungs to the rest of the body. Anemia comes in a variety of forms. Each has its own set of causes and therapies. Some types, such as the mild anemia that occurs during pregnancy, are not a major concern. However, some types of anemia can indicate a serious underlying medical condition. Anemia is also characterized by fatigue, weakness, pale or yellowish complexion, irregular heartbeats, shortness of breath, dizziness or lightheadedness, chest discomfort, and chilly hands and feet. Anemia is a severe global public health issue that disproportionately affects young children, menstrual teenage girls and women, pregnant and postpartum women, and pregnant and postpartum women. The World Health Organization (WHO) estimates that 40% of children aged 6 to 59 months, 37% of pregnant women, and 30% of women aged 15 to 49 are anemic globally [5]. Anemia may go unnoticed in many people owing to modest symptoms, especially in undeveloped places where there are fewer medical facilities, health care, and self-awareness. The usual method for determining whether a person is anemic or not involves measuring hemoglobin levels in the blood. A blood sample is taken from a person and the hemoglobin concentration is measured using a hematocrit test. This test is performed to determine the proportion of red blood cells in the blood.

The entire invasive process necessitates laboratory setup and testing, which takes time and may expose health professionals to blood-borne illnesses. Today, technology has advanced and has had a massive impact. The use of artificial intelligence (AI) and machine learning (ML) algorithms are becoming increasingly used for non-invasively estimating health-related data [6]. There is a need for effective non-invasive procedures for detecting anemia in order to prevent the complications caused by invasive methods. Some historical customs include checking the eye and nail pallor for an approximate estimate of a person's anemic status; however, most scientific approaches include collecting a patient's blood sample and then measuring the hemoglobin count. Invasive procedures have their own set of complications, such as a lack of attribution data, over-attribution to iron deficiency anemia, and biomarker complexity [7]. Due to cost and other administrative concerns, it is not reasonable to expect every individual in rural areas to choose a hemoglobin count test. It should be noted that various rural places have varied cultures and customs, with some considering the intrusive procedure to detect anemia to be a taboo. This makes detection in these regions challenging. Because of the risks and severity of anemia, particularly in children and pregnant women, it is critical to develop strong and quick machine learning models for non-invasive anemia detection. Numerous investigations have been undertaken with the use of non-invasive approaches such as machine learning algorithms in the identification of anemia, primarily with the use of the conjunctiva of the eyes, however the palpable palm is utilized far less frequently than the conjunctiva in most studies. This has resulted in a scarcity of research employing palpable palm images to detect anemia in a non-invasive approach. In comparison to the conjunctiva and fingernails, research has demonstrated that the use of palm images to detect anemia in a non-invasive way is significant [8]. As a result, in this study, several artificial intelligence and machine learning algorithms are employed to detect anemia in a non-invasive technique using palpable palm-based images, therefore alleviating the issues associated with the invasive method. The present paper presents the most recent method in the literature for detecting anemia in a non-invasive manner utilizing machine learning techniques. The following are summary of the expected work's contributions as adapted from [9, 10]:

- To develop a reliable, novel and robust machine learning non-invasive models to detect anemia.
- Use of palpable palm-based images for analyzing and detecting anemia through estimation of the hemoglobin level patient.
- Application of different machine learning algorithms for accurate detection of anemia based on hemoglobin levels.
- Use of different performance evaluation metrics, thus, accuracy, precision, recall, F1-score, and Area Under the Curve-Receiver Operating Characteristics (AUC-ROC) score.
- It is a novel contribution to the literature in which novel approaches for detection of anemia for effective and productivity outcomes for health workers.

The remainder of the paper is structured as follows: Section 2 provides a detailed review of previous studies that employ various machine learning models to detect anemia. Section 3 describes the suggested framework that will be used in the present study. The findings of the experiments are presented in Section 4. Section 5 contains a full discussion of the study's findings. Finally, Section 6 gives the study's general conclusion and future work.

2. Related Works

Artificial intelligence and machine learning have been used in various sectors various sectors such as health in prediction of kidney disease [11], cybersecurity [10], economic prediction of unemployment [12], network botnet detection [9], and agriculture [13]. To begin with, Appiahene et al. [14] in a study proposed models for the detection of anemia using various algorithms, which include convolutional neural network (CNN), k-Nearest Neighbor (k-NN), naïve bayes (NB), support vector machine (SVM), and decision tree (DT). Their study justified that the study supports other similar studies on the potency of the machine learning algorithm as a non-invasive method in detecting iron deficiency anemia. Their study focuses mostly on using conjunctiva images to detect anemia, but it adds to the field by advocating the use of deep learning models rather than machine learning models. In another study, Mitani et al. [15] proposed that anemia can be detected via machine-learning algorithms trained using retinal fundus images. Their study used metadata including race or ethnicity, age, sex and blood pressure of participants in their dataset. Their study showed that automated anemia screening on the basis of fundus images could particularly aid patients with diabetes undergoing regular retinal imaging and for whom anemia can increase morbidity and mortality risks. Additionally, a study by KILICARSLAN et al. [16] proposed two hybrid models using genetic algorithm (GA) and deep learning algorithms of Stacked Autoencoder (SAE) and Convolutional Neural Network (CNN) for the prediction of HGB-anemia, nutritional anemia, (iron deficiency anemia, B12 deficiency anemia, and folate deficiency anemia), and patients without anemia. Their study revealed that the performance of the proposed genetic algorithm-convolutional neural network (GA-CNN) algorithm whose layers trained separately and sequentially was found to be better than the performance of the studies proposed in the literature, by a 98.50% accuracy. It should be noted that their research adds to and supports the use of non-invasive ways to diagnose anemia in order to reduce the impact and repercussions of the condition.

Furthermore, Alzubaidi et al. [17] proposed lightweight deep learning models that classify the erythrocytes into three classes: circular (normal), elongated (sickle cells), and other blood content. To tackle this issue and optimize the performance, the transfer learning technique was utilized in their study. Their findings demonstrated that their model achieved state-of-the-art performance and exceeded the most recent approaches for detecting anemia. A study by Alsayegh et al. [18] proposed a robust multi-algorithm interpretable machine learning (ML) pipeline, and 20 relevant risk factors to fit predictive models to 4,044 patients presenting with acute coronary syndromes (ACS) between 2012 and 2013. Their model pipeline was able to elucidate ACS risk of individual patients based on their unique risk factors. They came to the conclusion that the usage of artificial intelligence technologies to identify anemia is vast and significant. Finally, Waisberg et al. [19] proposed that retinal fundus photography and deep learning may be utilized to help further understand this microgravity-induced, anemic process for future spaceflight. Their approach allowed for a non-invasive retinal photograph that can be done frequently during spaceflight as opposed to an invasive blood draw and subsequent tests.

3. Methodological Design

This section explains the proposed non-invasive anemia detection approach based on palpable palm-based images in general.

a) Data source and description

The study made use of 4260 publicly accessible dataset observations from the Mendeley data repository via "Justice Williams Asare. (2022)." <i>Anemia Detection using Palpable Palm Image Datasets from Ghana</i> [Dataset]. https://doi.org/10.17632/CCR8CM22VZ.1" accessed April 18, 2023. Clinical laboratory personnel from Komfo Anokye Teaching Hospital, Bolgatanga Regional Hospital, Kintampo Municipal Hospital, Ahmadiyya Muslim Hospital, Sunyani Municipal Hospital, Manhyia District Hospital, Ejusu Government Hospital, Seventh Day Adventist (SDA) Hospital, Nkawie-Toase Government Hospital, and Holy Family Hospital all worked together to compile the data. This study's dataset focuses on young children aged five and under. The laboratory personnel took all images, and the hemoglobin (Hb) values of the patients, age, sex, and a remark (anemic or non-anemic) based on the Hb value collected were uploaded by technicians or medical laboratory officers using standard high-quality cameras with a minimum of 12MP. Before the datasets were collected, the ethical committees from the various hospitals involved in this study approved the ethical consent. The University of Energy and Natural Resources (UENR), Sunyani's Committee for Human Research and Ethics (CHRE) provided an ethical letter for the research's conduct, with the identification CHRE/CA/042/22.

b) Preprocessing techniques

Real-world data is frequently inconsistent, which can have an impact on model performance [20]. Preprocessing data before feeding it into classifiers is an essential aspect of constructing a machine-learning model [21]. The original dataset that consists of images were converted to comma separated values (csv) files using the Squeeze function in matrix laboratory (MATLAB). The Fig. 1 depicts the data preprocessing techniques utilized in this study.

