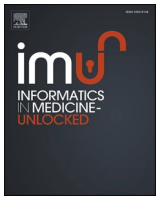
Informatics in Medicine Unlocked 45 (2024) 101451

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu



Application of machine learning approach for iron deficiency anaemia detection in children using conjunctiva images

Justice Williams Asare [a,c,*], William Leslie Brown-Acquaye [a], Martin Mabeifam Ujakpa [a], Emmanuel Freeman [a], Peter Appiahene [b]

[a] Ghana Communication Technology University (GCTU), Accra, Ghana
[b] University of Energy and Natural Resources (UENR), Sunyani, Ghana
[c] The West African Examinations Council (WAEC), Accra, Ghana

ARTICLE INFO

Keywords:
Iron deficiency
CIELAB colour space
Non-invasive
Invasive
Algorithms
Anaemia

ABSTRACT

Iron deficiency is commonly referred to as anaemia which is a general public health problem that normally occurs as a result of a reduction in red blood cells which is common in developing countries such as Africa. In this study, machine learning algorithms such as CNN, k-NN, Naïve Bayes, Decision Tree and SVM were utilized for the study to detect anaemia in children using conjunctiva images. The images were segmented into their various CIELAB colour space components and the ROI from each image was retrieved. The dataset was split randomly into 70:10:20, which were then used to train, validate, and test the models, as appropriate. The CNN achieved the highest accuracy (98.45 %). The findings of this study demonstrate that non-invasive techniques are essential for detecting anaemia in children. This study deploys a cost-effective mechanism, and result-orientated, to detect anaemia in developing communities where health facilities, resources, and personnel are scarce.

1. Introduction

One of the most common public health problems that confront many countries around the world is iron deficiency which is commonly known or referred to as anaemia, and mostly affects young children below the age of 6 [1,2]. Studies have shown that everyone, that is people of all ages is at risk of iron deficiency, however, children below the age of 59 months and pregnant women are at higher risk [3,4], and is commonly recorded in developing countries such as Africa [5]. Symptoms of iron deficiency include fatigue, weakness, pale skin, shortness of breath, and dizziness. Early detection of iron deficiency is essential for proper treatment and prevention of further complications [6] since it has significant economic implications and hinders the growth of the nation by reducing the labour capacity of individuals and entire populations [7]. Vital organs including the heart, brain, liver, and kidneys receive more blood when an individual suffers from iron deficiency, whereas less important organs receive less blood [8].

Iron deficiency occurs when the haemoglobin value in the blood vessels is diminished below the normal threshold of blood in the human body, which is commonly referred to as red blood cell deficiency. The process of detecting and diagnosing iron deficiency (anaemia) is performed by examining the conjunctiva of the eyes which is performed by medical officers or physicians. This method or technique is often referred to as clinical signs. However, studies show that this procedure is not reliable and efficient because from time to time it usually depends on the discretion of the medical officer considering the extent of paleness and colour of the conjunctiva of the eyes [9,10].

The detection or diagnosis of iron deficiency can be confirmed through the assessment of the quantity of haemoglobin in the blood or the hematocrit, which measures the proportion of red blood cells to total blood volume. Iron deficiency (anaemia) is thought to be present in a patient whose haemoglobin or hematocrit readings are more than two standard deviations below the normal range [11]. A patient with a low RBC mass who is simultaneously experiencing hypovolemia-caused dehydration-induced plasma volume loss may not have blood that accurately reflects the severity of the iron deficiency since the haemoglobin and hematocrit levels are likely to be within the normal range [8].

Conversely, the detection of iron deficiency is mostly done by using the invasive approach, in which blood samples are taken to the laboratory for an iron deficiency test to confirm whether a patient is iron deficient or not. This invasive approach is expensive, time-consuming,

* Corresponding author. The West African Examinations Council (WAEC), Accra, Ghana.
E-mail address: justwills15@gmail.com (J.W. Asare).

https://doi.org/10.1016/j.imu.2024.101451
Received 12 January 2024; Received in revised form 19 January 2024; Accepted 20 January 2024
Available online 28 January 2024

and not applicable in rural communities where there are no proper health facilities, resources, and personnel, such as physicians and biomedical scientists to perform such tests.

In addition, it sometimes endangers healthcare professionals such as biomedical scientists and laboratory technicians who take blood samples from potential people with iron deficiency for testing, which could lead to blood infection diseases [12]. Current approaches in machine learning have provided outstanding solutions to concerns in many different fields, including Computer Vision (CV), medical imaging (medical image processing and analysis), and natural language processing (NLP) [13]. Impressive performances have been demonstrated in convolutional neural networks in terms of medical imaging and segmentation [14], but there has been little research on segmenting conjunctiva [15]. Despite this, CNNs have shown good results in the image processing domain due to their inherent abilities to process the spatial context.

As a method of overcoming this challenge, scientists have proposed the application of a non-invasive technique, which includes a machine learning approach, which is effective and less costive and reliable in detecting iron deficiency [16]. In precis, recent advances in machine learning techniques and approaches have enabled researchers to develop automated systems for the detection of iron deficiency (anaemia) from the conjunctiva of the eye images. These systems have achieved good accuracy and are useful for the early detection of iron deficiency mainly in areas where medical facilities and personnel are scarce.

In light of this, the current work proposed the use of machine learning models for the detection of iron deficiency using a primary medical dataset (pictures of the conjunctiva of the eyes) collected from chosen Ghanaian hospitals. That is, to compare the performance of the various machine learning algorithms such as support vector machine, decision tree, Naïve Bayes, and convolutional neural to facilitate the detection of iron deficiency in the health centres or facilities, especially where medical officers or physicians are scarce in Ghana. This would require quick detection of iron deficiency and timely treatment due to its efficiency of performance, timely outcome orientated, and cost-effectiveness.

2. Related works

Numerous studies have been conducted for iron deficiency (anaemia) detection utilizing a non-invasive approach such as machine learning models with the use and application of conjunctiva of the eye images, which have proven to be successful. Recent advances in medical imaging technology have allowed researchers to develop methods and models for iron deficiency detection with the application of medical imaging.

In 2022, Zhang et al. [17] developed an automated system that uses computer vision and machine learning techniques for the detection of iron deficiency (anaemia) using conjunctiva of eye images. The system was trained on a dataset of 362 conjunctiva images and achieved an accuracy of 82.37 %. Chen et al. [18] developed a machine learning model to detect iron deficiency by predicting Hb levels using conjunctiva images. The system was trained on the conjunctiva of the eye images and achieved a confidence interval (C.I) of 95. Similarly, Sarsam et al. [19] in their study proposed a novel mechanism for diagnosing patients with iron deficiency based on the relationship between the symptoms of iron deficiency and patients' emotions posted on the Twitter platform. The k-means and Latent Dirichlet Allocation (LDA) algorithms were used for the study which achieved a prediction accuracy of 98.96 %. Shahzad et al. [20] proposed a three-tier deep convolutional fused network (3-TierDCFNet) to extract optimum morphological features with the use of iron deficiency images to predict the rate of the severity of iron deficiency. The proposed model covered two modules: Module-I classified the input image into two classes, that is, iron deficient and non-iron deficient, while Module-II detected the rate of the severity level of the iron deficiency and categorized it into mild or chronic.

Using 99 datasets (images), a study by Jain et al. [21] employed a non-invasive technique for the classification of iron deficiency by using images of the conjunctiva of the eyes. Due to the study's success in achieving an accuracy of 97 % with the Artificial Neural Network (ANN) method, the dataset size was enhanced by applying image augmentation techniques.

Furthermore, the authors in Ref. [12] conducted a study to detect anaemia using the thresholding algorithm, which was an automated anaemia detection system. In their study, the colour of the conjunctiva was quantified when captured by a camera, and then the RGB colour components were obtained which achieved an accuracy of 78.90 %. Another study by Magdalena et al. [22] used CNN for anaemia detection utilizing conjunctiva of the eye images and produced an accuracy of 94 % while precision, recall and F1-score were used to evaluate the performance obtained in the study. In addition, a Convolutional Neural Network (CNN) was utilized by Delgado-Rivera et al. [23] by utilizing conjunctiva images after successfully segmenting the images and achieved a sensitivity of 77.58 % when compared to the laboratory test results.

With regards to Noor et al. [24], the authors did a comparative study to assess the performance of support SVM, NB, k-NN and DT to detect anaemia using 104 images of the conjunctiva of the eyes. Using a confusion matrix to evaluate the models' performance, the Decision Tree outperformed the other models by attaining an accuracy of 82.61 %. In the study conducted by Agrawal [25], the author used 400 retinal images to detect anaemia using CNN, which resulted in an accuracy of 62.23 %. Furthermore, the study conducted by Peksi et al. [26] used the Naïve Bayes algorithm with 20 conjunctivas of the eye images and had an accuracy of 90 %.

The use of a developed mobile phone application by Dimauro et al. [2] utilized the k-NN classification algorithm and used images of the conjunctiva of the eye to detect anaemia which gave significant results when tested on healthy (non-anaemic) people when compared with their real Hb values and with the results of the developed mobile application. Additionally, images of the retinal fundus with their correlated Hb values were used to detect anaemia using a deep learning model that gave significant results with a confidence interval of 90.5 %.

The study by Agrawal [25] used a CNN approach to detect anaemia with the use of 400 medical images. The rotate, zooming, translate and flip techniques were applied to augment the original datasets. However, the overall quantity (size) of the dataset attained after augmenting the images was not indicated. The CNN achieved an accuracy of 74 %, AUC attained 62.23 % and a precision of 59.00 %. Moreover, a SVM approach was used to detect anaemia which achieved an accuracy of 93 % when Bauskar et al. [27] conducted a study on anaemia detection.

A study conducted by Jain et al. [21] utilized ANN to detect anaemia using 99 image datasets that were enhanced with the application of mirroring, rotation, and translation techniques. The proposed ANN model achieved an accuracy of 97 % when evaluated with specificity, sensitivity and confusion metrics. The Naïve Bayes algorithm was applied by Peksi et al. [26] regarding anaemia detection with the use of 20 images (dataset) without image augmentation. The proposed model achieved an accuracy of 90 %, with the use of specificity and sensitivity to evaluate the model's performance.

