Detection of anemia using ensemble model

Dr. D N Varshitha, Prof. D U Latha, K S Chirag, K Madan Kumar, M Manisha, K Sriharsha

*Vidyavardhaka college of engineering, Mysuru. Karnataka, India.*

ABSTRACT: In this paper we are proposing a new ensemble method for anemia detection using machine learning and a combination of data sources. By analysing both medical information that are symptoms, history, vitals and smartphone images of the eye conjunctiva, the approach aims to surpass traditional blood tests. The analysis involves extracting informative visual cues from the images using Convolutional Neural Networks (CNNs). Importantly, data from different sources is weighted based on significance for anemia detection before being combined. This method, along with other machine learning algorithms, is expected to improve detection accuracy and enable earlier diagnosis. Additionally, the accessibility of this non-invasive approach has the potential to empower individuals and transform preventative healthcare. However, further research is needed to validate this approach and ensure its effectiveness in real- world settings.

*Keywords:* Anemia, Ensemble learning, conventional Neural Network.

1. INTRODUCTION

Anemia, marked by a deficiency in red blood cells or haemoglobin, significantly impacts global health. The World Health Organization estimates that 42% of children under six and 40% of pregnant women suffer from this condition, often due to iron deficiency. Untreated anemia can lead to severe consequences, including irreversible organ damage and even death. Early detection is vital, yet current diagnostic methods, primarily blood tests, have notable limitations such as invasiveness and resource constraints, particularly in low-resource settings. This paper proposes an innovative solution: a machine learning model leveraging smartphone technology for non-invasive anemia detection. By integrating traditional data—symptoms, patient history, and vitals—with smartphone-captured images of the conjunctiva (the inner eyelids), deep learning algorithms can analyse these images to detect subtle visual indicators of anemia. This multifaceted risk assessment model promises enhanced accuracy and accessibility.

However, to ensure the model’s effectiveness across diverse populations, rigorous testing is essential. Variations in skin tone and lighting conditions can impact the accuracy of visual cues, necessitating robust datasets that include children, pregnant women, and individuals with varying skin tones. This inclusive approach is crucial for equitable access to this potentially transformative technology. Moreover, addressing regulatory and privacy concerns and providing adequate training for healthcare staff are necessary for successful integration into healthcare systems. This vision of a non-invasive, AI-powered model for anemia detection holds promise for early intervention and better health outcomes, making anemia diagnosis more accessible and less invasive worldwide

1. RELATED WORK

The paper [1] outlines a skincare product recommendation system using facial analysis from photos, leveraging CNNs and traditional image techniques to assess skin features like color, texture, and acne. It highlights the need for improved accuracy, fairness, and privacy protection to address current flaws such as errors and biases.

The paper [2] discusses an AI-powered app for estimating iron levels from nail photos, using color and texture analysis linked to a database. Despite its potential for non-invasive anemia testing, it faces challenges with image quality, lighting, skin color variability, and privacy concerns, necessitating improvements in accuracy, fairness, and data security.

The paper [3] explores using the VGG16 deep learning model to detect anemia from smartphone photos of the eye's white part by analysing color and patterns against a labelled database. While offering a convenient anemia screening method, challenges include image quality, lighting, variability across skin tones, and privacy concerns, necessitating improvements in accuracy, bias reduction, and data security.

The paper [4] discusses the limitations of traditional anemia tests and highlights the advantages of using computer programs for anemia detection, such as being non-invasive, cost-effective, and time-saving. It reviews various methods, including blood extraction, eye whiteness, nail color, palm examination, and smartphone devices, noting factors like image size, data preparation, and program accuracy, while addressing issues such as inter-observer variability, sensitivity of eye whiteness, insufficient anemia data, and current information challenges.

The paper [5] evaluates machine learning methods to predict anemia in children, finding Random Forest achieves the highest accuracy at 98.4%. Despite testing feature selection and ensemble learning, Random Forest remained superior. The study measured various performance metrics but noted limitations such as data from a single hospital and a small feature set.

1. METHODOLOGIES

The methodology that we have used in the project is a collection of multiple symptoms related to anemia and predict probability of anemia individually on each of them and then combine them into a result with different importance to each element. This will make sure that we have taken all the different symptoms into consideration according to their importance.