- MATLAB Squeeze function: The Squeeze command in MATLAB was used to determine the red, green and blue (RGB) percentiles. The values were stored in an array where the RGB values were extracted into an csv file. The RGB values extracted were matched with the corresponding status (anemic or non-anemic). The squeeze theorem is a mathematical proof that demonstrates the value of a limit by squeezing a hard function between two equal and known values [22]. The equations 1 and 2 below shows the Squeeze function utilized in this study.

$$h(x) \le f(x) \tag{1}$$

when x is near a, except possibly at a, and

$$\lim_{x \to a} f(x) = \lim_{x \to a} h(x) = L \tag{2}$$

All this means is that if $g(x)$ is squeezed between $f(x)$ and $h(x)$ near a, and if $f(x)$ and $h(x)$ have the same limit L at a, then $g(x)$ is trapped and must have the same limit L as well.

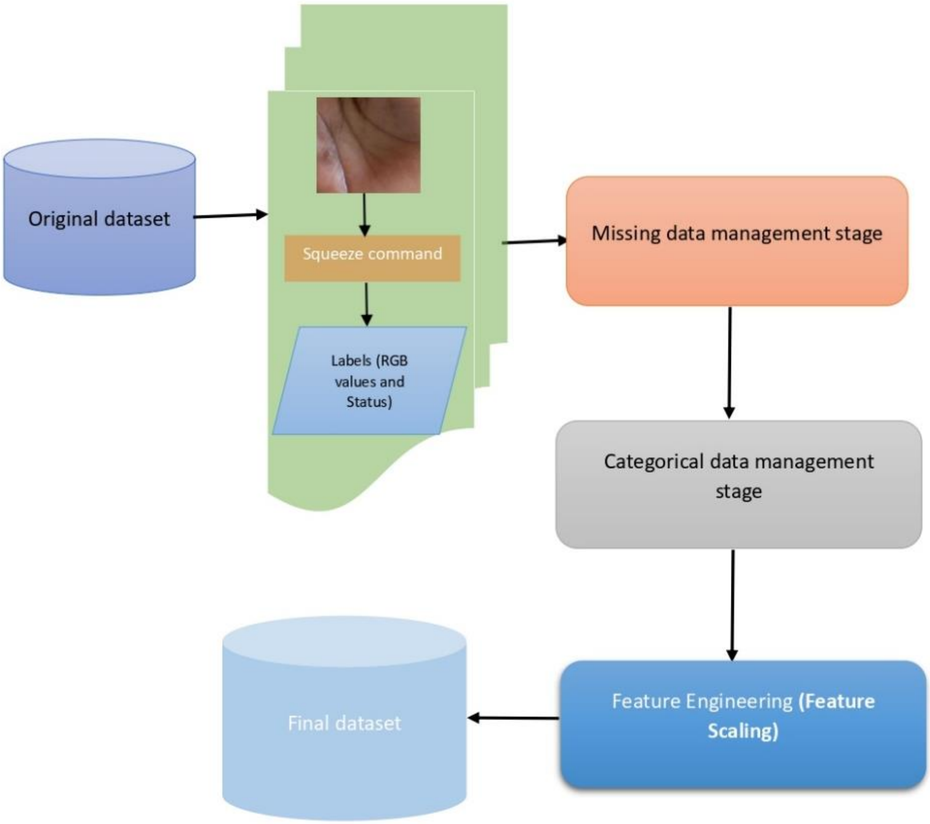

Fig. 1. Anemia disease dataset preprocessing stage.

- Handling missing values: Real-world data frequently contains a large number of missing values. Missing values can be caused by data corruption or a failure to record data [23]. Missing data management is critical during dataset preparation since many machine learning algorithms do not handle missing values. It should be noted that the data used in this study had no missing values. The clinical laboratory personnel labeled the data such that there were no missing values.
- Handling categorical values: The dependent variable was changed into the necessary format at this stage. The nominal data were transformed into numerical data in the form of 0 (non-anemic) and 1 (anemic).
- Feature Engineering: Feature engineering is an important stage in the development of accurate and successful machine learning models. Scaling, normalization, and standardization are important aspects of feature engineering since they require changing data to make it more suited for modeling. The feature scaling method was employed in the present study. Feature scaling is a data preparation technique that includes scaling down the values of features or variables in a dataset [24]. This is done to guarantee that all characteristics contribute equally to the model and that bigger values do not dominate the model.

c) Proposed model building flow

The final dataset is utilized to train the proposed models of k-Nearest Neighbor (KNN), decision tree (DT), logistic regression (LR), naïve bayes (NB), random forest (RF), and kernel-support vector machine (K-SVVM). The Fig. 2 presents the proposed model framework.

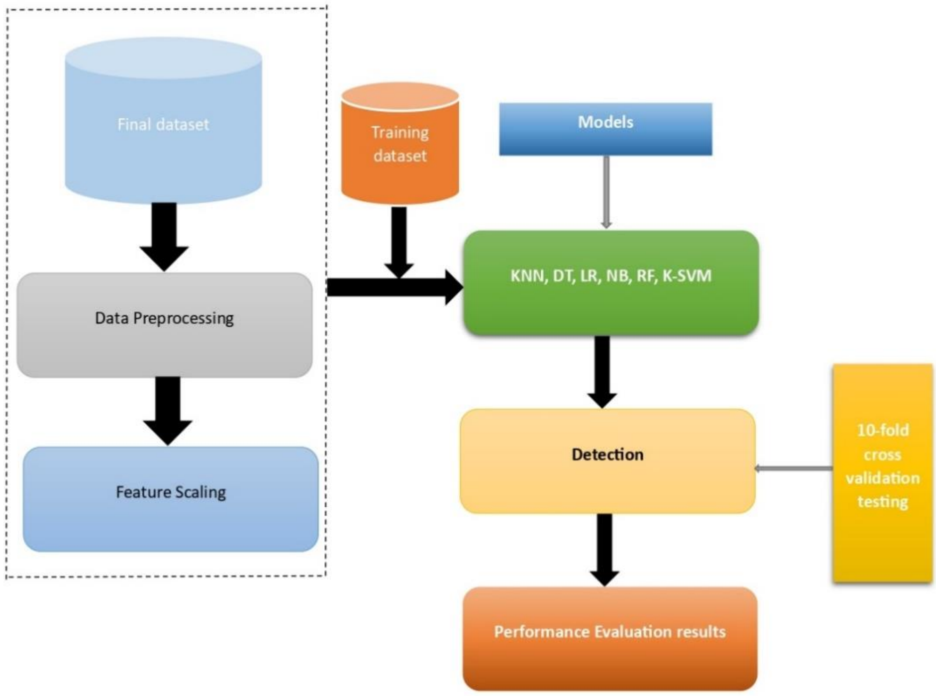

Fig. 2. Proposed model building flow.