The study by Ref. [28] applied several machine learning algorithms which include CNN, SVM, DT, and NB for anaemia detection with the use of palpable palm images. The images were extracted, and augmented, for training, validation and testing on their proposed models. The CNN achieved an accuracy of 99.96 %, while SVM had the lowest accuracy of 96.34 % when the models were evaluated. In the study by Dhalla et al. [15], five pre-trained segmentation architectures utilizing 162 images of the conjunctiva eyes of pediatric patients were utilized for the detection of anaemia and to analyze the effectiveness of their proposed models (UNet, UNet++, FPN, PSPNet and LinkNet) and segmented the palpable images of the conjunctiva from the eye images.

The study suggests that the LinkNet model was better suited for the segmentation of the palpable conjunctiva images which achieved an accuracy of 94.17 %.

Ghosh et al. [29] used ANN to estimate the Hb levels using colour information taken from blood sample images collected from 86 participants. The colour intensity values calculated from the blood drop images were used as feature descriptors for the samples. The features obtained from their samples were used to feed the ANN and a sensitivity of 95.50 % with a specificity of 52.00 % A study by Dimauro et al. [30] denoted that screening for anaemia using a non-invasive and cost-effective mechanism is currently a significant research challenge because it spares patients encountering discomfort from physical examinations and clinical laboratory tests. The study used "the Eyes-defy-anemia dataset" comprises 218 conjunctivae of eye images from Italy and India. The system was trained using palpebral conjunctiva images, which achieved an accuracy of 88 %, 0.66 sensitivity, and 0.91 specificity on the Italian dataset and 75 % accuracy, 0.79 sensitivity, and 0.74 specificity on the Indian dataset.

Acar et al. [31] proposed a system for diagnosing cataract disease based on images of the colour of the fundus. For the accurate diagnosis of cataract disorders, deep learning-based models were used for the study, and to detect anomalies in the descriptive regions of the human eye, two trained robust architectures, the VGGNet and DenseNet architectures were utilized. The proposed system achieved 97.94 % accuracy by the VGGNet and 95.07 % accuracy by DenseNet in the diagnosis of cataract disease. In a study by Dimauro et al. [32] a novel pipeline was proposed for the estimation of anaemia comprising three key contributions: a sclera segmentation algorithm applied to near-taken digital images of the eye, an algorithm for a vessel extraction, and a classifier to predict the anaemic and non-anaemic status of a person. The proposed sclera segmentation algorithm had an F1 score of 87 % when compared the performance to a state-of-the-art algorithm.

From the above studies, it is induced that the application of non-invasive approaches like machine learning algorithms that are effective, less costive, require less time, and achieve great results for anaemia detection. In addition, the conjunctiva of the eyes is one of the vital sites in the detection of anaemia using machine learning algorithms [33].

The conjunctiva of the eye, in particular, is a useful and efficient part of the human body for anaemia detection, due to this, we employed five (5) machine learning algorithms, namely, convolutional neural network, support vector machine, decision tree, k-nearest neighbor, and the naïve Bayes for anaemia detection using images of the conjunctiva [21].

In addition, the main contribution of this study is the use of a large number of datasets (527 conjunctiva images) and diverse machine learning models for the detection of anaemia since most of the studies used a smaller number of datasets and a single machine learning model, as one of the critical limitations of applying machine learning algorithms in anaemia detection is the need for a large and diverse dataset [21,28]. Moreover, the motivation of this study is to devise a mechanism and approach that is cost-effective, result-orientated, and a stress-free procedure to detect anaemia in developing communities where health facilities, resources, and personnel are scarce. This would help to diagnose and detect anaemia and give early treatment to anaemic patients on time.

2.1. Limitations on the proposed models

The application of machine learning algorithms in anaemia detection has been of interest to researchers over the past years. Several studies have been conducted to evaluate the performance of these algorithms in detecting anaemia, with significantly varying results [28]. The need for a sizable and varied dataset is one of the fundamental drawbacks of machine learning methods in anaemia diagnosis [21]. Anaemia is a complex disease, and its symptoms and causes can vary greatly between individuals. As such, machine learning algorithms require sufficient data points to accurately detect anaemia, which may not always be available. However, individually, all machine learning algorithms have limitations or drawbacks in their operations.

Support Vector Machines (SVMs) are prevailing supervised machine learning algorithms which have been successfully applied to a wide range of classification and regression tasks as with any machine learning algorithm, however, certain limitations must be considered when using SVMs [34]. One of the primary limitations of SVM is its reliance on a fixed kernel function. While there are many different kernel functions available, each kernel has its own set of parameters, which must be tuned for optimal results.

This requires a great deal of trial and error and can be time-consuming. In addition, SVMs can be slow to train when dealing with large datasets. This is because they require a high degree of computational complexity, which can lead to long training times and are not well suited for handling large amounts of noise in the data. SVMs are not designed to handle data with non-linear relationships, although certain techniques can be used to handle non-linear data, they can be slow and computationally expensive [35].

Naive Bayes is a popular and robust machine learning algorithm that has been widely used for a variety of tasks, however, there are some limitations to its use. The most significant limitation of Naive Bayes is its reliance on the assumption of independence between features. This means that it assumes that each feature is independent of the other features, which may not always be the case. Naive Bayes is vulnerable to the curse of dimensionality, which means that it is unable to accurately predict data points with a large number of features [36].

Decision trees are dominant and vigorous for decision-making and classification in both supervised and unsupervised learning, although decision trees have several limitations which have been identified in recent research [36]. These include their inherent instability, their difficulty of interpretability, and their susceptibility to overfitting. Decision tree instability occurs when small changes to training data can lead to large changes in the tree structure. This can lead to poor generalization performance. Furthermore, decision trees can be difficult to interpret, as the structure of the tree can be complex and difficult to understand [34,35].

Several studies have highlighted the limitations of Convolutional Neural Networks (CNNs) for various challenging tasks. Most notably, these limitations include difficulty in learning long-term dependencies, lack of interpretability, and difficulty in generalizing to unseen data. CNNs are limited in their ability to learn long-term dependencies in sequential data [37]. This is because CNNs are designed to capture local patterns in their receptive fields and do not capture dependencies that extend beyond the local region [34,37]. In addition, CNNs are often considered to be a black box due to their lack of interpretability. This makes them difficult to debug, as it is hard to understand why a particular decision is being made.

k-NN is a robust technique for classification and regression problems, but it has certain limitations. It is sensitive to the choice of the value of k. That is, if the value of k is too small, the model may overfit to some of the data points and thus may lead to inaccurate predictions, and, if the value of k is too large, it may lead to inaccurate predictions due to the "curse of dimensionality" [34,36]. The k-NN algorithm also requires that the entire training dataset be stored in memory, which makes it computationally expensive and sensitive to noisy data and outliers.

One of the main gaps identified in the related works is the size of the dataset used. A huge dataset is required for machine learning training and testing because it allows the algorithm to learn patterns and trends in the data that can be used to make predictions. The more data you have to train and test on, the more accurately the algorithm can learn and make predictions. Additionally, having a large dataset can help to reduce the risk of overfitting, which is when the algorithm learns the details of the training data too well and fails to generalize to new data. The proposed study used a larger size of dataset of 527 conjunctiva images and was then augmented to 2635. Moreover, the origin of the dataset used by Refs. [22,26] was not indicated, while the source or

origin of our dataset for the study was indicated in the study. In addition, most of the reviewed studies mainly in Table 1 used only one machine learning model for their various studies.

Comparing algorithms in machine learning for disease diagnosing or detection is recommendable because it is a way to measure the performance of each algorithm or model and determine which is the most accurate and reliable. This helps to ensure that the correct diagnosis is made and that the best possible treatment is provided. Additionally, the comparison will reveal potential weaknesses and strengths in the algorithms, which can be used to further improve them. However, this study compared five (5) machine learning algorithms to detect anaemia with the use of the conjunctiva of the eye images which seeks to address and bridge the gaps identified in the related works.

2.2. Limitations encountered during this study

The collection of anaemia datasets is generally challenging and charged with difficulties. The majority of the anaemia datasets in Table 1 are made up of fewer datasets whereas the non-anaemic dataset predominates the anaemic datasets. Despite having a larger number of datasets with a balanced mix of anaemic and non-anaemic datasets, this study encountered various difficulties before and during the gathering of the dataset for the study.

i. **Ethical Clearance:** Because human lives are elaborated in this research, it has become necessary to obtain ethics approval for the study from the "Committee for Human Research and Ethics" and the various hospitals. This via numerous steps and took a long time. Despite the ethical clearance being obtained, some meetings with hospital administrators had to be postponed because of their busy work schedules.
ii. **Sample Selection:** Gathering the datasets was time-consuming, especially when it came to obtaining the conjunctiva images. This has become necessary since all the applicants involved were minors or children (age range of 6–59 months), which called for extra care and cautiousness to avert the finger to get into their eyes.
iii. **Cost:** The dataset gathering was costly as a result of various laboratory (lab) tests, laboratory officers training, and hospital visitations to have a series of meetings with health officers and administrators.
iv. **Inadequate Rate of Responses:** mostly, situations about anaemia are less informed as a result of inadequate education and awareness. Due to this, it became a challenge to gather the anaemia dataset for the study. The majority of parents and guardians also expressed reluctance to permit their children to participate in the survey, though they did not explain why. However, following several meetings and thorough explanations

Table 1
Comparative study of other related works.

