Anemia Detection using Conjunctiva Images

Manas Satish Bedmutha
University of California San Diego
La Jolla, CA 92093
mbedmutha@ucsd.edu

Jainish Nileshkumar Chauhan
University of California San Diego
La Jolla, CA 92093
jchauhan@ucsd.edu

Siddhant Saoji
University of California San Diego
La Jolla, CA 92093
ssaoji@ucsd.edu

Abstract

Anemia is a condition where the blood has insufficient Hemoglobin (Hb) concentration to effectively carry oxygen to all parts of the body. It affects 1.62 billion people globally. Anemia can be detected via Hb level and other data of the patient such as gender, age etc. The only gold standard for testing anemia is through blood tests which are not practically possible at scale. Medically the conjunctiva of the eye acts as a good indicator of the Hb level. With increasing research interest in digital eye images, we aim to develop an automated system that can predict the Hb level as well as classify anemic patients using deep learning. We also aim to validate our findings by visualizing the network attention maps to explain the physical correctness of the model.

1 Background & Related Work

Our goal is to use images of the conjunctiva from the eyes to detect Hb levels and thus anemia. We primarily use the datasets from [1] to develop ML models for Hb prediction.

Image Features The primary source of literature in this space has been around image capture and preliminary image processing methods to enhance the way data has been captured. [9] also showed the effectiveness of the Lab color space in extracting the red region. However this method fails since the color tone varies drastically with image capture settings and participants.

In [3], the authors utilize the effect of hue saturation in HSI color space to filter out meaningful regions and extract intensities before passing through a Support Vector Machine. [4] uses mean values of Red and Green channels with an experimental threshold to classify anemic and non-anemic samples. [10] extended this further by introducing a multilayer perceptron to learn this threshold from data and fine tune the results.

Classical Learning Ideas Some of the work in this area has also been around identifying models that rely on image statistics and deploying classical machine learning systems around it. [5] uses a Kalman filtering setup to analyse image intensities. Another interesting idea was the use of K-means clustering to cluster image pixels to get anemia classification as shown in [2] and [7].

Deep Learning Approaches Most of the prior work has mainly relied on manually segmented regions to detect anemia using custom features. Many recent works have also tried to automate the segmentation pipeline itself. [8] and [9] use contours and image features to first segment the

ECE228. Milestone report.

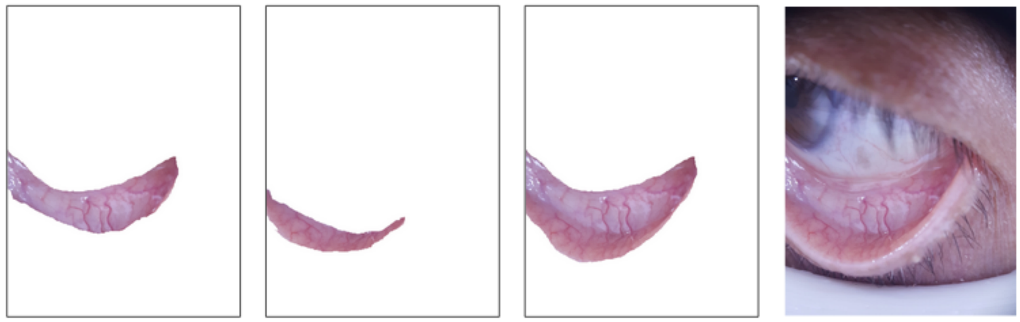

Figure 1: The figure shows a representative sample from the image. The images are the palpebral, forniceal, total conjunctiva and original image in the left to right order.

conjunctiva and then apply ML models. [6] uses a U-Net based semantic segmentation network to extract the conjunctiva before processing further.

2 Data Processing

The dataset is a mix of data collected by different research groups in Italy and India. Some datapoints are also missing or corrupted. We developed an ETL pipeline to allow us to access data across different structures at once. First all the contents of the folders are traversed. Once the annotations are discovered, all values are normalized to the same format.

Images with different names, viz. base_palpberal, base_forniceal, base_forniceal_palpebral, base (where base is the filename of the original image) in each folder are segregated and loaded into different lists. All lists of images are saved as numpy data files for further processing. Also, for consistency across all images we have down sampled all the images to 640*480*3 dimensions. Along with this we have also done some preprocessing of the labels to like bringing all of them to appropriate data type as well handling some of the Null values. Note: Some folders have annotated images missing.

3 Data Visualization and Analysis

3.1 Dataset Statistics

Eyes-defy-anemia dataset is used [1] for our task. This dataset contains 218 images of eyes, in particular conjunctivas, which can be used for research on the diagnosis and estimation of anemia based on the pallor of conjunctiva. Each image in the dataset is labeled for the Gender, Age and Hemoglobin (Hb) level of the person whose eye's image was taken. Below are the results of the analysis of the data that we performed. 2.

The average hemoglobin level in the dataset is 12.7975. The maximum and the minimum values of the Hb level across the dataset are 17.4 and 7 respectively. From above plots it is evident that there are more samples from males compared to females. Also, the average hemoglobin level for male is higher compared to females. This is the reason why the average Hb level across the dataset is closer to the maximum Hb level value. As we can see from the above plots there is no explicit relationship between age and Hb level but in general the average value of Hb level is higher in younger people compared to older people. Some of these relations are motivation for trying multitask learning approach in future.

Sample images of the eyes from the dataset can be seen in 1. The dataset also contains segmentation masks of conjunctiva and the exposed parts of the sclera and iris.