The final dataset is separated into two parts: training (70%) and testing (30%). The dataset is used to train the various machine learning classifiers. The detection of the models is validated using 10-fold cross validation. Cross-validation is a strategy for testing ML models that involves training numerous ML models on subsets of the available input data and then evaluating them on the complementary subset [25]. To achieve the objectives, the proposed model architecture leverages transfer learning from the same domain as the target dataset as recommended to solve the shortage of training data in anemic disease classification tasks. This approach has the potential to improve performance substantially. In addition, using the transfer-learning approach, primary data that are comparable to the target task images were acquired. The proposed models' detection results are assessed using evaluation measures.

d) Machine learning models

To train the models in this study, six (6) different machine learning (ML) techniques are used. The k-Nearest Neighbor (KNN), decision tree (DT), logistic regression (LR), naïve bayes (NB), random forest (RF), and kernel-support vector machine (KSVM) are the ML models used in this study. The ML models are detailed as below:

- K-Nearest Neighbor (KNN) is a basic ML method that uses the supervised learning approach. The KNN algorithm assumes similarity between new and existing data and places the new data in the category that is most similar to the existing categories. The KNN algorithm maintains all available data and uses similarity to classify new data points [26]. This implies that as fresh data arrives, it may be quickly sorted into a well-suited category using the KNN algorithm.
- A decision tree (DT) algorithm is a machine learning method that predicts using a decision tree. It follows a tree-like model of decisions and their potential outcomes. The method divides the data recursively into subsets depending on the most significant attribute at each node of the tree [27]. A popular machine learning approach that may be used for both regression and classification applications is decision trees. The classification feature is used to examine the data in this study. The aim of machine learning is to reduce the amount of uncertainty or disorder in the dataset. The entropy metric is used to assess the uncertainty in a dataset or as a measure of disorder. The equation 3 represents the entropy formula. Entropy is a measure of a node's impurity. Impurity is a measure of randomness; it indicates how random our data is.

$$E(S) = -p_{(+)}logp_{(+)} - p_{(-)}logp_{(-)} \tag{3}$$

Here $p_{(+)}$ is the probability of positive class, $p_{(-)}$ is the probability of negative class, S is the subset of the training example.

- A supervised learning approach used to predict a dependent categorical target variable is logistic regression (LR). The logistic regression model forecasts a dependent variable by examining the connection between one or more existing independent variables. It enables machine learning systems to categorize incoming input based on past data. As more relevant data is added, the algorithms improve their ability to predict classifications within datasets. The LR is essential because it reduces complicated probability calculations to a simple arithmetic task.
- The naïve bayes (NB) method is a supervised learning technique that uses the Bayes theorem to solve classification issues. The Bayes theorem is used to evaluate the likelihood of a hypothesis given prior knowledge. It is determined by the conditional probability. The equation 4 represents the Bayes theorem formula.

$$P(A|B) = \frac{P(B|A)P(B)}{P(B)} \tag{4}$$

Where $P(A|B)$ is the posterior probability, $P(B|A)$ is the likelihood probability, $P(A)$ is the prior probability, and $P(B)$ is the marginal probability.

The NB converts the input dataset into frequency tables, which then build a likelihood table by calculating the probabilities of the given features. To compute the posterior probability, the NB employs the Bayes theorem.

- The random forest (RF) constructs decision trees by segmenting samples and seeks the best conclusion based on the voting forecast; it is difficult to overfit while dealing with regression difficulties [28]. One of the most essential characteristics of the RF algorithm is that it can handle data sets with both continuous variables, as in regression, and categorical variables, as in classification. This study employs the classification feature of the RF algorithm.
- The support vector machine (SVM) techniques rely on a collection of mathematical functions known as the kernel. The kernel's function is to receive data as input and transform it into the desired form [25]. The kernel functions compute the inner product of two points in a given feature space [29]. The sigmoid kernel function was utilized in the present study. The idea of SVM is to generate a hyperplane in D-dimensional feature space that approximates *h(x)*, which is the target (Fig. 3). The data is given in equation 5.

$$S = (X_i , Y_i)_{i=1}^{N} \tag{5}$$

The SVM considers approximating functions of form based on the knowledge of S in equation 6,

$$f(x, w) = \sum_{i=1}^{D} w_i \varphi_i(x) + b \tag{6}$$

where $\varphi_i(x)$ are the features and w_i are the coefficients which are estimated by minimizing in equation 7.

$$R(C) = \frac{1}{N} \sum_{i=1}^{N} |y_i - f(x_i, w)|_\varepsilon + \lambda \|w\|^2 \tag{7}$$

where λ is the constant term and $|y_i - f(x_i, w)|_\varepsilon$ is the robust error function. The equation of the main separator line is called a hyperplane equation.

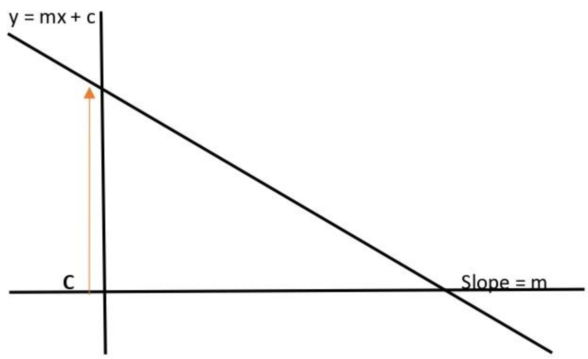

Fig. 3. The hyperplane equation separator.

The hyperplane equation dividing the points (for classifying) is easily summarized in equation 8.

$$H: w^T(x) + b = 0 \tag{8}$$

Here: b = Intercept and bias term of the hyperplane equation. In D dimensional space, the hyperplane would always be D -1 operator. For example, for 2-D space, a hyperplane is a straight line (1-D).

e) Performance evaluation measures

The evaluation of performance is a vital stage in constructing an accurate machine-learning model. The prediction model must be examined to ensure that it matches the dataset and works well on untested data. The goal of an evaluation of performance is to quantify a model's generalization accuracy on unseen/out-of-sample data [5]. Cross-Validation (CV) is a performance evaluation approach that divides data into divisions to evaluate and compare models [30]. The final dataset was divided into folds, which are equal-sized subsamples. In this investigation, tenfold cross-validation was utilized. The model was trained using nine folds, and the ML models were tested or validated using one-fold. Accuracy, precision, recall, F1-score, and Area Under the Curve-Receiver Operating Characteristics (AUC-ROC) are among the performance evaluation metrics evaluated. The performance evaluation measures are generated using the computed data's true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. The ability of the classification algorithm to properly anticipate the classes of the data set is referred to as accuracy [31]. It quantifies how close the projected value is to the real or theoretical value. In general, accuracy is defined as the ratio of correct predictions to total number of instances. The equation 9 shows the formula of the accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{9}$$

Precision is calculated by subtracting the true values that were successfully predicted from the total predicted values in the real class. Precision measures a classifier's ability to avoid labeling a negative example as positive. The formula of the precision is presented in equation 10.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{10}$$

The micro averaging method is used because it assigns equal weight to each sample. The equation of the micro average precision is shown in equation 11.

$$\text{PrecisionMicroAverage} = \frac{TP_1 + TP_2 + \cdots + TP_n}{TP_1 + TP_2 + \cdots + TP_n + FP_1 + FP_2 + \cdots + FP_n} \tag{11}$$

The rate of properly categorized positive values is measured by recall. Recall solves the issue of how many true positives are accurately categorized. Equation 12 presents the formula for the recall.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{12}$$

Because the micro average is used to calculate model recall, the micro average recall is calculated using equation 13.

$$\text{RecallMicroAverage} = \frac{TP_1 + TP_2 + \cdots + TP_n}{TP_1 + TP_2 + \cdots + TP_n + FN_1 + FN_2 + \cdots + FN_n} \tag{13}$$

The F1 score is a machine learning assessment statistic that measures the accuracy of a model. It combines the precision of a model and recall scores. The F1-score formula is shown in equation in 14.

$$F1_{score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{14}$$

The AUC - ROC curve is a performance metric for classification issues at various threshold levels. AUC is the degree or measure of separability, whereas ROC is a probability curve. The higher the AUC, the better the model predicts 0 classes as 0 and 1 classes as 1. The ROC curve is displayed using true positive rate (TPR) vs false positive rate (FPR), with TPR on the y-axis and FPR on the x-axis.