S/N	Paper/ Article	Utilized Algorithm	Evaluation Metrics	Image Augmentation	Image Augmenting Technique	Dataset Source	Dataset Size	Achieved Results (Accuracy)	Remarks
1	Jain et al. [21]	ANN	Accuracy, sensitivity, specificity, and confusion metric	Applied	Translate, Rotate and Mirroring	Italy and Idia	099	Achieved an accuracy of 97.00 %	The ANN was able to detect anaemia with 97.00 % accuracy.
2	Asare et al. [38]	CNN, SVM, k-NN, Decision Tree, Naïve Bayes,	Accuracy, F1-Score, AUC, Recall and Precision	Applied	Translation, Rotation and Flipping	Ghana (10 medical facilities)	527	The CNN achieved the highest accuracy of 99.12 %, while the lowest accuracy of 95.4 % was achieved by the SVM	The study indicated that the palpable palm has a higher potential of detecting anaemia in children with the application of non-invasive approaches such as machine learning models
3	Magdalena et al. [22]	CNN	Precision, recall, accuracy and f1-score	Applied	Not indicated	Not indicated	2000	An accuracy of 94.00 % was achieved	The Adam optimizer was utilized with a learning rate of 0.0001.
4	Appiahene et al. [28]	Decision Tree, Naïve Bayes, k-NN, and CNN	Accuracy, F1-Score, AUC, Recall and Precision	Applied	Flipping (horizontal and vertical), Rotation (90^{o} and 180^{o}) and Translation	Ghana	527	99.92 % accuracy was achieved by the CNN and a recall of 99.98 % achieved	The palm is one of the ideal spots of the human body to detect anaemia in children since the eye may be exposed to falling objects.
5	Agrawal [25],	CNN	Precision, AUC, weighted log loss metric	Applied	Flip, Zoom, Rotate and Translate	India	400	74.00 % accuracy was achieved by CNN which was the highest.	The CNN achieved 74.00 % accuracy while the AUC and precision of 62.23 % and 59 % respectively.
6	ppiahene [39]	CNN, Logistic Regression, and Gaussian Blur	Accuracy, Specificity and Sensitivity	Not applied	Not indicated	Ghana	710	The study achieved an accuracy of 92.50 % when embedded in a smartphone application	On average performance of 50 s, the smart application can detect anaemia using a real-time dataset.
7	Peksi et al. [26]	Naïve Bayes	Accuracy, specificity and sensitivity	Not applied	Not indicated	Not indicated	020	90.00 % accuracy was achieved	The Naïve Bayes achieved a signi-ficant accuracy of 90.00 %
8	Bauskar et al. [27]	SVM	Accuracy, specificity, and sensitivity Accuracy	Not applied	Not indicated	India	099	Achieved an accuracy of 93.00 %	The SVM was able to achieve 93.00 % accuracy

of the study's advantages to the medical community, the majority of parents or guardians gave their approval.

Regards these, we advise that ethical consent clearance should be attained on time before progressing to undertake studies involving human participants.

3. Methodology

In this section, the method used for this study is detailed and discussed with the various approaches applied to the proposed models. This section is made up of three phases/steps which are detailed in phases/steps I, II and III respectively. Also, a concise step or phase of the proposed methodology has been indicated from i-iii below.

i. Collection of datasets, which are conjunctiva images detailed in sections 3.1 and 3.2.
ii. Preprocessing of images (dataset), which goes through the extraction phase, enhancing the images, segmenting the various ROI and deducing the conjunctiva photos to obtain their diverse colour components (CIELAB colour space), which is detailed in sections 3.3–3.5.
iii. Developing the proposed algorithms to detect the anaemia conditions with the application of convolutional neural network, naïve Bayes, k-nearest neighbor, support vector machine and decision tree as detailed in sections 3.6–3.12.

These models were considered because they have proven to be robust in clinical disease diagnosis or detection which include anaemia, cancer, diabetes, etc., mostly with the application of clinical images through the utilization of machine learning approaches.

Phase/Step I; Dataset Collection.

3.1. Application of ethical and consent for the study

Before the commencement of the study, the "Ethics Committee for Human Research of the University of Energy and Natural Resources, Ghana approved the collection of datasets to be used for this work". Furthermore, because the participants (patients) in the study were minors, ethical consent was obtained from their parents or guardians and the purpose and objectives of the study were explained to them, along with the benefits it would bring to health services [28,38,40]. Additionally, the designations and participants' names were not captured during the image-gathering process, which makes their identities anonymous.

3.2. Dataset collection App development, training of officers, and dataset collection

The Kobo Collect applications system was developed for dataset collection by training licensed biomedical scientists on how to use the application platform for the datasets to be collected. The designed application contained a form to take patient biodata such as age, sex, and a comment based on the value of the haemoglobin measured at the time of the laboratory assessment. Afterwards, the image of the eye is taken to extract the ROI which is the conjunctiva.

The dataset for the study was aimed at children below the age of six, that is, from 6 to 59 months with the haemoglobin range $<$11gd/L for anaemic while $\geq$ 11gd/L for non-anaemic. Table 2 shows samples of a demographic description of participants (patients) for the dataset collection.

All images were taken by licensed biomedical scientists with a camera with high-quality standards with 12.00 Megapixels minimal resolution.

To capture the image of the conjunctiva of the eyes, the lower eyelid was gently pushed back with the thumb and possibly with the support of the index finger. Furthermore, the cameras' spotlights were turned off when capturing the photographs to prevent excessive shine effects created by the picture quality, that is to eradicate ambient light, which dramatically influences detection by the models. This method is a great technique to remove the effect of ambient light on photographs in datasets (see Fig. 1 shows the conceptual framework of the meethodology utilized for the study).

Table 2
Demographic details of some participants (patients) for the dataset collection.

Image Number	Haemoglobin Level	Age (in months)	Gender	Remark
Image1	11.9	48	Female	Non-anaemic
Image2	9.9	12	Female	Anaemic
Image3	9.9	48	Male	Anaemic
Image4	6.1	24	Female	Anaemic
Image5	11.16	12	Female	Non-anaemic
Image6	6.8	12	Female	Anaemic
Image7	11.2	12	Male	Non-anaemic
Image8	8.6	12	Female	Anaemic
Image9	11.2	24	Female	Non-anaemic
Image10	8.7	59	Female	Anaemic
Image11	11.2	48	Female	Non-anameic
Image12	8.7	59	Male	Anaemic
Image13	11.3	2	Male	Non-anaemic
Image14	8.7	5	female	Anaemic

The following hospitals, all of which are in Ghana, provided the datasets for this study: "Komfo Anokye Teaching Hospital, Bolgatanga Regional Hospital, Kintampo Municipal Hospital, Ahmadiyya Muslim Hospital, Sunyani Municipal Hospital, Manhyia District Hospital, Ejusu Government Hospital, SDA Hospital, Nkawie-Toase Government Hospital, and Holy Family Hospital". To collect the datasets and upload them to the database, the biomedical researchers received instructions on how to utilise the Kobo Collect application. Fig. 2 displays an example of the unprocessed conjunctiva of the eye photographs, with [A] designating anaemic images and [B], non-anemic images [28,38–40].

Phase/Step II; Dataset (Image) Preprocessing.

3.3. Extraction of Region of Interest (ROI)

After successfully extracting the ROI from each image, we rotated, flipped, and translated the images to enhance the size of the original datasets because limited data sizes might lead to overfitting [21,41]. The ROI of the conjunctiva of the eyes were extracted after the application of the triangle thresholding algorithm, in combination with the entropy grayscale image algorithm. The triangle thresholding algorithm is one of the simplest methods for separating an image's foreground and background based on the various hues or intensities [42]. To locate the features in the other domains, transform techniques are employed [43]. A sample of the marked ROI is shown in Fig. 3, while Fig. 4 also shows samples of extracted images of the conjunctiva of the eyes where [A] and [B] are for anaemic and non-anaemic images respectively.

3.4. Augmenting the images

Using image augmentation techniques, the original dataset's size was enhanced since machine learning models work better with large datasets [44]. This helped prevent models from performing poorly [27,45]. The rotate, translate, and flip augmentation techniques were utilized in this work to improve the initial dataset. The mean intensity component of the cropping and Gaussian augmentation procedures is altered when

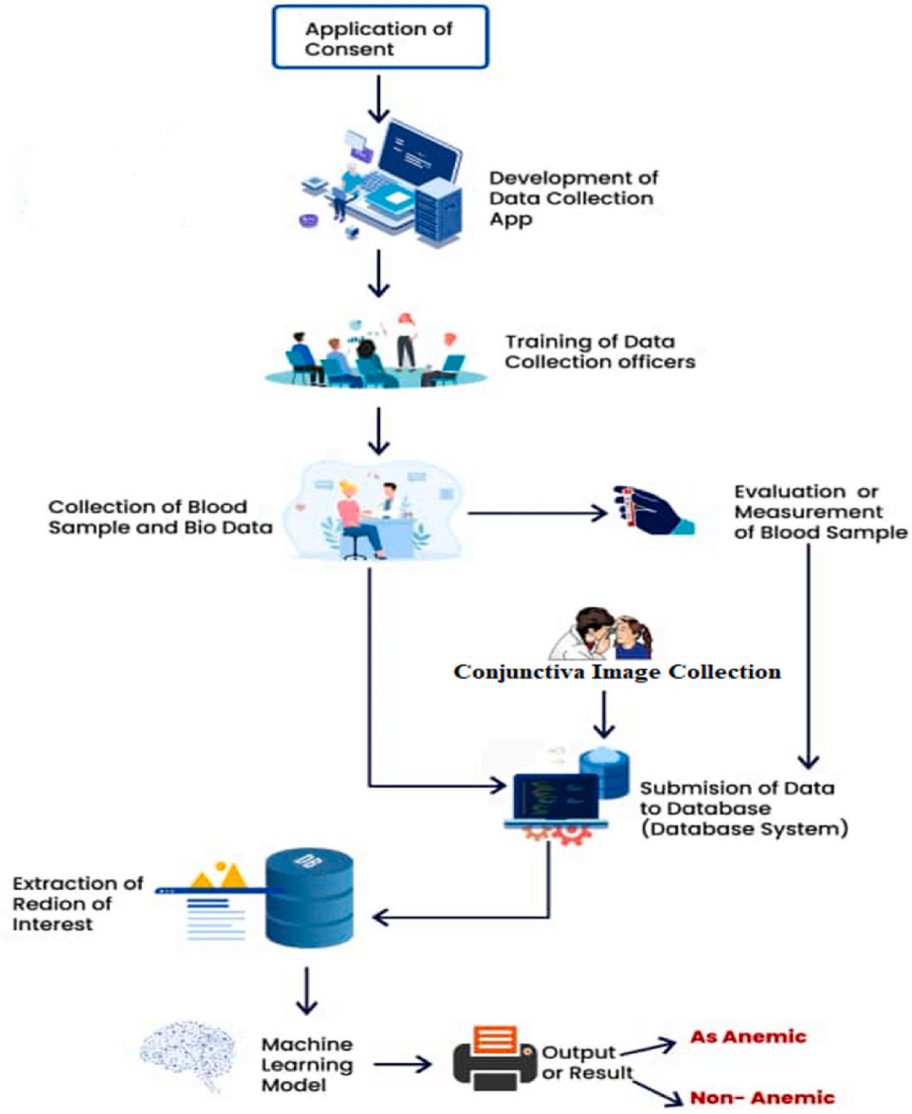


Fig. 1. A proposed framework of the conceptualized methodology [38].