* 1. *Medical Information*

Patient data including symptoms (fatigue, shortness of breath), family history, and vitals. Statistical or machine learning techniques like feature selection or dimensionality reduction extract relevant patterns from the medical data. This might involve identifying correlations between specific symptoms and anemia.

* 1. *Visual Data*

Smartphone-captured images of the conjunctiva (eye's inner lining) - potentially including the palm in future iterations. Both medical information and visual data undergo pre-processing to ensure consistency. Convolutional Neural Networks (CNNs) analyse the conjunctival images, focusing on extracting features like subtle color variations that might indicate anemia. A similar CNN might analyse palm images in the future.

* 1. *Weighted Feature Fusion*

Weights are assigned to each data source (medical information, eye image, palm image) based on their relative importance in anemia detection. These weights can be determined by medical expertise or the model itself through feature importance analysis. The extracted features from each source are then multiplied by their corresponding weights before being combined. This ensures information deemed more crucial by the model contributes more to the overall analysis.

* 1. *Machine Learning Model*

DNNs (Deep Neural Networks) are a type of ML algorithm with multiple layers that progressively extract higher- level features from the data. In our case, DNNs can analyse conjunctival images to identify subtle patterns indicative of anemia. CNNs are a specific type of DNN particularly adept at processing image data. Their ability to automatically learn features from raw images extracting relevant information A Convolutional Neural Network (CNN) is a specialized deep learning model used for analysing visual data like images. It includes layers of neurons that perform operations like convolution, pooling, and activation. CNNs are great for tasks such as image classification, object detection, and facial recognition because they can learn hierarchical features from pixel values. They have transformed computer vision and are used across different industries like healthcare, autonomous vehicles, and security systems. Random Forest is a method in machine learning that creates multiple decision trees and then combines their predictions to make a final decision. It is widely used because it can improve accuracy and generalization in many different types of tasks. Random Forest is known for being robust, versatile, and resistant to overfitting, which is why it is a popular choice for both classification and regression tasks

* 1. *Proposed Architecture*

In this machine learning model for anemia detection, multiple data sources are merged using a weighted fusion approach. Medical data, including symptoms and vital signs, and smartphone-captured images of the conjunctiva and possibly the palm, undergo pre-processing. CNNs extract features from these images, focusing on variations indicative of anemia. Statistical techniques analyse the medical data. Weights are assigned to each data source based on their relevance, either determined by medical expertise or the model. The weighted features are combined, and an SVM assesses this data to predict anemia risk. This method leverages diverse data sources, prioritizing the most relevant information for accurate detection.

* 1. *Eye Model Training*

The initial step involves preparing the dataset for model training. This includes defining the paths to directories containing anemic and nonanemic images. Using the os.listdir() function, images and their corresponding labels are loaded into lists. These image paths and labels are then combined into a pandas DataFrame for convenient handling. Subsequently, the DataFrame is split into training and testing sets using the train\_test\_split() function from scikit- learn. Feature extraction is a critical preprocessing step where relevant information is extracted from the raw image data. In this project, the extract\_features() function is defined to perform this task. Within this function, images are resized to a standard size of 224x224 pixels using the resize() function from the Python Imaging Library (PIL). Additionally, pixel values are normalized by dividing by 255.0 to ensure uniformity across the dataset. The extracted features are reshaped into the required format for input into the CNN model. The core of the anemia detection system lies in the design of the CNN architecture. Utilizing the Sequential API of Keras, a CNN model is constructed layer by layer. The architecture typically comprises convolutional layers for feature extraction, followed by max-pooling layers to reduce spatial dimensions. Additional convolutional layers with increasing filter sizes capture higher-level features. Fully connected layers with Rectified Linear Unit (ReLU) activation function are incorporated for classification, with dropout regularization to prevent overfitting. The output layer consists of a single neuron with sigmoid activation for binary classification.