In machine learning, ensuring the reliability and precision of findings requires a mix of thorough data preparation, model selection, training, assessment, and continual monitoring [30]. The study strikes a careful balance between model complexity and interpretability, and the approaches used are tailored to the unique issue and situation. In this study, cross-validation is used to estimate how effectively the model generalizes to new, unknown data, as well as the relevant assessment criteria used. Accuracy may not be the optimal statistic in many circumstances, particularly for unbalanced datasets; nonetheless, metrics like as precision, recall, F1-score, ROC-AUC, are used to provide a broader view of model performance.

4. Experimental Results

This section describes the performance of the ML models used and the measures used to evaluate their performance.

4.1 Results from the machine learning models

The experiments are carried out using proposed ML models, and the experimental outcomes are calculated using various performance evaluation measures. The proposed ML models predict hemoglobin levels using k-Nearest Neighbor (KNN), decision tree (DT), logistic regression (LR), naïve bayes (NB), random forest (RF), and kernel-support vector machine (KSVM) classification models. The data was divided into test and training sets, and both were analyzed using the confusion matrix model. This was important since assessing the feasibility of the accuracy of classification methods. Table 1 emphasizes the manner of value distribution based on the confusion matrix.

Table 1. Adapted from [9, 32] about confusion matrix.

Real		
Classification prediction	True	False
True	TP	FN
False	FP	TN

According to Table 1, positive data correctly classified by the models is known as the "true positive" (TP), negative data correctly identified as negative is known as the "true negative" (TN), negative data incorrectly perceived as positive is known as "false negatives" (FN), and "positive" data incorrectly recognized as "positive" is known as "false positives" (FP). The Fig. 4 depicts the confusion matrix distribution of the six (6) classification machine learning models.

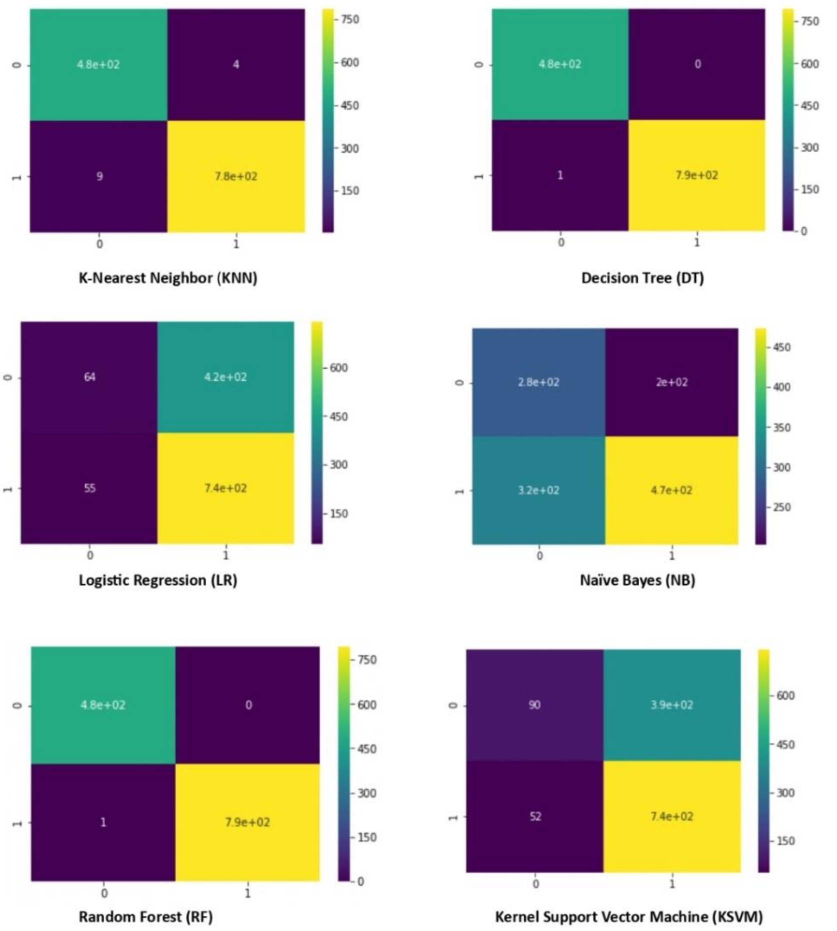

Fig. 4. The confusion matrix distribution of the ML models.

Table 2 displays the accuracy, precision, recall, and F1-score values from six (6) classification models. Table 2 shows that the decision tree and random forest had the same outcomes in terms of accuracy (0.9992), precision (1.000), recall (0.9987), and F1-score (0.9994). Because random forests are built on decision trees, the classification results of decision trees and random forests were similar. To promote generalization and avoid overfitting, random forests use the notion of averaging predictions from several trees. Random forests produce more robust and accurate classification results than single decision trees because they combine the predictions of individual decision trees. Because of the robustness of the former (random forest), the authors recommend using random forest rather than decision tree for detecting anemia.

Table 2. The performance evaluation measure scores of the ML models.

Model	Accuracy (%)	Precision	Recall	F1-score
K-Nearest Neighbor (KNN)	**0.9898**	**0.9949**	**0.9887**	**0.9918**
Decision Tree (DT)	**0.9992**	**1.000**	**0.9987**	**0.9994**
Logistic Regression (LR)	0.6283	0.6376	0.9307	0.7568
Naïve Bayes (NB)	0.5900	**0.6997**	0.5957	0.6435
Random Forest (RF)	**0.9992**	**1.000**	**0.9987**	**0.9994**
Kernel-Support Vector Machine (KSVM)	**0.6510**	0.6532	**0.9345**	**0.7689**

[2] **Green** is the highest performing measure, **blue** denotes the second-best performing measure, and **purple** denotes the third best performing measure.

In all performance evaluation measures, the KNN came in second place. Despite having the third best accuracy (0.6510), recall (0.9345), and F1-score (0.7689), the KSVM failed to achieve the best precision score. The naïve bayes classifier earned the third highest precision score, but the lowest accuracy (0.5900). The NB is the lowest performing ML classifier in terms of accuracy (0.5900), but it has a better precision (0.6997) score than the KSVM (0.6532) and the LR (0.6376). Based on the dataset, the various machine learning models performed well in detecting anemia. The findings demonstrate how important machine learning models have been utilized to forecast disease in this sector. The outputs of the ML models may be used to improve risk stratification, thus, prioritize patients (anemic patients) for further examination and treatment, hence maximizing healthcare resources and improving patient management. The results from the six (6) machine models are graphically represented in Fig. 5.