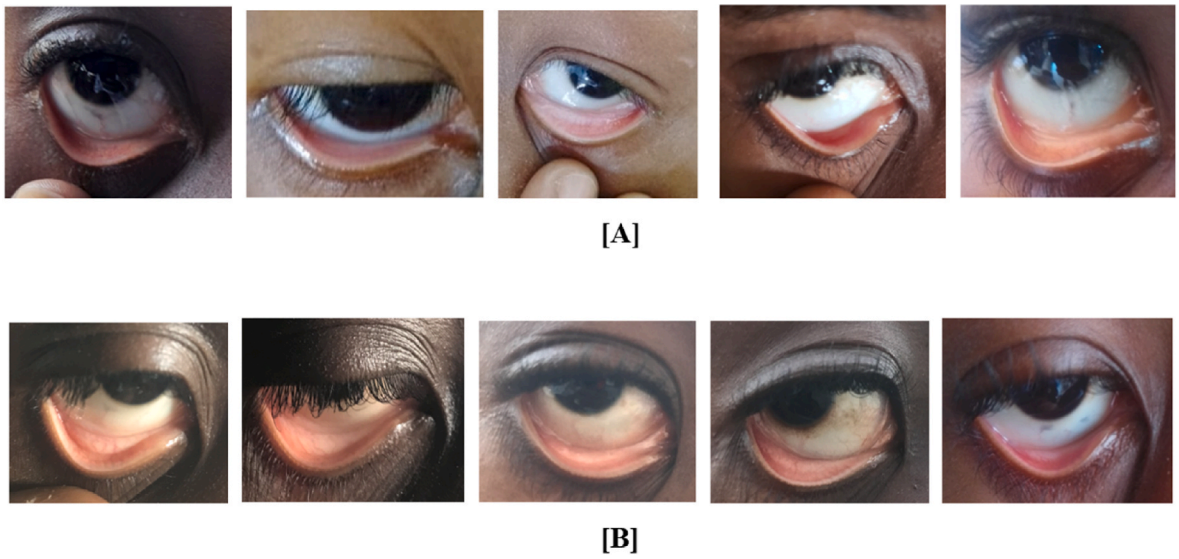


Fig. 2. Unextracted raw image samples, where anaemic images and non-anaemic images are represented by [A] and [B] respectively [40].

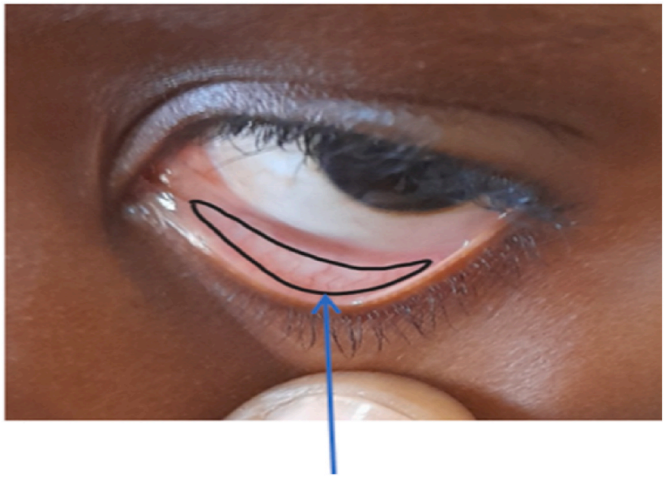


Fig. 3. Region of interest (ROI) [38].

used, hence they were not used [21]. The application of data augmentation in machine learning for medical disease detection or diagnosis has the advantage of enhancing the detection and classification accuracy of images [46,47].

A 90-degree and 270-degree rotation were utilized as a procedure to augment the initial images, while the vertical and horizontal techniques were also used to flip the initial images, and lastly, translate them to the X and Y positions. The technique utilized for image augmentation in this study is shown in Fig. 5 below, while the mathematics functions used to augment the images are specified in equations (1)–(5) (Fig. 6 shows a graphical representation of a 3-dimensional CIELAB colour space).

The rotation angle θ for the equations for the latest coordinates of a pixel in an anti-clockwise rotation is defined as:

$$x' = x\cos\theta - y\sin\theta \quad (1)$$

$$y' = x\sin\theta + y\cos\theta \quad (2)$$

In a related interpretation, every pixel's position is (x', y') and can be used as rotation and vector-matrix in the current image as shown in

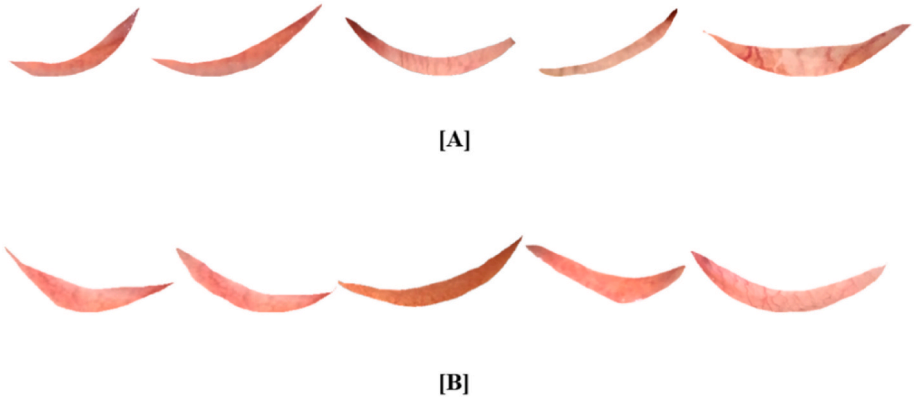


Fig. 4. Image extraction samples where anaemic and non-anaemic images are represented as [A] and [B] respectively [38].

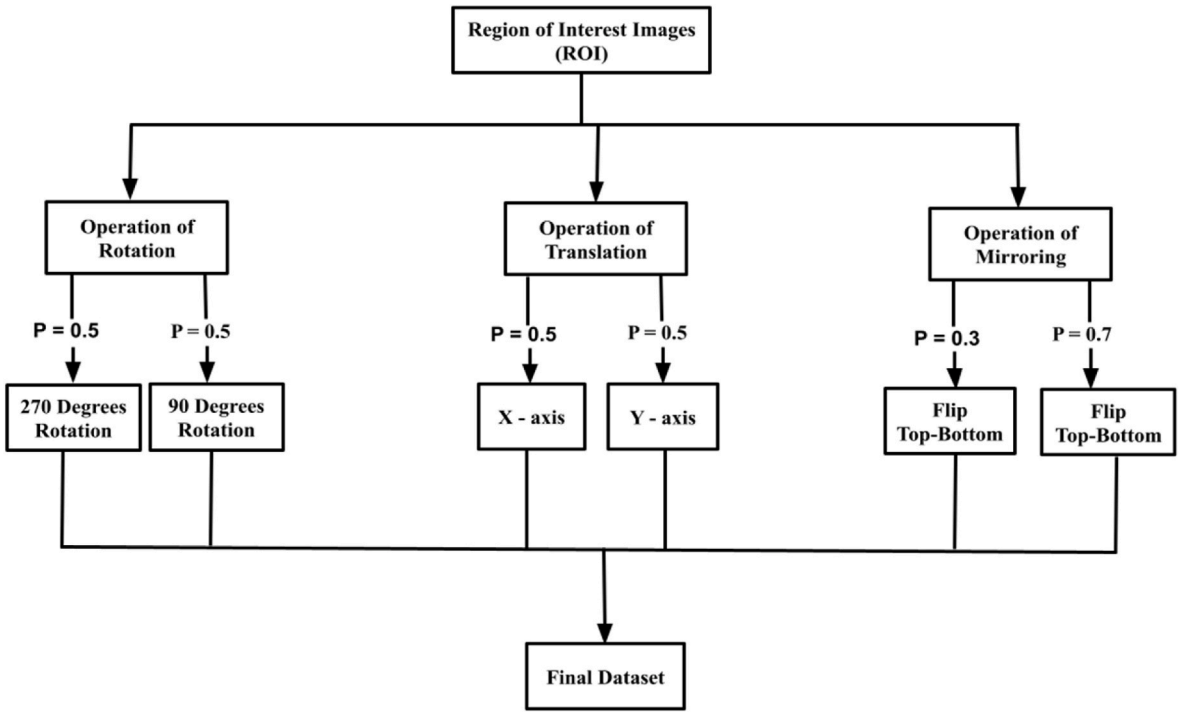


Fig. 5. Graphical representation of the image augmentation technique, P is the probability [28].

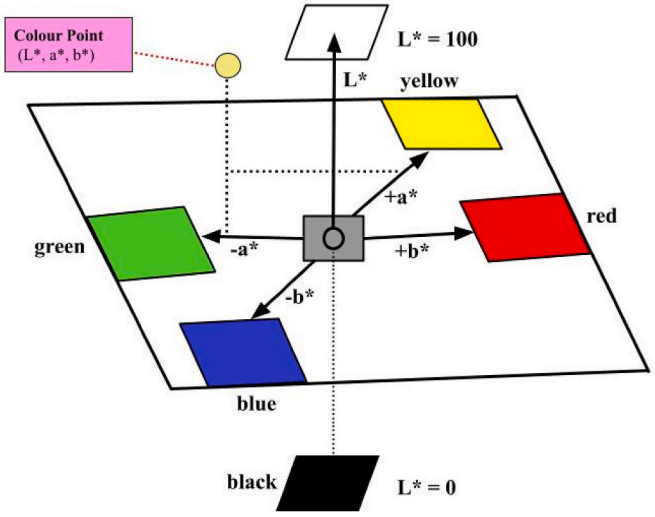


Fig. 6. A graphical representation of a 3-dimensional CIELAB colour space [49]. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

equation (3);

$$\begin{matrix} x' \\ y' \end{matrix}\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos\,\theta & -sin\,\theta \\ sin\,\theta & cos\,\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \tag{3}$$

The rotation (x_0, y_0) coordinate the point of derivation $(0, 0)$ and also rotate the image at the midpoint or around a specific point. For the image to be rotated around the point (x_0, y_0), it is expressed as shown in equations (4) and (5).

$$x' = x_0 + (x - x_0)cos\,\theta + (y - y_0)sin\,\theta \tag{4}$$

$$y' = y_0 + (x - x_0)Sin\,\theta + (y - y_0)cos\,\theta \tag{5}$$

The pixel's position is located at (x_0, y_0) does not move and always remains fixed. An equation that does not need to be solved maintains its original image at *x and y*.