Once the model architecture is defined, it is compiled and trained using the specified optimizer, loss function, and training parameters. In this project, the Adam optimizer and binary cross-entropy loss function are chosen. Training is performed using the fit() method, specifying parameters such as batch size, number of epochs, and validation split. During training, the model learns to minimize the loss function by adjusting its weights and biases to make accurate predictions. Training and validation accuracy and loss metrics are monitored and stored for later analysis.

* 1. *Palm model Training*

Extract\_features() function is defined to extract both color and texture features from pre-processed palm images. Initially, the images are converted to the HSV color space to capture color information effectively. For each channel in the HSV image, statistical features such as mean, standard deviation, and percentiles are computed and added to the feature list. Additionally, texture features are extracted using the Local Binary Pattern (LBP) method applied to the grayscale version of the image. The script loads palm images from directories containing anemic and healthy samples. For each image, features are extracted using the defined function, and corresponding labels (1 for anemic, 0 for healthy) are assigned. The extracted features and labels are stored in separate lists, forming the feature matrix (X) and target vector (y) for model training. The feature matrix (X) and target vector (y) are split into training and testing sets using the train\_test\_split() function from scikit-learn. A RandomForestClassifier model with 100 decision trees is initialized and trained on the training set using the fit() method. The trained model is then ready for evaluation and deployment.

* 1. *Fingernail model Training*

The script begins by defining a function load\_data() to load and pre-process the images from the specified folder path. Images are resized to 224x224 pixels using the target\_size parameter of image.load\_img(). For each image, its corresponding label (1 for Anemic and 0 for Non- Anemic) is assigned, and both images and labels are stored in numpy arrays X and y, respectively. The pre-trained VGG16 model is loaded with weights from the ImageNet dataset, excluding the fully connected layers (include\_top=False). Feature extraction is performed by passing the input images through the base VGG16 model and flattening the output. The resulting features are stored in the features variable using the predict() method.

The trained model is evaluated on the testing dataset, with its performance assessed using metrics like accuracy, precision, recall, and F1 score. Predictions are made on pre-processed test images, and accuracy and loss curves are plotted using matplotlib. This evaluation reveals the model's generalization capability and effectiveness in detecting anemia.

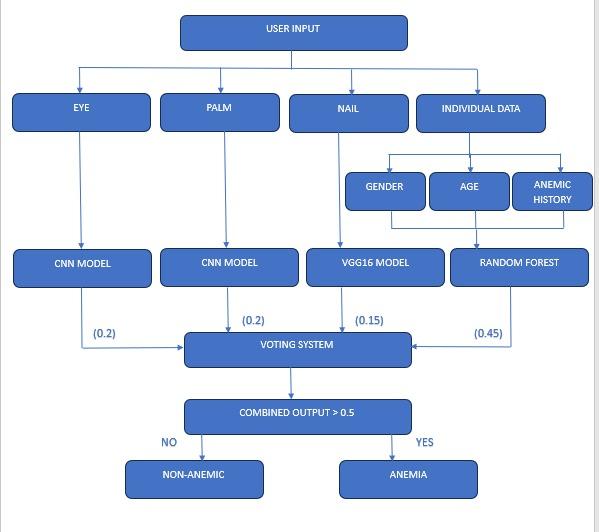


Fig.1 System Architecture

1. RESULTS:

The final output of the model is a prediction of the patient's anemia risk.

* Binary Classification: Anemic or not anemic.
* Probability Score: Likelihood of having anemia (ranging from 0 to 1).
* Improved Accuracy and Generalizability: Data fusion with weighted features leverages strengths of various sources and prioritizes the most relevant information, potentially leading to more accurate anemia detection compared to relying on single data sources.
* Early Detection: By providing a non-invasive and potentially accessible approach, early detection of anemia becomes possible, enabling timely interventions and improved health outcomes.
* Democratization of Healthcare: Empowering individuals and communities with the ability to assess their anemia risk using smartphones puts preventative healthcare within reach.

We have observed an accuracy of 71.11 for the eye training model. We have observed an accuracy of 90.7% for the finger training model. We have observed an accuracy of 99.6% for the palm training model. The accuracy of the final model is 86.33%. This is calculated using the average of the other models.