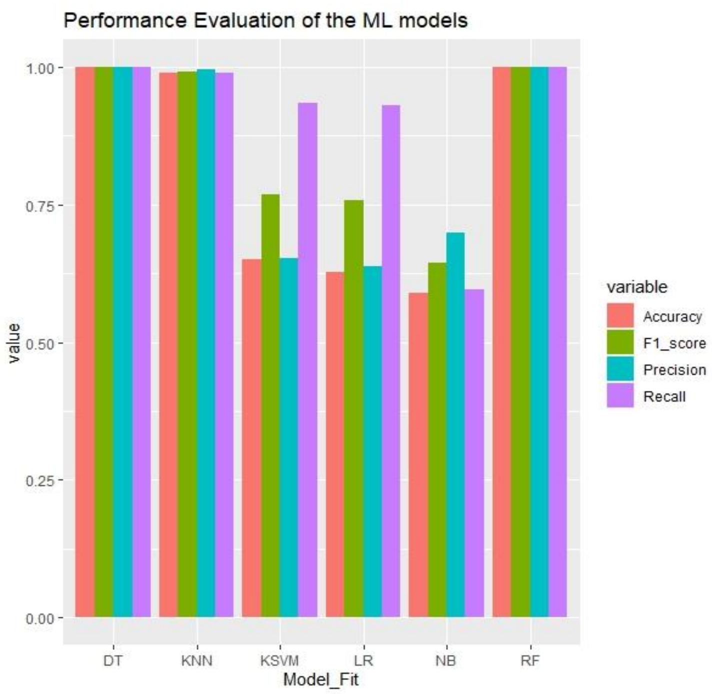

Fig. 5. Performance evaluation scores of the ML models.

4.2 AUC-ROC evaluation results of the ML models

The ML models' Area Under the Curve-Receiver Operating Characteristics (AUC-ROC) scores are used to determine how efficient the models are. The AUC-ROC score is a popular machine learning evaluation measure for binary classification problems. It assesses a classification model's ability to distinguish between positive and negative data across multiple classification criteria. The AUC, which is used as a summary of the ROC curve, is utilized for distinguishing between the binary (anemic and non-anemic) data used in this study. Table 3 summarizes the AUC-ROC scores of the ML classifiers employed. An AUC value greater than 0.5 indicates that the classifier is capable of identifying more True positives and True negatives than False negatives and False positives.

Table 3. The AUC-ROC scores of the ML Models.

Model	**AUC-ROC**
K-Nearest Neighbor (KNN)	**0.9902**
Decision Tree (DT)	**0.9994**
Logistic Regression (LR)	0.5315
Naïve Bayes (NB)	**0.5881**
Random Forest (RF)	**0.9994**
Kernel-Support Vector Machine (KSVM)	0.5602

[3] **Green** is the highest performing measure, **blue** denotes the second-best performing measure, and **purple** denotes the third best performing measure.

The random forest and decision tree models have the highest AUC-ROC (0.9994), followed by k-Nearest Neighbor (0.9902). The third AUC-ROC produced by the naive bayes was 0.5881. It is obvious that the models used in the present study obtained substantial AUC-ROC values over 0.5. Fig. 6 presents the graphical representation of the AUC-ROC of the models employed in this study.

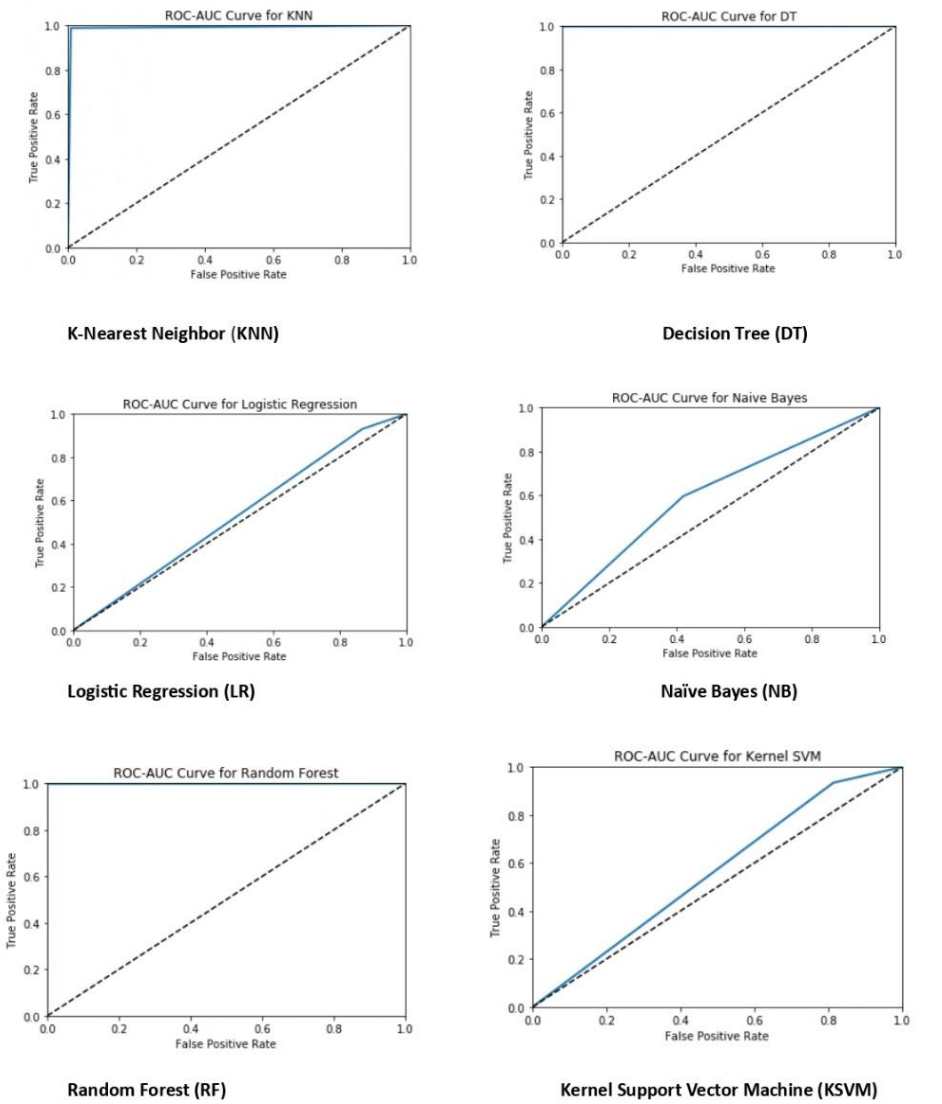

Fig. 6. AUC-ROC of the ML models.

The Fig. 7 below shows the graphical representation of the AUC-ROC values achieved by each ML model. When compared to other metrics such as accuracy, the AUC-ROC score is less affected by imbalanced class distributions. Accuracy can be deceptive in cases when one class dominates the dataset, but AUC-ROC gives a more trustworthy assessment. It is crucial to remember, however, that the AUC-ROC score has a few drawbacks. As a result, in this study, a variety of assessment measures was used to acquire a thorough knowledge of model performances.

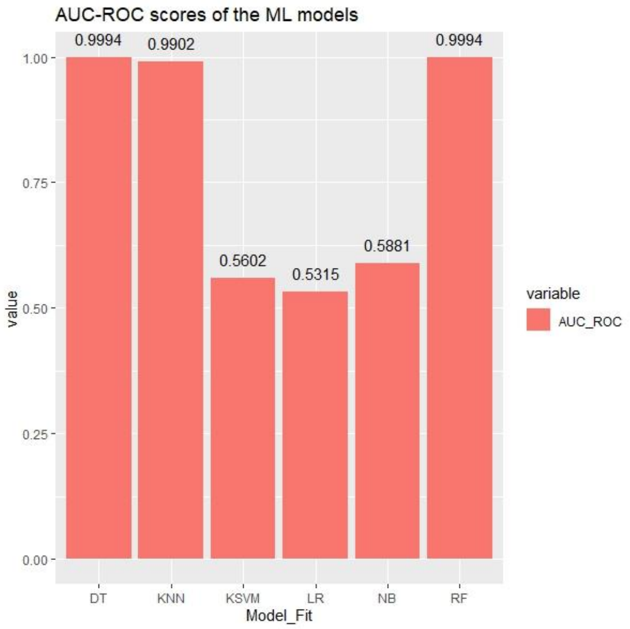

Fig. 7. Graphical representation of AUC-ROC values of the ML models.