3.5. CIELAB (CIE L*a*b*) colour space acquisition process

The position (i, j) of each pixel of the entropy is calculated by the values of the pixel with a centred region of a 2-dim at (i, j) in which the image entropy is calculated. Mostly, the size of the region is configured to be $(2N \times 2N) = (10, 10)$. Subsequently, the extracted regions of interest (ROI) are separated into metrics using the image intensity value of the CIELAB colour space.

For the quality of datasets to be acquired, the datasets needed to be preprocessed. During the preprocessing, 38 images (for both anaemic and non-anaemic images) were eliminated or rejected for low quality and with too much light to get the true colour of the mean component of the CIE L*a*b* of the conjunctiva. Each picture's ROI was transformed into the CIELAB (also known as CIE L*a*b*) colour space model based on the results of image preprocessing and dataset construction. The purpose of the L*a*b colour space was to replicate human perception or vision.

It is depicted using the standard deviation of the ROI a* components' mean value. The a* components with average values are red (a* > 0) and green (a* 0) [2]. Early studies in this area indicate that a significant correlation between the a* components and the haemoglobin values can be calculated with the application "Pearson's correlation index". Many studies in this area have demonstrated that higher haemoglobin values a mean component amount of a* ≥160, and an average value of a* < 142 are aligned with patients with low haemoglobin values [48].

The average values of the a* component are therefore known to be quite selective. That is, anaemic and non-anaemic individuals may be distinguished better by the mean intensity of the red and green components. The records are evaluated to detect anaemia. To begin, image colour characterization was conducted using the CIELAB colour space. The specific colour space translates all noticeable colours into an arithmetic operation space of three 3-dimensions specifying an independent approach for device-independent numerical illustration.

Comparative differences among two colours in the L*, a* and b* are assessed because the components indicated above are associated with the general variation in the supposed colour. By considering each hue as a point in three dimensions with L*, a*, and b* as its three components, the Euclidean distance between each hue was determined. The Colour components a* and b* are represented by L* (Lightness), which stands for the darkest black at 0 and the lightest white at 100.

The "Cartesian coordinate system (a*, b*)" is defined by the existence of non-alignment grey in it. The b* axis displays the colours of the opposition, with yellow at b* >0 and blue at b* 0. Red and green are at a* >0 and a* 0, respectively, on the a* axis, representing the colours of the enemy. After multiple attempts, the greatest performance classified was acquired by absorbing a characteristic of a 3-component, that is "a*, b*, and the G value produced from the RGB component images". The relationship between the conjunctiva of the eye and the haemoglobin level is estimated using the colour scheme.

The purpose of converting the images from RGB space to CIE L*a*b* space in this study was to analyze the colour data of the images more accurately. This conversion eliminates the influence of different displays and lighting conditions on the colour data, allowing for better comparison between images. The study found that converting the images from RGB space to CIE L*a*b* space resulted in more accurate and consistent colour data, leading to more reliable results. Comparative results showed that when using CIE L*a*b* space, differences in colours between images were much smaller than when using RGB space. This suggests that CIE L*a*b* space is more accurate and reliable for comparing colours between images.

To filter the input data, the values of the RGB and L components were retained. Following that, the images of the conjunctiva of the eye were categorized as either anaemic or non-anemic depending on the haemoglobin result from the laboratory test. The initial and augmented images of the conjunctiva of the eyes are thoroughly described in Table 3, and Table 4 gives the statistical description and the level of concentration of the haemoglobin value (g/dL) on the actual dataset.

Equations (6)–(9) show the mathematics approach for the CIELAB colour space

$$L^* = 116f\left(\frac{Y}{Yn}\right) - 16, \tag{6}$$

$$a^* = 500\left\{f\left(\frac{X}{Xn}\right) - f\left(\frac{Y}{Yn}\right)\right\}, \tag{7}$$

$$b^* = 200\left\{f\left(\frac{X}{Xn}\right) - f\left(\frac{Z}{Zn}\right)\right\}, \tag{8}$$

Where, $f(s) = s^{\frac{1}{3}}$, for $s > 0.008856$, and $f(s) = 7.787s + \frac{16}{116}$, for $s \leq 0.008856$.

The difference (ΔE) between the CIELAB colour space is defined as;

$$\Delta E = \sqrt{\left(L_2^* - L_1^*\right)^2 + \left(a_2^* - a_1^*\right)^2 + \left(b_2^* - b_1^*\right)^2} \tag{9}$$

Table 3
Datasets (Original and Augmented) Description after preprocessing.

	Anaemic	Non-anaemic	Total Dataset
Original Dataset	304	223	527
Dataset after Augmentation	1520	1115	2635

Table 4
Statistical analysis and haemoglobin concentration leve (g/dL) [38,40].

	Patient Class		
	Anaemic	Non-anaemic	Total
Patients	304 (58.00 %)	223 (42.00 %)	527 (100 %)
Female	180 (60.00 %)	118 (40.00 %)	298 (57.00 %)
Male	124 (54.00 %)	105 (46.00 %)	229 (43.00 %)
Anaemia Classification (diagnosis) from 6–59 months			
Classification of Anaemia	Anaemic	Non-anaemic	
Haemoglobin Value	<11g/dL	≥11g/dL	

Where a* and b* represent the colour channels, L* is the darkest colour and ΔE *unity* values represent the JND (just a noticeable difference).

Phase/Step III; Proposed Models (Machine Learning) Development and Anaemia Detection.

3.6. Convolutional neural network (CNN)

Convolutional Neural Network (CNN) is a type of machine learning algorithm in which data in the form of images are extracted and processed by using classification [22]. CNN is made up of two main components: feature extraction, which uses various descriptions to improve the accuracy of the data being processed. After the extraction of the data characteristics, it seeks to extract crucial information from the data and the classification layer. Fully linked neurons are used in this process to change the data's dimension [50,51].

In this study, CNN applied the AlexNet architecture and the Stochastic Gradient Descent (SGD) optimization and ReLu was used to train the models. The activation function used a regularization of $a = 0.001$, with a minimal iteration of 15. The AlexNet architecture comprised 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer. AlexNet has a neat architecture, which has a structure similar to VGGNet in CNN's deep learning-based architecture [31].

The application of deep learning often include training a neural network on a collection of matched inputs and outputs, to minimise the gap between the projected values and some known right outputs [52]. Large data sets are frequently needed for this, which means they must be handled manually and are prone to inaccuracy, however, this can be avoided by minimizing the residuals of the predetermined set of training points and the governing equations [52]. Selecting an activation function for the network's output layer and hidden layers is a crucial decision to be made throughout the neural network development process. After applying an additional "bias" and the activation function, the weighted sum of the inputs makes up a layer's net output [53].

The primary goal of the activation function was to change the nodes' signal input into a signal output. Without the activation function, the CNN is reduced to a linear regression. In other words, the CNN couldn't train sophisticated models without the activation function.

The mathematical model for the CNN is represented as:

$$z = w^T \cdot x \cdot b \tag{10}$$

Where x is the input, w is the weight, and b is the bias. The proportion is chosen at random to begin the matrix = w. The back-propagation rule, on the other hand, has the mathematical model denoted as;

$$f^{'}(x) = (1 + e - x) - 1[1 - (1 + e - x) - 1] \tag{11}$$

$$f^{'}(x) = sigmoid(x)[1 - sigmoid(x)] \tag{12}$$

$$\frac{\partial O}{\partial Z2} = (O)(1 - O) \tag{13}$$

The model's performance is greatly impacted by the input data normalization methods and neural network architecture selection. Attaining optimal outcomes necessitates experimentation, hyperparameter adjustment, and comprehension of the unique features of the dataset.

3.7. K-nearest neighbor (k-NN)

Regarding the comparison to train, validate and test the dataset, the k-NN algorithm classifies similar data [48]. The feature vectors are classified using k-NN to classify the best performance of the model where k = 1 [54]. The k-NN contains a small positive integer which is represented as "k". Due to the Euclidean metric and uniform weight being used, the k-NN has 100 neighbors [28]. There is the application of vote which contains a majority of neighbors used in the experimental setup. The mathematical notation for the k-NN algorithm is as shown in equations (14) and (15):

$$\mathrm{d}(x, x^{'}) = \sqrt{(x, -x_1)^2 + div + \left(x_n - x_n^{'}\right)^2} \tag{14}$$

$$P(y = j | x = x) = \frac{1}{k} \sum_{i \in A} I\left(y^{(i)} = j\right) \tag{15}$$

The input x is assigned to the maximum possible value of the class.

3.8. Naïve Bayes

With the use of algorithm classifiers which are independent in assumption, the theorem of Bayes is applied in Naïve Bayes. A small portion of the training dataset is deemed suitable for computing the mean and variance of the variables associated with it using the Naïve Bayes algorithm while dividing fields that continue into discrete bin and target value fields. This method works by assuming that there is no correlation between the presence or absence of an attribute [50]. The mathematical model of the Naïve Bayes is shown in equation (16) as;

$$P(c/x) = \frac{P(x/c)P(c)}{P(x)} \tag{16}$$

where $P(c)$ represents the Prior Probability of the class, $P(x)$ represent the Prior Probability of the Predictor, $x(c)$ represent the likelihood and P represent the Posterior probability.

3.9. Support vector machine (SVM)

Extensively, the Support Vector Machine (SVM) is utilized to classify numerous elements which are merged from initially used methods. The assumption that the "data are separable" that is, that they can be divided into groups by a functional separator—underlies SVM's similarity to discriminant analysis [55].

Sigmoid was employed with the SVM, with a limit of 100 iterations, cost (c) equal to 100, and an epsilon of regression (ε) of 1.10 assigned. The 0.1000 value was chosen for the numerical tolerance. Fundamentally, SVM is based on theorems for statistical learning and intensely enlarges the annotation of "separability" centred on numerous principles.

For the SVM, a hyperplane is represented by a vector weight (w) and a bias (b) that are parameterized as shown by equations (17)–(21);

$$w \cdot x + b = 0 \tag{17}$$

$$w \cdot x + b = -1 \tag{18}$$

$$w \cdot x + b = 0 \tag{19}$$

The hyper-plane function is emitted for the datasets utilized to train, validate and test the models can be categorized as shown in equation (20);

$$f(x) = sign(w \cdot x + b) \tag{20}$$

If the kernel function is used, the above-mentioned function could be

stated as;

$$f(x) = sign\left(\sum_{i=1}^{N} a_i y_i k(x_i, x) + b\right) \quad (21)$$

3.10. Decision tree

When utilizing the Decision Tree approach to discretely compute target-valued functions, the datasets are presented using a structured tree. When tree examples are ordered from the root to the leaf nodes, classification is utilized [48]. The binary tree's maximum was used for smaller subsets of less than 5, and subsets with less than five were not divided. The leave's minimum number of occurrences was set at 10, and there had to be at least 100 trees. Each branch of the tree represented the value of each attribute, and the attribute also represented each node of the tree. The Decision Tree is mathematically expressed as shown in equations (22) and (23);

$$H(s) = [-P\log 2\ \ (P+) - [-P\log 2\ \ (P-)] \quad (22)$$

The information gain is defined as;

$$Gain(S, A) = H(s)\sum_{/s/}^{/Sv/} H(Sv) \quad (23)$$

Where $(P+)$ represents the percentage of the positive class, $(P-)$ represent the percentage of the negative class and P represent the probability.