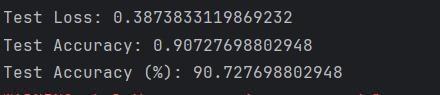


Fig. 2 Accuracy of Eye Model Fig.3 Accuracy of Finger Nails model

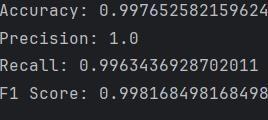


Fig.4 Accuracy of Palm model

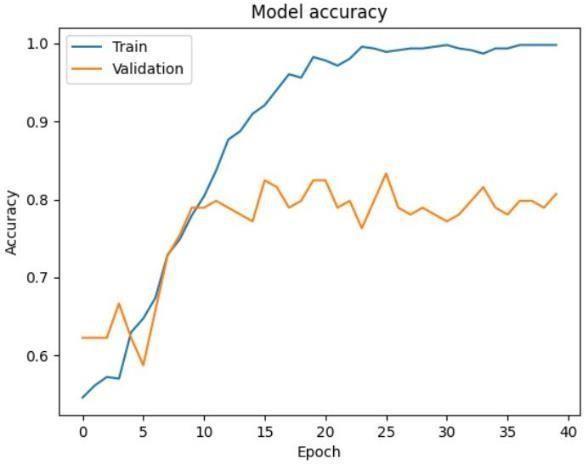
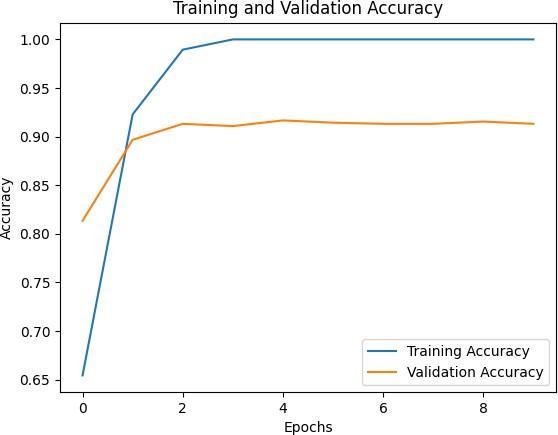
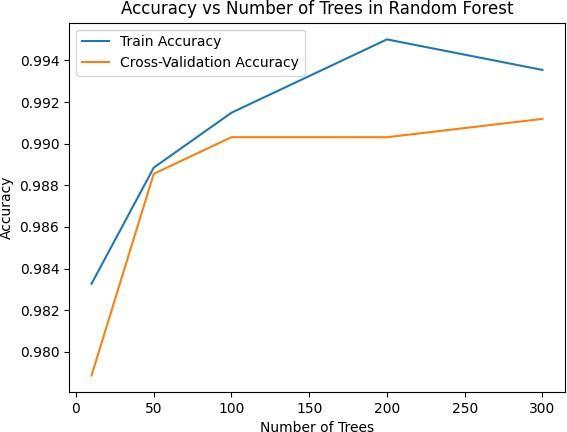


Fig.5 Processing of eye image Fig.6 Processing of palm image Fig.7 Processing of nail image

4.1. *UI/UX of the website*

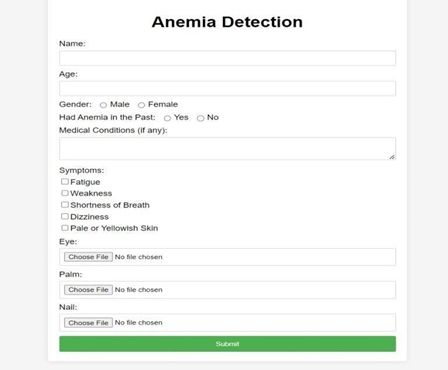
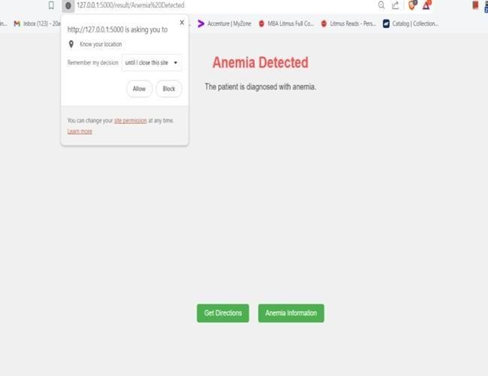
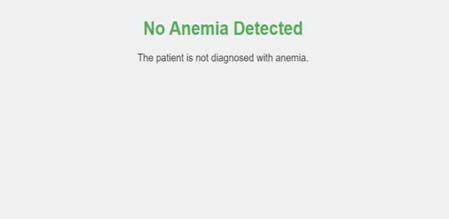