4.3 Data visualization performance of the ML models

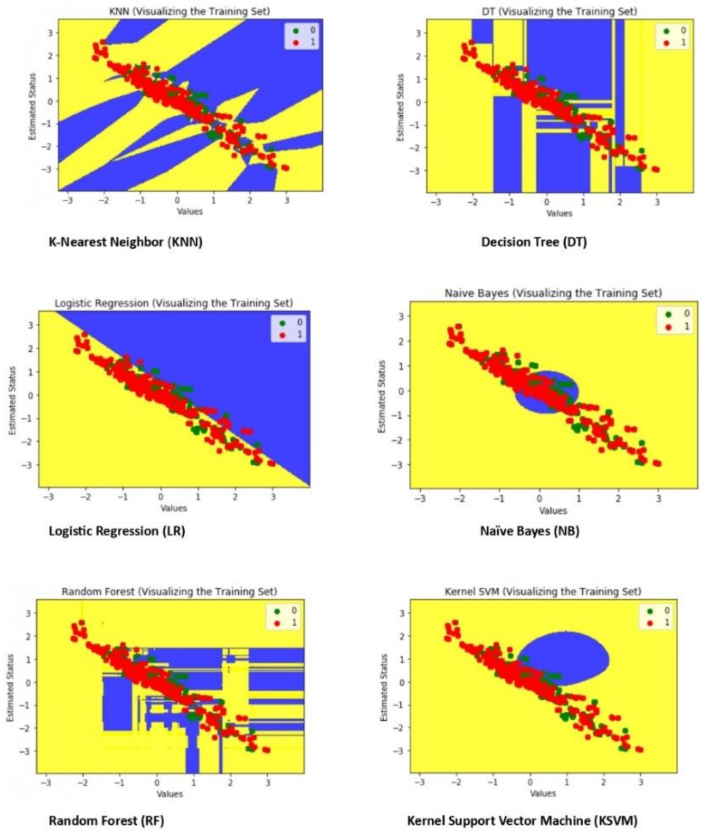

Fig. 8. Data visualization of the training sets of the ML models.

Data visualization aids in understanding how each machine learning model distributes data equitably. The variables in the dataset are non-anemic (0) and anemic (1). This contributes to the development of a relationship between the two variables, namely non-anemic (0) and anemic (1). Fig. 8 and Fig. 9 show the data visualization for the training and testing sets, respectively.

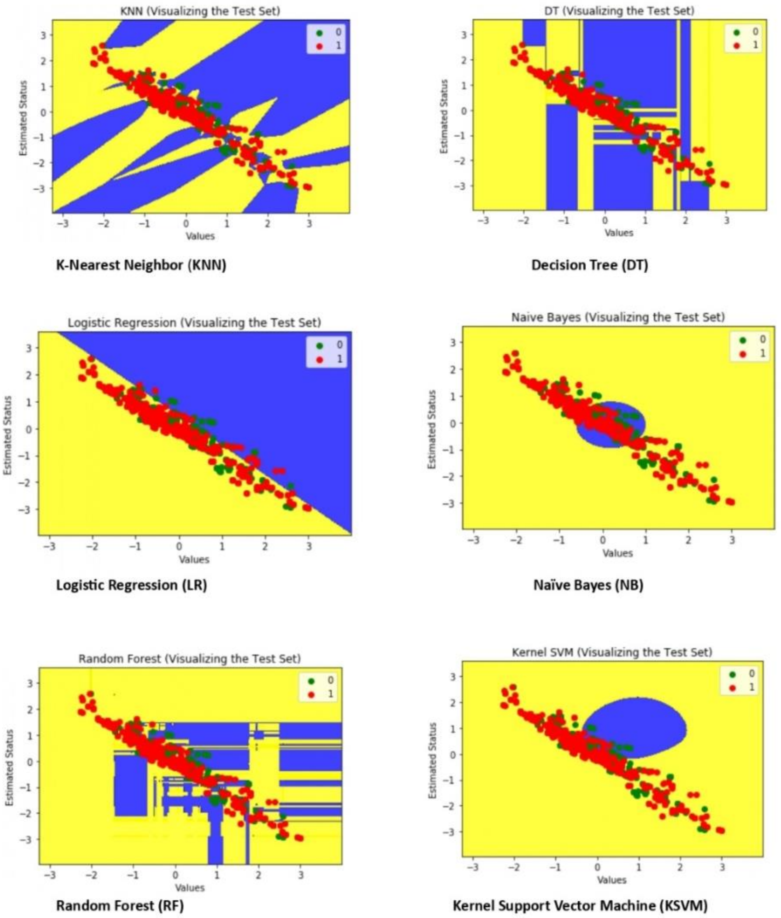

Fig. 9. Data visualization of the testing sets of the ML models.

According to the visualizations in Fig. 7 and Fig. 8, the data are uniformly distributed in both the training and testing sets. The machine learning models employed in this study provide a considerable distribution that gives insight into the non-anemic (0) and anemic (1) variables.

5. Discussion

Anemia is a severe global public health issue that disproportionately affects young children, menstrual teenage girls and women, pregnant and postpartum women, and pregnant and postpartum women. According to the WHO, 40% of infants aged 6 to 59 months, 37% of pregnant women, and 30% of women aged 15 to 49 are anemic globally. The invasive technique, which involves blood tests, endoscopy, colonoscopy, and bone marrow biopsy, is the most often used method of identifying anemia. There are several obstacles that impede the successful identification of anemia, including fear and customs from various religious and/or ethnic groups. The following machine learning models were used in this study: k-Nearest Neighbor (KNN), decision tree (DT), logistic regression (LR), naïve bayes (NB), random forest (RF), and kernel-support vector machine (KSVM). The study included 4260 data observations from Ghanaian hospitals of non-anemic (0) and anemic (1) people. The MATLAB Squeeze function was used to extract the red, green, and blue (RGB) values from the image dataset. The final dataset, hence the carefully labeled dataset, was trained with the study's machine learning models. The decision tree and random forest both obtained an accuracy of 0.9992, followed by the k-Nearest Neighbor with 0.9898. Based on performance, all of the machine learning models used in this study produced significant values with higher accuracies, with kernel-support vector machine obtaining an accuracy of 0.6510, logistic regression achieving an accuracy of 0.6283, and naïve bayes achieving an accuracy of 0.5900. Precision, recall, F1-score, and AUC-ROC scores were used to assess the machine learning models used in this study. The decision tree and random forest had the highest precision, recall, F1-score, and AUC-ROC scores. Data visualization was applied to the training and testing sets of the various machine learning models. The data visualization revealed that the variables, thus non-anemic (0) and anemic (1) data, were uniformly distributed, indicating a significant link between the variables. The study's findings indicate that machine learning algorithms are useful in detecting anemia in a non-invasive manner. It is advised that the non-invasive way of identifying anemia improves health quality, which contributes to Sustainable Development Goal 3 (SDG 3), and hence to excellent health and well-being. Table 4 compares the performance of the models used in this study to other studies that used machine learning models in the domain.

Table 4. The AUC-ROC scores of the ML Models.