Improving generalization is an essential part of creating strong machine learning models. The capacity of a model to perform effectively with unknown data and be relevant to a variety of scenarios is known as generalization. Image augmentation, a 10-fold cross validation, normalization and hyperparameter tunning are the mechanisms used to generalize the machine learning models in this study. Normalization reduces the sensitivity of the model to the magnitude of input characteristics, which helps improve generalization to previously unknown data.

3.11. Dataset utilized to train, validate and test the models

The datasets used to train, validate, and test the models were separated into 70 %, 10 %, and 20 % accordingly by the use of a random sampling strategy, which was used after data augmentation for the entire dataset of 2635. The models were validated using a 10-fold cross-validation procedure. Table 5 outlines the segmentation of the dataset used to develop, test, and validate the models. The 10-fold cross-validation implementation during hyperparameter tuning is utilized to find the global minimum. This involves an outer cross-validation loop for model evaluation and an inner loop for hyperparameter tuning. It provides a more robust estimate of the model's generalization performance.

3.12. Performance measures evaluations

We evaluated the performance of all the proposed machine learning models to assess the results achieved by every individual model with the use of evaluation matrices which are, precision, recall, F1-score and AUC using 10-fold cross-validation in the anaemia detection.

Table 5
A Tabular narration of Segmented Dataset to Train, Validate and Test Models.

	Conjunctiva of the Eye Images		Total Images
	Anaemic	Non-anaemic	
Dataset for Training (70.00 %)	1064	781	**1845**
Dataset for Validating (10.00 %)	152	111	**263**
Dataset for Testing (20.00 %)	304	223	**527**
Total Dataset (100 %)	**1520**	**1115**	**2635**

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (24)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (25)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (26)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (26)$$

$$\text{AUC (TPR)} = \frac{TP}{TP + FN} \quad (27)$$

$$\text{F1-Score} = \frac{2(P.R)}{P + R} \quad (28)$$

Precision is denoted by P, recall by R, and the terms TNR and TPR, which, respectively, stand for True Negative and True Positive Rate. The number of samples that are True Positive (TP) once valid values have been anticipated, discovered, and/or are positive. False Negative is abbreviated FN while True Negative is abbreviated TN.

In this study, our assumption was to initially gather a larger dataset sample and achieve significant accuracy in the detection of anaemia using the proposed machine learning models. This encouraged us to use more machine learning models to compare their performance and that the models would be able to generalize to unseen data regarding the detection of anaemia. The dataset employed to train, validate and test the models is representative of the population of patients with the age range of 6–59 months, and the parameters used to develop the models are predictive of the target disease class.

Additionally, it is assumed that the models are trained, validated, and tested using appropriate methods, such as cross-validation, to ensure that the results are valid and generalised. The assumptions made during the machine learning training, validation, and testing stage for anaemia detection of this research work could be evaluated by the accuracy of the dataset used and the selection of the parameters used to train the models. It is also important to consider whether the models are appropriately validated to ensure they are generalizable to the population of patients.

Furthermore, if the data is not properly collected or labelled, the results of the model may be inaccurate. In addition, the accuracy of the model is also dependent on the quality of the data. If the data is noisy or contains outliers, the model may not be able to accurately predict the disease. Overall, while the assumption of a representative dataset is important, it is also important to verify that the data is of high quality and properly labelled. This will help ensure that the machine learning model is accurate and reliable.

Finally, it is important to consider the potential ethical implications of using machine learning for anaemia detection using conjunctiva of the eye images, including potential bias in the models, privacy concerns, and potential misuse of the results. However, the names and faces of participants were not exposed during the image capturing, which makes their identities unknown.

4. Results and discussion

The models proposed by this study compared five machine learning techniques, that is, convolutional neural network, k-nearest neighbor, Naive Bayes, support vector machine, and Decision Tree, used to identify anaemia. The CNN model was trained using the optimization and activation functions, regularization of SGD and ReLu, and a = 0.0001. Because the CNN operates as a linear regression without the activation function, transforming the signal node input to the signal output, these procedures are carried out by the activation function. With a maximum

of 100 iterations, the sigmoid was used for SVM operations. The numerical tolerance was 0.1000, while the regression epsilon (ε) was set at 1.10.

The restriction was set at a minimum of 100 trees, and the minimum number of neighbors with equal weight was set at 100. The Euclidean metric was created, and the Decision Tree's binary tree was generated using the fewest possible instances in the leave to 10. Subsets below five were not divided.

The orange data mining software was used for the experimental setups. Using Nave Bayes, k-NN, and Decision Tree, the study by Naik & Samant [56] used orange data mining tools to diagnose liver disorders with impressive accuracy. Additionally, to identify and diagnose diabetes, Peker et al. [57] employed the Random Forest, Artificial Neural Network, k-nearest Neighbor, Support Vector Machine, and Decision Tree Algorithms in the orange data mining software. The ANN demonstrated the best accuracy (90.27 %), while the SVM had the lowest accuracy (64.66 %) among the other methods. This shows that the orange data mining software is a valuable tool for diagnosing or detecting medical diseases, of which anaemia detection is no exception.

Figs. 7–9 illustrate how detection accuracy, AUC, precision, recall and f1-score were taken into consideration while assessing the performance of the machine learning models. Before testing the models on the datasets, 10-fold cross-validation was employed for validating to prevent overfitting the performance of the models.

4.1. Results

Having trained, verified, and tested the models using 2635 data, that is, 1520 are anaemic and 1115 are non-anaemic, the proposed models achieved a good result (presented in Table 6) for the detection of anaemia using images of the conjunctiva of the eyes (Table 7 shows results of related studies in comparison with the proposed study.).

4.2. Comparison of results of related studies

4.3. Discussions

Iron deficiency anaemia (IDA) detection in children is an important healthcare task with great potential for applying machine learning techniques, particularly with conjunctiva of the eye images. From the results achieved by this study, the convolutional neural network

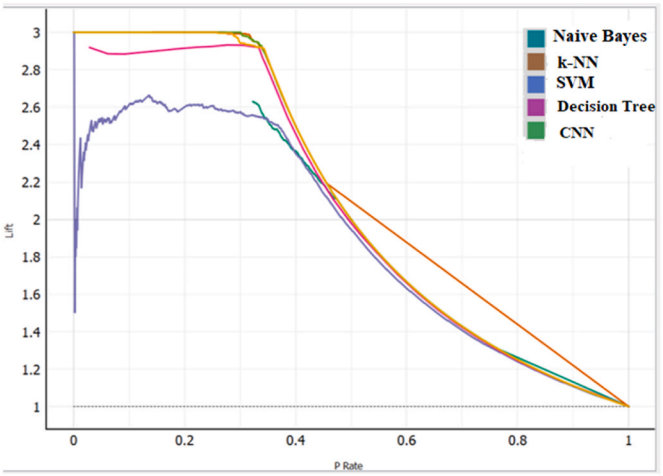


Fig. 8. Lift Curve for the models' performance.

achieved the highest accuracy of 98.45 % and a running time of 6 s. However [22], achieved 94.00 % accuracy when a convolutional neural network technique was used to detect anaemia, the results achieved in their study were higher than those achieved by Ref. [23]. Also, a deep-learning-based approach was utilized by Ref. [31] to detect anaemia which achieved a significant result of 97.94 % by the VGGNet and 95.07 % by the DenseNet. However, the proposed CNN model used in this study performed better (98.45 %) than [22,23,31] with an AUC performance of 999.93 % with an execution time of 6 s. A high Area Under the ROC Curve (AUC) suggests excellent discrimination between classes, with a minimal chance of misclassification, while the short running time is advantageous, especially in real-time or resource-constrained applications.

The k-NN achieved an accuracy of 97.96 % and a running time of 17 s. An accuracy of 97.96 % was achieved by the decision tree algorithm and was executed in 8 s as the running time. The Naïve Baye was executed in 10 s and achieved an accuracy of 94.94 % as compared to Ref. [26] which achieved 90.00 % accuracy when the naïve Bayes model was employed for anaemia detection with the use of the conjunctiva of the eye images as shown in Table 6.

Even though the Support Vector Machine achieved an accuracy of 89.45 % which is the lowest among all the models, it had a better running time of seconds than the k-NN and Naïve Bayes running time. In

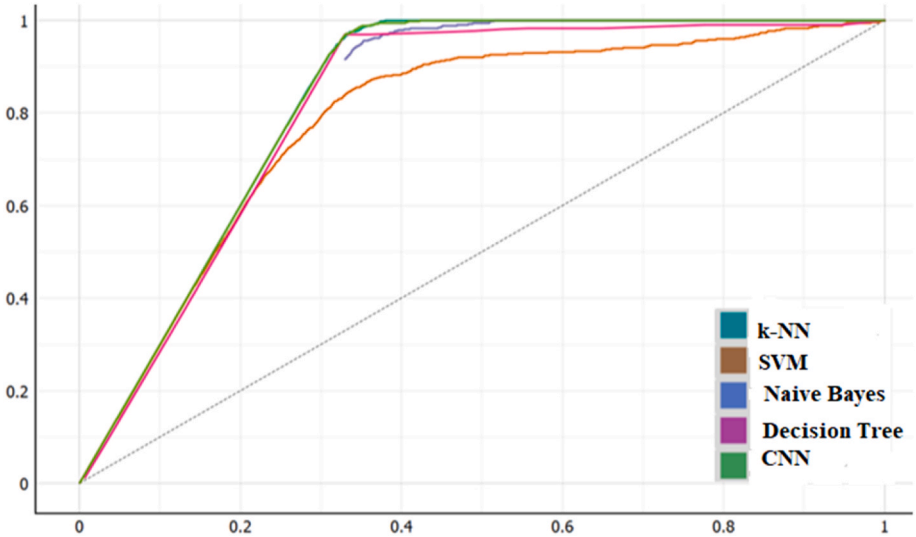


Fig. 7. Auc analysis of the anemic detection by the models in a graphic illustration.