Fig.8 Details of the user Fig.9 Results of the user



Fig.10 Non-anemic Images Fig.10 Anemic Images

5 CONCLUSIONS

The global burden of anemia necessitates innovative solutions, and the potential of converging technologies and data sources to revolutionize its diagnosis and management. Evidence shows the pervasiveness of anemia, particularly among vulnerable populations like children and pregnant women, emphasizing the urgent need for accessible and accurate detection methods. While traditional blood tests have been fundamental, their limitations in accessibility and resource demands call for alternative approaches. Machine learning (ML) emerges as a promising solution, offering a non- invasive, AI-powered pathway to early detection. Our survey highlights various ML techniques, notably Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs), showcasing their ability to analyse conjunctival images and extract subtle visual cues indicative of anemia. These findings not only demonstrate the accuracy and generalizability of prediction but also signify the potential for democratizing healthcare. Practical Implications include the potential for ML-driven approaches to improve accessibility and accuracy of anemia diagnosis, particularly in underserved populations. Leveraging smartphones and readily available data, this non-invasive approach holds promise in reaching regions with limited healthcare infrastructure, empowering individuals and communities with early detection capabilities. Future research should focus on optimizing ML algorithms for diverse demographic groups, integrating ML-based diagnostic tools into existing healthcare systems, and assessing the long-term impact of ML-driven early detection on anemia management and public health outcomes. By addressing these areas, we can further advance proactive health management and build resilience against anemia.

6REFERENCES

1. Yang, N., Li, X., Wu, X., Lin, F., Wu, X., Zhu, W., et al. (2022). A deep learning approach for facial image-based skincare product recommendation system. Frontiers in Public Health, 10, 964385. https://doi.org/10.3389/fpubh.2022.964385
2. Tilburg University, Digital Sciences for Society Program. (n.d.). NAIL: Nail based Artificial Intelligence image Learning for measuring iron status. Project website.
3. Dimauro, G., Guarini, A., Caivano, D., Girardi, F., Pasciolla, C., & Iacobazzi, A. (2019). Detecting clinical signs of anaemia from digital images of the palpebral conjunctiva. IEEE Access, 7, 113488-113498. https://doi.org/10.1109/ACCESS.2019.2932274
4. Appiahene, P., Asare, J. W., Donkoh, E. T., Dimauro, G., & Maglietta, R. (2023). Detection of iron deficiency anemia by medical images: A comparative study of machine learning algorithms. BioData Mining, 16(1), 2. https://doi.org/10.1186/s13040-023-00319-z
5. Dhakal, P., Khanal, S., & Bista, R. (2023). Prediction of anemia using machine learning algorithms. International Journal of Computer Science and Information Technology, 15.
6. Dimauro, G., Caivano, D., Di Pilato, P., Dipalma, A., & Camporeale, M. G. (2020). A systematic mapping study on research in anemia assessment with non-invasive devices. Applied Sciences, 10(14), 4804. <https://doi.org/10.3390/app10144804>
7. An, R., Huang, Y., Man, Y., Valentine, R. W., Kucukal, E., Goreke, U., Sekyonda, Z., Piccone, C., Owusu-Ansah, A., Ahuja, S., Little, J. A., & Gurkan, U. A. (2021). Emerging point-of-care technologies for anemia detection. Lab on a Chip, 21(10), 1843-1865. https://doi.org/10.1039/d0lc01235a
8. Shekhar, M., Berk, D. T., Mohammed, M., Mustafa, K., & Freeman, C. (2023). Anemia detection through non-invasive analysis of lip mucosa images. Frontiers in Big Data, 6. <https://doi.org/10.3389/fdata.2023.124>