Reference Authors	Research Year	Model Employed	Accuracy (%)
Zhang et al. [33]	2023	Random Forest	91.50
Appiahene et al. [14]	2023	k-Nearest Neighbor	97.92
Asare et al. [34]	2023	k-Nearest Neighbor	98.92
Alsayegh et al. [18]	2022	Logistic Regression	60.19
Proposed Model		**RF, DT, KNN, LR**	**99.92, 99.92, 98.98, 62.83**

When the proposed models are compared to current research in similar domains, they outperform the existing ones. The proposed models share commonalities with previously published studies in terms of the operationalization of the algorithms, however they differ in terms of the core dataset utilized to train the models. It should be noted that the use of machine learning algorithms to detect anemia in a non-invasive manner has improved; nonetheless, this study gives the latest developments in the field. Based on the accuracy and performance evaluation measures, the machine learning employed in the study demonstrates resilience above previously known models.

6. Conclusion and Future Works

a) Conclusion

Early detection is essential for both professionals and patients in preventing and progressing anemia, particularly in children and pregnant women. The invasive way to detecting anemia has several obstacles that impede community health and well-being. With the growth of technology and its massive influence, it is essential to create a non-invasive method of detecting anemia. The non-invasive method of detecting anemia can save money and time. The proposed algorithms are built using six (6) different machine learning models: k-Nearest Neighbor, decision tree, logistic regression, naïve bayes, random forest, and kernel-support vector machine. With tenfold cross-validation, the dataset was divided into training (70%) and testing (30%) halves. The performance of the built models was assessed using accuracy, precision, recall, F1-score, and AUC-ROC scores after tenfold cross-validation. The final dataset was used to train the machine learning models, and the results were substantial. The decision tree and random forest had the greatest accuracy of 99.92% each, followed by k-Nearest Neighbor at 98.98%. With 65.10% accuracy, the kernel-support vector machine placed third. The naïve bayes model has the lowest accuracy of 59.00%. The decision tree and random forest models performed best in terms of precision, recall, and F1-score. The use of the palpable palm in the identification of iron deficiency anemia with substantial findings, that is, the greater performance of the models, is the most novel characteristic of this study, as earlier research employed the conjunctiva of the eyes [2]. According to Appiahene et al. [14], when the palpable palm, color of the fingernails, and conjunctiva of the eyes were utilized to diagnose anemia in children, the palm had a better accuracy than the conjunctiva and the fingernails. Furthermore, the current study outperforms their performance as well as other state-of-the-art in the field of study. This demonstrates that the palpable palm is an important area for anemia detection. Although the proposed framework demonstrates improved sensitivity with an acceptable range of errors and classification accuracy, it attempts to evaluate the repeatability of samples acquired from primary sources over time. We feel that machine learning is critical to contributing to the identification of anemia in a non-invasive method based on the substantial outcomes of the various models.

b) Future works

It is envisaged that the study would utilize diverse datasets to train the models and compare the outcomes. Furthermore, ensembles of machine learning models will be employed to train the dataset. To train the dataset, deep learning methods such as convolutional neural networks and capsule networks will be employed.

References

[1] G. Dimauro, M. E. Griseta, M. G. Camporeale, F. Clemente, A. Guarini, and R. Maglietta, “An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset,” *Artif. Intell. Med.*, vol. 136, no. August 2022, p. 102477, 2023, doi: 10.1016/j.artmed.2022.102477.

[2] P. Appiahene, E. J. Arthur, S. Korankye, S. Afrifa, J. W. Asare, and E. T. Donkoh, “Detection of anemia using conjunctiva images: A smartphone application approach,” *Med. Nov. Technol. Devices*, vol. 18, no. February, p. 100237, 2023, doi: 10.1016/j.medntd.2023.100237.