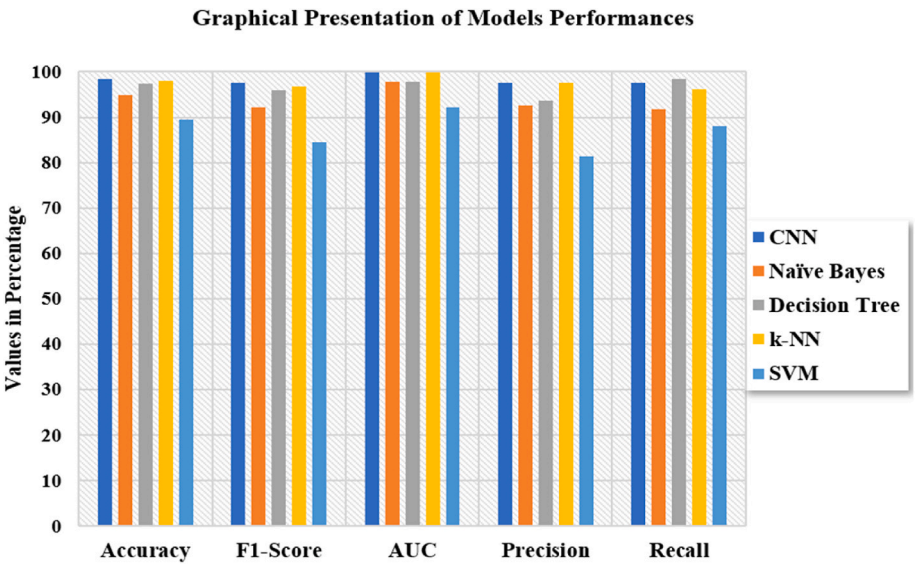


Fig. 9. A graphical presentation of the models with their corresponding evaluation metrics.

Table 6
Results of the performance of the proposed models with the corresponding evaluation metrics and running time.

S/N	Algorithm Used	Models Performances (%)					Running Time
		Accuracy	F1-Score	AUC	Precision	Recall	
1	Convolutional Neural Network (CNN)	98.45 %	97.63	99.93	97.64	97.63	6 s
2	Naïve Bayes (NB)	94.94 %	92.24	97.74	92.64	91.84	10 s
3	Decision Tree (DT)	97.32 %	96.02	94.70	93.67	98.49	8 s
4	k-Nearest Neighbor (k-NN)	97.96 %	96.86	99.86	97.60	96.13	17 s
5	Support Vector Machine (SVM)	89.45 %	84.53	92.16	81.34	87.98	8 s

Table 7
Results of related studies in comparison with the proposed study.

Paper	Utilized Algorithm	Results Achieved (Accuracy)
Tamir et al. [12]	Support Vector Machine	78.90 %
Jain et al. [21]	Artificial Neural Network	97.00 %
Magdalena et al. [22]	Convolutional Neural Network	94.00 %
Delgado-Rivera et al. [23]	Convolutional Neural Network	77.58 %
Peksi et al. [26]	Naïve Bayes	90.00 %
Acar et al. [31]	Convolutional Neural Network (VGGNet)	97.94 %
	Convolutional Neural Network (DenseNet)	95.07 %
Irum et al. [58]	Support Vector Machine	78.90 %

addition, the SVM performs better than the accuracy achieved by Refs. [12,58] which both achieved an accuracy of 78.90 % each. Table 6 and Fig. 9 give details of the results of all modes with their corresponding evaluation metrics. Figs. 7 and 8 were used to compare the performance of the models, assess the accuracy of the models, and identify areas for improvement. The higher the curve, the more accurate the model is at detecting positive cases. Involving medical professionals within the model creation process guarantees conformity to clinical procedures and amplifies the model's practicality. It is necessary to research to verify the model's predictions through clinical evaluations to ascertain its dependability and efficiency.

Model performance evaluation in the context of anaemia diagnosis depends heavily on metrics like sensitivity, specificity, accuracy, and area under the curve (AUC). Improved model generalisation and resilience can be achieved by augmenting the dataset with original picture modifications (such as flips, zooms, and rotations). From the study it is evidence that conjunctiva of the eye images offers a convenient and non-intrusive data source. However, to reduce discomfort and possible hazards connected with invasive procedures, non-invasive screening approaches are especially beneficial in the field of paediatric healthcare.

4.4. Conclusion and Future works

In this work, we used images of the conjunctiva of the eyes to examine how well convolutional neural network, k-nearest neighbor, decision tree, Naive Bayes, and support vector machine performed in detecting anaemia. To prevent overfitting, which mostly happens when using limited datasets for machine learning techniques, the datasets for this study were enlarged from 527 to 2635 by applying picture augmentation. Results from each of the study's proposed models were significant, with the convolutional neural network achieving the highest accuracy (98.45 %), the support vector machine having the lowest accuracy (89.45), the k-nearest neighbor, the Decision tree, and Nave Bayes having the highest detection accuracy (97.96 %, 97.32 %, and 94.94 %, respectively), and their corresponding evaluation metrics being shown in Fig. 9.

As shown in Table 6 and Fig. 9, the models were assessed using Recall, Precision, F1-score, and AUC. Additionally, on 10 % of the total datasets, the models were verified using 10-fold cross-validation. The findings of this study suggest that by using images of the conjunctiva of the eyes, the convolutional neural network is effective in diagnosing anaemia. The utilization of conjunctiva images, machine learning can be a viable approach for detecting iron deficiency anaemia. Effective, accurate, and morally sound solutions that can greatly improve paediatric healthcare require collaboration between data scientists, physicians, and ethical concerns. The success of such applications will largely depend on ongoing research, validation, and integration with clinical

operations.

Since the palpable palm is easier to access than the conjunctiva, for further studies on this subject, the same machine learning models will be employed for anaemia detection using the palpable palm. Additionally, it can be challenging to focus on the conjunctiva of the eye while shooting pictures, especially of children below the age of 59 months, and items can fall on the eye when the conjunctiva is exposed. In addition, implementing Machine Learning-based conjunctiva image for detection and analysis of anaemia on a large scale can significantly improve public health by addressing issues connected to anaemia, especially in areas where the condition is more prevalent.

Institutional approval

The "Ethics and Consent Committee for Human Research of the University of Energy and Natural Resources, Sunyani, Ghana" approved this study before its commencement.

Funding acquisition

No funding was received for this study.

Availability of datasets

The datasets utilized for this study are available the Mendeley dataset repository as "Asare, Justice Williams; Appiahene, Peter; Donkoh, Emmanuel (2023), "CP-AnemiC (A Conjunctival Pallor) Dataset from Ghana", Mendeley Data, V1, doi: 10.17632/m53vz6b7fx.1"

CRediT authorship contribution statement

Justice Williams Asare: Writing – original draft, Visualization, Software, Resources, Methodology, Data curation, Conceptualization. **William Leslie Brown-Acquaye:** Writing – review & editing, Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Martin Mabeifam Ujakpa:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Formal analysis, Conceptualization. **Emmanuel Freeman:** Writing – review & editing, Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Peter Appiahene:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank all hospital administrators and medical laboratory scientists for assisting in the clinical evaluation and image acquisition process.

References

[1] Mitani A, et al. Detection of anaemia from retinal fundus images via deep learning. Nat Biomed Eng Jan. 2020;4(1):18–27. https://doi.org/10.1038/s41551-019-0487-z.
[2] Dimauro G, Caivano D, Girardi F. A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva. IEEE Access 2018;6:46968–75. https://doi.org/10.1109/ACCESS.2018.2867110.
[3] World Health Organization. Overview of anemia. Accessed: Jul. 16, 2022. [Online]. Available: https://www.who.int/health-topics/anaemia.
[4] Al-alimi AA, Bashanfer S, Morish MA. Prevalence of iron deficiency anemia among University Students in hodeida province, Yemen. Anemia 2018;2018:1–7. https://doi.org/10.1155/2018/4157876.
[5] Sevani N, Fredicia, V Persulessy GB. Detection anemia based on conjunctiva pallor level using *k-means a* lgorithm. IOP Conf Ser Mater Sci Eng Oct. 2018;420:012101. https://doi.org/10.1088/1757-899X/420/1/012101.
[6] Dimauro G, et al. Estimate of anemia with new non-invasive systems—a moment of reflection. Electronics (Basel) May 2020;9(5):780. https://doi.org/10.3390/electronics9050780.
[7] Nithya R, Nirmala K. Detection of Anaemia using Image Processing Techniques from microscopy blood smear images. J Phys Conf Ser Aug. 2022;2318(1):012043. https://doi.org/10.1088/1742-6596/2318/1/012043.
[8] Saputra DCE, Sunat K, Ratnaningsih T. A new artificial intelligence approach using extreme learning machine as the potentially effective model to predict and analyze the diagnosis of anemia. Healthcare Feb. 2023;11(5):697. https://doi.org/10.3390/healthcare11050697.
[9] Mazzu-Nascimento T, et al. Smartphone-based photo analysis for the evaluation of anemia, jaundice and COVID-19. International Journal of Nutrology Aug. 2021;14(2). https://doi.org/10.1055/s-0041-1734014. e55–e60.
[10] Chen Y-M, Miaou S-G. A kalman filtering and nonlinear penalty regression approach for noninvasive anemia detection with palpebral conjunctiva images. J Healthc Eng 2017;2017:1–11. https://doi.org/10.1155/2017/9580385.
[11] Dhakal P. Prediction of anemia using machine learning algorithms. Int J Comput Sci Inf Technol Feb. 2023;15(1):15–30. https://doi.org/10.5121/ijcsit.2023.15102.
[12] Tamir A, et al. Detection of anemia from image of the anterior conjunctiva of the eye by image processing and thresholding. In: 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). IEEE; Dec. 2017. p. 697–701. https://doi.org/10.1109/R10-HTC.2017.8289053.
[13] Waisberg E, Ong J, Zaman N, Kamran SA, Lee AG, Tavakkoli A. A non-invasive approach to monitor anemia during long-duration spaceflight with retinal fundus images and deep learning. Life Sci Space Res May 2022;33:69–71. https://doi.org/10.1016/j.lssr.2022.04.004.
[14] Vitek M, et al. SSBC 2020: sclera segmentation benchmarking competition in the mobile environment. In: 2020 IEEE International Joint Conference on Biometrics (IJCB),. IEEE; Sep. 2020. p. 1–10. https://doi.org/10.1109/IJCB48548.2020.9304881.
[15] Dhalla S, et al. Semantic segmentation of palpebral conjunctiva using predefined deep neural architectures for anemia detection. Procedia Comput Sci 2023;218:328–37. https://doi.org/10.1016/j.procs.2023.01.015.
[16] World Health Organization. Anemia Treatment, prevalence and data status. Accessed: Jul. 16, 2022. [Online]. Available: https://www.who.int/health-topics/anaemia#tab=tab_3; Apr. 2019.
[17] Zhang A, et al. Prediction of anemia using facial images and deep learning technology in the emergency department. Front Public Health Nov. 2022;10. https://doi.org/10.3389/fpubh.2022.964385.
[18] Chen Y, Zhong K, Zhu Y, Sun Q. Two-stage hemoglobin prediction based on prior causality. Front Public Health Nov. 2022;10. https://doi.org/10.3389/fpubh.2022.1079389.
[19] Sarsam SM, Al-Samarraie H, Alzahrani AI, Shibghatullah AS. A non-invasive machine learning mechanism for early disease recognition on Twitter: the case of anemia. Artif Intell Med Dec. 2022;134:102428. https://doi.org/10.1016/j.artmed.2022.102428.
[20] Shahzad M, et al. Identification of anemia and its severity level in a peripheral blood smear using 3-tier deep neural network. Appl Sci May 2022;12(10):5030. https://doi.org/10.3390/app12105030.
[21] Jain P, Bauskar S, Gyanchandani M. Neural network based non-invasive method to detect anemia from images of eye conjunctiva. Int J Imaging Syst Technol 2019;30(1):112–25. https://doi.org/10.1002/ima.22359.
[22] Magdalena R, Saidah S, Ubaidah IDS, Fuadah YN, Herman N, Ibrahim N. Convolutional neural network for ANEMIA detection based on conjunctiva palpebral images. Jurnal Teknik Informatika (Jutif) 2022;3(2):349–54.
[23] Delgado-Rivera G, et al. Method for the automatic segmentation of the palpebral conjunctiva using image processing. In: 2018 IEEE International Conference on automation/XXIII Congress of the Chilean association of automatic Control (ICA-ACCA). IEEE; 2018. p. 1–4.
[24] Bin Noor N, Anwar Md S, Dey M. Comparative study between decision tree, SVM and KNN to predict anaemic condition. In: 2019 IEEE International Conference on biomedical Engineering, computer and information technology for health (BECITHCON). IEEE, Nov.; 2019. p. 24–8. https://doi.org/10.1109/BECITHCON48839.2019.9063188.
[25] Agrawal A. Detecting anemia from retinal images using deep learning. Kolkata, India: Indian Statistical Institute; 2021.
[26] Peksi NJ, Yuwono B, Florestiyanto MY. Classification of anemia with digital images of nails and palms using the naive Bayes method. Telematika Mar. 2021;18(1):118. https://doi.org/10.31315/telematika.v18i1.4587.
[27] Bauskar S, Jain P, Gyanchandani M. A noninvasive computerized technique to detect anemia using images of eye conjunctiva. Pattern Recogn Image Anal 2019;29(3):438–46. https://doi.org/10.1134/s1054661819030027.
[28] Appiahene P, Asare JW, Donkoh ET, Dimauro G, Maglietta R. Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. BioData Min Jan. 2023;16(1):2. https://doi.org/10.1186/s13040-023-00319-z.
[29] Ghosh A, Mukherjee J, Chakravorty N. A low-cost test for anemia using an artificial neural network. Comput Methods Programs Biomed Feb. 2023;229:107251. https://doi.org/10.1016/j.cmpb.2022.107251.
[30] Dimauro G, Griseta ME, Camporeale MG, Clemente F, Guarini A, Maglietta R. An intelligent non-invasive system for automated diagnosis of anemia exploiting a

novel dataset. Artif Intell Med Feb. 2023;136:102477. https://doi.org/10.1016/j.artmed.2022.102477.
[31] Acar E, Turk O, Ertugrul F, Aldemir E. Employing deep learning architectures for image-based automatic cataract diagnosis. Turk J Electr Eng Comput Sci Oct. 2021; 29(SI-1):2649–62. https://doi.org/10.3906/elk-2103-77.
[32] Dimauro G, Camporeale MG, Dipalma A, Guarini A, Maglietta R. Anaemia detection based on sclera and blood vessel colour estimation. Biomed Signal Process Control Mar. 2023;81:104489. https://doi.org/10.1016/j.bspc.2022.104489.
[33] Hasan MK, Haque M, Sakib N, Love R, Ahamed SI. Smartphone-based human hemoglobin level measurement analyzing pixel intensity of a fingertip video on different color spaces. Smart Health 2018;5:26–39. https://doi.org/10.1016/j.smhl.2017.11.003.
[34] Wu J-H, Liu TYA, Hsu W-T, Ho JH-C, Lee C-C. Performance and limitation of machine learning algorithms for diabetic retinopathy screening: meta-analysis. J Med Internet Res Jul. 2021;23(7):e23863. https://doi.org/10.2196/23863.
[35] Rosa MJ, Seymour B. Decoding the matrix: benefits and limitations of applying machine learning algorithms to pain neuroimaging. Pain May 2014;155(5):864–7. https://doi.org/10.1016/j.pain.2014.02.013.
[36] Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. Br J Radiol Aug. 2019;92(1100):20190001. https://doi.org/10.1259/bjr.20190001.
[37] Eckart L, Eckart S, Enke M. A brief comparative study of the potentialities and limitations of machine-learning algorithms and statistical techniques. E3S Web of Conferences Jun. 2021;266:02001. https://doi.org/10.1051/e3sconf/202126602001.
[38] Asare JW, Appiahene P, Donkoh ET, Dimauro G. Iron deficiency anemia detection using machine learning models: a comparative study of fingernails, palm and conjunctiva of the eye images. Engineering Reports May 2023. https://doi.org/10.1002/eng2.12667.
[39] Appiahene P, Arthur EJ, Korankye S, Afrifa S, Asare JW, Donkoh ET. Detection of anemia using conjunctiva images: a smartphone application approach. Med Nov Technol Devices Jun. 2023;18:100237. https://doi.org/10.1016/j.medntd.2023.100237.
[40] Appiahene P, Chaturvedi K, Asare JW, Donkoh ET, Prasad M. CP-AnemiC: a conjunctival pallor dataset and benchmark for anemia detection in children. Med Nov Technol Devices Jun. 2023;18:100244. https://doi.org/10.1016/j.medntd.2023.100244.
[41] Karagül Yıldız T, Yurtay N, Öneç B. Classifying anemia types using artificial learning methods. Engineering Science and Technology, an International Journal 2021;24(1):50–70. https://doi.org/10.1016/j.jestch.2020.12.003.
[42] K.T N, Prasad K, Singh BMK. Analysis of red blood cells from peripheral blood smear images for anemia detection: a methodological review. Med Biol Eng Comput Sep. 2022;60(9):2445–62. https://doi.org/10.1007/s11517-022-02614-z.
[43] Appiahene P, et al. Application of ensemble models approach in anemia detection using images of the palpable palm. Med Nov Technol Devices Dec. 2023;20: 100269. https://doi.org/10.1016/j.medntd.2023.100269.
[44] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data Dec. 2019;6(1):60. https://doi.org/10.1186/s40537-019-0197-0.
[45] Asare JW, Appiahene P, Donkoh ET. Detection of anaemia using medical images: a comparative study of machine learning algorithms – a systematic literature review. Inform Med Unlocked 2023;40:101283. https://doi.org/10.1016/j.imu.2023.101283.
[46] Gbenga Emmanuel Dada, Opeoluwa David Oyewola, Bassi Stephen Joseph. Deep convolutional neural network model for detection of sickle cell anemia in peripheral blood images. Communication in Physical Sciences Mar. 2022;8(1): 9–22.
[47] Alomar K, Aysel HI, Cai X. Data augmentation in classification and segmentation: a survey and new strategies. J Imaging Feb. 2023;9(2):46. https://doi.org/10.3390/jimaging9020046.
[48] Dalvi PT, Vernekar N. Anemia detection using ensemble learning techniques and statistical models. In: 2016 IEEE International Conference on recent trends in Electronics, information and Communication technology, RTEICT 2016 - Proceedings. Institute of Electrical and Electronics Engineers Inc.; Jan. 2017. p. 1747–51. https://doi.org/10.1109/RTEICT.2016.7808133.
[49] J. Beetsma, "The CIELAB L*a*b* system – the method to quantify Colors of Coatings.".
[50] Joker PT. Binary image classifier CNN using TensorFlow. Accessed: Jun. 26, 2022. [Online]. Available: https://medium.com/techiepedia/binaryimage-classifier-cnn-using-tensorflowa3f5d6746697.
[51] Putra AT, Usman K, Saidah S. Webinar student presence system based on regional convolutional neural network using face recognition. Jurnal Teknik Informatika (Jutif) 2021;2(2):109–18.
[52] Anitescu C, Atroshchenko E, Alajlan N, Rabczuk T. Artificial neural network methods for the solution of second order boundary value problems. Comput Mater Continua (CMC) 2019;59(1):345–59. https://doi.org/10.32604/cmc.2019.06641.
[53] Samaniego E, et al. An energy approach to the solution of partial differential equations in computational mechanics via machine learning: concepts, implementation and applications. Comput Methods Appl Mech Eng Apr. 2020;362: 112790. https://doi.org/10.1016/j.cma.2019.112790.
[54] Acar E. Extraction of texture features from local iris areas by glcm and iris recognition system based ON KNN. 2016.
[55] Djuric N, Grbovic M, Vucetic S. Distributed confidence-weighted classification on Big data platforms. 2015. p. 145–68. https://doi.org/10.1016/B978-0-444-63492-4.00007-1.
[56] Naik A, Samant L. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Comput Sci 2016; 85:662–8.
[57] Peker M, Özkaraca O, Şaşar A. Use of orange data mining Toolbox for data analysis in clinical decision making. 2018. p. 143–67. https://doi.org/10.4018/978-1-5225-5149-2.ch007.
[58] Irum A, Akram M, Ayub SM, Waseem S, Khan MJ. Anemia detection using image processing. In: The International Conference on digital information processing. Electronics, And Wireless Communications; 2016.