[3] Q. Chen, Y. Zhang, M. Zhang, Z. Li, and J. Liu, "Application of Machine Learning Algorithms to Predict Acute Kidney Injury in Elderly Orthopedic Postoperative Patients," *Clin. Interv. Aging*, vol. 17, no. March, pp. 317–330, 2022, doi: 10.2147/CIA.S349978.
[4] A. Chakraborty, S. Mitra, M. Bhattacharjee, D. De, and A. J. Pal, "Determining human-coronavirus protein-protein interaction using machine intelligence," *Med. Nov. Technol. Devices*, vol. 18, no. December 2022, p. 100228, 2023, doi: 10.1016/j.medntd.2023.100228.
[5] WHO.int, "Anaemia." Accessed: Apr. 30, 2023. [Online]. Available: https://www.who.int/health-topics/anaemia#tab=tab_1
[6] E. I. Abd El-Latif and N. E. Khalifa, "A Model based on Deep Learning for COVID-19 X-rays Classification," *Int. J. Image, Graph. Signal Process.*, vol. 15, no. 1, pp. 36–46, 2023, doi: 10.5815/ijigsp.2023.01.04.
[7] S. E. Wilson *et al.*, "Comprehensive framework for integrated action on the prevention , diagnosis , and management of anemia : An introduction The WHO Interdepartmental Anaemia Working Group : María Barreix 2," pp. 1–5, 2023, doi: 10.1111/nyas.14999.
[8] A. Kesarwani *et al.*, "Non-invasive anaemia detection by examining palm pallor: A smartphone-based approach," *Biomed. Signal Process. Control*, vol. 79, no. July 2022, 2022, doi: 10.1016/j.bspc.2022.104045.
[9] S. Afrifa, V. Varadarajan, P. Appiahene, T. Zhang, and E. A. Domfeh, "Ensemble Machine Learning Techniques for Accurate and Efficient Detection of Botnet Attacks in Connected Computers," *MDPI Eng*, pp. 650–664, 2023, doi: https://doi.org/10.3390/eng4010039.
[10] S. Afrifa and V. Varadarajan, "Cyberbullying Detection on Twitter Using Natural Language Processing and Machine Learning Techniques," *Int. J. Innov. Technol. Interdiscip. Sci.*, vol. 5, no. 4, pp. 1069–1080, 2022, doi: 10.15157/IJITIS.2022.5.4.1069-1080.
[11] A. Mohammed, P. S. B. Podila, R. L. Davis, K. I. Ataga, J. S. Hankins, and R. Kamaleswaran, "Using machine learning to predict early onset acute organ failure in critically ill intensive care unit patients with sickle cell disease: Retrospective study," *J. Med. Internet Res.*, vol. 22, no. 5, pp. 1–10, 2020, doi: 10.2196/14693.
[12] W. K. Adu, P. Appiahene, and S. Afrifa, "VAR , ARIMAX and ARIMA models for nowcasting unemployment rate in Ghana using Google trends," *J. Electr. Syst. Inf. Technol.*, pp. 1–16, 2023, doi: 10.1186/s43067-023-00078-1.
[13] S. Afrifa, T. Zhang, P. Appiahene, and V. Vijayakumar, "Mathematical and Machine Learning Models for Groundwater Level Changes : A Systematic Review and Bibliographic Analysis," *Futur. Internet*, 2022, doi: https:// doi.org/10.3390/fi14090259.
[14] P. Appiahene, J. W. Asare, E. T. Donkoh, G. Dimauro, and R. Maglietta, "Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms," *BioData Min.*, vol. 16, no. 1, pp. 1–20, 2023, doi: 10.1186/s13040-023-00319-z.
[15] A. Mitani *et al.*, "Detection of anaemia from retinal fundus images via deep learning," *Nat. Biomed. Eng.*, vol. 4, no. 1, pp. 18–27, 2020, doi: 10.1038/s41551-019-0487-z.
[16] S. KILICARSLAN, M. CELIK, and Ş. SAHIN, "Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification," *Biomed. Signal Process. Control*, vol. 63, no. September 2020, 2021, doi: 10.1016/j.bspc.2020.102231.
[17] L. Alzubaidi, M. A. Fadhel, O. Al ‐ shamma, J. Zhang, and Y. Duan, "Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis," *Electron.*, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030427.
[18] F. Alsayegh, M. A. Alkhamis, F. Ali, S. Attur, N. M. Fountain-Jones, and M. Zubaid, "Anemia or other comorbidities? using machine learning to reveal deeper insights into the drivers of acute coronary syndromes in hospital admitted patients," *PLoS One*, vol. 17, no. 1 January, pp. 1–15, 2022, doi: 10.1371/journal.pone.0262997.
[19] E. Waisberg, J. Ong, N. Zaman, S. A. Kamran, A. G. Lee, and A. Tavakkoli, "A non-invasive approach to monitor anemia during long-duration spaceflight with retinal fundus images and deep learning," *Life Sci. Sp. Res.*, vol. 33, no. April, pp. 69–71, 2022, doi: 10.1016/j.lssr.2022.04.004.
[20] P. Appiahene, V. Varadarajan, T. Zhang, and S. Afrifa, "Experiences of sexual minorities on social media : A study of sentiment analysis and machine learning approaches," *J. Auton. Intell.*, vol. 6, no. 2, pp. 1–15, 2023, doi: 10.32629/jai.v6i2.623.
[21] P. Appiahene, Y. M. Missah, and U. Najim, "Evaluation of information technology impact on bank's performance: The Ghanaian experience," *Int. J. Eng. Bus. Manag.*, vol. 11, no. March, 2019, doi: 10.1177/1847979019835337.
[22] B. Jalili, A. Rezaeian, P. Jalili, D. D. Ganji, and Y. Khan, "Squeezing flow of Casson fluid between two circular plates under the impact of solar radiation," *ZAMM Zeitschrift fur Angew. Math. und Mech.*, no. October 2022, pp. 1–11, 2023, doi: 10.1002/zamm.202200455.
[23] M. A. Junior, P. Appiahene, and O. Appiah, "Forex market forecasting with two ‐ layer stacked Long Short ‐ Term Memory neural network (LSTM) and correlation analysis," *J. Electr. Syst. Inf. Technol.*, vol. 1, pp. 1–24, 2022, doi: 10.1186/s43067-022-00054-1.
[24] S. Roy, "Island Loss for Improving the Classification of Facial Attributes with Transfer Learning on Deep Convolutional Neural Network," *Int. J. Image, Graph. Signal Process.*, vol. 12, no. 1, pp. 18–29, 2020, doi: 10.5815/ijigsp.2020.01.03.
[25] S. Afrifa, T. Zhang, X. Zhao, P. Appiahene, and M. S. Yaw, "Climate change impact assessment on groundwater level changes: A study of hybrid model techniques," *IET Signal Process.*, vol. 17, no. April, p. e12227, 2023, doi: 10.1049/sil2.12227.
[26] S. Shastri, "An Efficient Approach for Text-to-Speech Conversion Using Machine Learning and Image Processing Technique," *Int. J. Educ. Manag. Eng.*, vol. 13, no. 4, pp. 44–49, 2023, doi: 10.5815/ijem.2023.04.05.
[27] R. Hasan and A. S. M. Shafi, "Feature Selection based Breast Cancer Prediction," *Int. J. Image, Graph. Signal Process.*, vol. 15, no. 2, pp. 13–23, 2023, doi: 10.5815/ijigsp.2023.02.02.
[28] G. Rossin *et al.*, "Artificial Intelligence in Bladder Cancer Diagnosis: Current Applications and Future Perspectives," *BioMedInformatics*, vol. 3, no. 1, pp. 104–114, 2023, doi: 10.3390/biomedinformatics3010008.
[29] J. Lötsch and B. Mayer, "A Biomedical Case Study Showing That Tuning Random Forests Can Fundamentally Change the Interpretation of Supervised Data Structure Exploration Aimed at Knowledge Discovery," *BioMedInformatics*, vol. 2, no. 4, pp. 544–552, 2022, doi: 10.3390/biomedinformatics2040034.
[30] J. Lötsch and A. Ultsch, "Enhancing Explainable Machine Learning by Reconsidering Initially Unselected Items in Feature Selection for Classification," *BioMedInformatics*, vol. 2, no. 4, pp. 701–714, 2022, doi: 10.3390/biomedinformatics2040047.

[31] S. C. Christopoulou, "Machine Learning Tools and Platforms in Clinical Trial Outputs to Support Evidence-Based Health Informatics: A Rapid Review of the Literature," *BioMedInformatics*, vol. 2, no. 3, pp. 511–527, 2022, doi: 10.3390/biomedinformatics2030032.
[32] P. Appiahene, Y. M. Missah, and U. Najim, "Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks," *Adv. Fuzzy Syst.*, vol. 2020, 2020, doi: 10.1155/2020/8581202.
[33] F. Zhang, J. Yang, Y. Wang, M. Cai, J. Ouyang, and J. X. Li, "TT@MHA: A machine learning-based webpage tool for discriminating thalassemia trait from microcytic hypochromic anemia patients," *Clin. Chim. Acta*, vol. 545, no. February, p. 117368, 2023, doi: 10.1016/j.cca.2023.117368.
[34] J. W. Asare, P. Appiahene, E. T. Donkoh, and G. Dimauro, "Iron Deficiency Anemia Detection using Machine Learning Models : A Comparative Study of Fingernails , Palm and Conjunctiva of the Eye Images .," *Eng. Reports*, no. February, pp. 1–28, 2023, doi: 10.1002/eng2.12667.

Authors' Profiles

Stephen Afrifa received a BSc. degree in Information Technology from University of Energy and Natural Resources, Ghana in 2020. He is a currently a postgraduate student leading to MSc. Information and Communication Engineering from Tianjin University, China. He currently provides community services to school children through imparting the technological experience for IT entrepreneurial skills for sustainable development. His current research interest includes image and signal processing, deep learning, machine learning, IoT, cloud computing, and network security. His current thesis is focused on using mathematical and machine learning models to estimate climate change impact on groundwater level changes for sustainable environment.

Dr. Peter Appiahene is an Assistant Professor (Senior Lecturer) in the Department of Information Technology and Decision Sciences at the University of Energy and Natural Resources, Ghana. Dr. Peter Appiahene has a PhD in Computer Science and is the current Head of the Department. Designs and Analysis of Algorithm, Deep Learning, Image Processing, Social Network Analysis, and Optimizations are his current research areas.

Professor Tao Zhang received the B.Sc., M.Sc., and Ph.D. degrees from Tianjin University, Tianjin, China, in 1998, 2001 and 2004 respectively. He is an Associate Professor with the School of Electrical and Information Engineering, Tianjin University, China. His current research interests include image, video, and acoustic signal processing, auditory model and machine learning.

Professor Dr. Vijayakumar Varadarajan has completed Diploma with First Class Honors. He has completed BE CSE and MBA HRD with First Class. He has also completed ME CSE with First Rank Award. He has completed his PhD from Anna University in 2012. He is an Adjunct Professor in School of Computer Science and Engineering, University of New South Wales, Sydney, Australia. His current research interest includes image and signal processing, deep learning, machine learning, natural language processing, cloud computing, and IoT.

How to cite this paper: Stephen Afrifa, Peter Appiahene, Tao Zhang, Vijayakumar Varadarajan, "Machine Learning Algorithms for Iron Deficiency Anemia Detection in Children Using Palm Images", International Journal of Education and Management Engineering (IJEME), Vol.14, No.1, pp. 1-15, 2024. DOI:10.5815/ijeme.2024.01.01