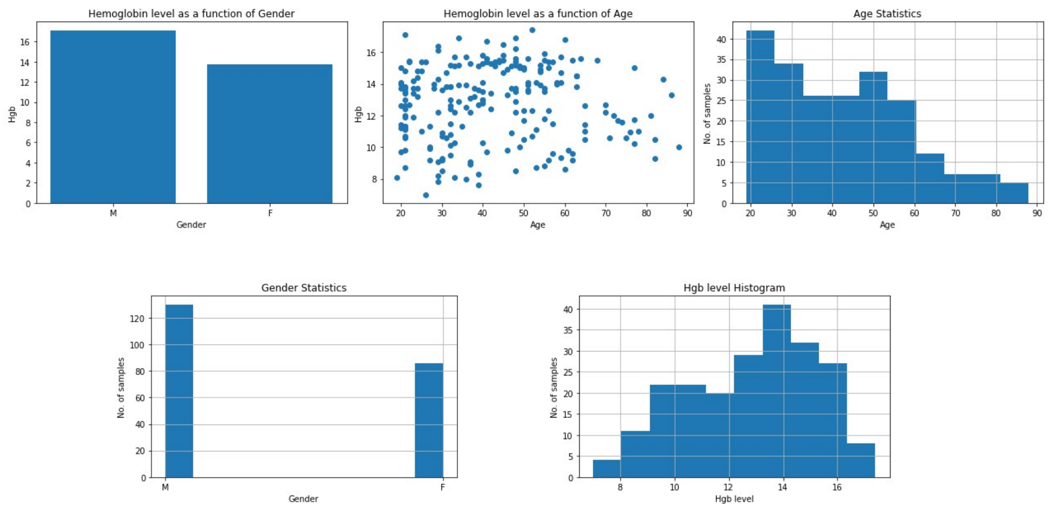

Figure 2: Dataset Visualisation (labels self-explanatory). Affirming to real world knowledge, we find average Hb levels are higher for men that women and higher for younger people. Additionally, the dataset follows a normal distribution and can be assumed to capture population dynamics.

3.2 Image Analysis

Raw images of the eye cannot directly be used for processing since they contain noise. We sought to find ways to preprocess the image such that the conjunctival region could be enhanced. We denoised the image using Gaussian Blur and then also performed Otsu's Thresholding 3.

Figure 3: Image preprocessing results: denoised image using Gaussian Blur with kernel size 5 and thresholded image to highlight conjunctival region

4 Problem Formulation

The given task can be solved via two different approaches. The first as a classification approach where we train the model using certain features extracted from images and the labels of the images for anemia / non-anemia. For getting anemia / non-anemia labels we have used simple thresholding on our data before training the model. As we have Hb level labels we label all the images with Hb level greater than some threshold as non-anemia and others as anemia.

The second approach is the Regression+Classification approach where we first train regression models using given images and their labels of Hb level. Then use this predicted Hb level to classify the given image into anemia / non-anemia. We can also use other features like age and gender of the person in this last classification step.

In both approaches we can either use whole eye images or just the segmented palpebral+forniceal part of the eye image.

5 Preliminary Results

In this section we have reported results of the sub tasks required for the two approached that we discussed in the above section.

First we have performed the task of predicting Hb level using different models for regression task. Table 1 summarizes the results of the same.

Next, we have performed the task of reproducing results of anemia / non-anemia classification using existing methods [3] and [10] on our data. Table 2 summarizes the results of the same.

Technique	MSE (3 Channels)	MSE (Red)	MSE (Green)	MSE (Blue)
Linear Regression	7.389	7.445	8.064	8.808
Vanilla CNN	9.448	12.324	14.987	13.107
XGBoost	12.986	15.982	16.535	16.974

Table 1: The Mean square Error of different approaches for hemoglobin prediction using raw image data before extensive preprocessing or segmentation. Linear regression was performed directly on flattened images, Vanilla CNN with 2 Convolution and 2 Fully Connected layers. XGBoost with 25 estimators and a max depth of 7

F1/Dataset	Raw Image	Segmented Image
HHR + PVM + SVM [3]	0.82	0.88
Channel Mean + MLP [10]	0.77	0.60

Table 2: The table shows the F1 scores for classification using segmented images using recent approaches. The first approach uses hue based features of the image passed through an SVM and the second shows channel mean based features regressed over a 2 layered multilayer perceptron

From the preliminary results we can infer that Linear regression gave the best result of the regression based approaches as compared to a vanilla CNN or XGBoost. This can be attributed to the fact that the dataset is small to properly train a CNN while avoiding over-fitting. The best results were found when all colour channels of the image were given. This proves the hypothesis that the red channel is not the only necessary channel to detect hemoglobin level. The classification based approaches also were not able to classify anemia with satisfactory accuracy. The possible problems for the unsatisfactory results are the small size of the dataset, the limited preprocessing of image data and using the RGB colour space instead of the LAB colour space used in some of the related work.

6 Future Goals and Plans

Based on our progress so far and the results attained, we want to learn from the mistakes in our approach and the feedback from our TA to improve the model to best suit our task. As discussed in results sections dataset's small size leads to undesirable results. To overcome this we have planned to use different approaches like Data augmentation, data oversampling using Techniques like SMOTE, transfer learning and semi-supervised learning. We will try to find another dataset related eyes and train a ML model on seperate task on that data and then use this pretrained model and fine tune it on our data for defined task. We will also try formulating the problem in different manner and using segmentation masks of the eye images. The table 3 shows our tentative plan.

Task	Expected Date
Data Augmentation + SMOTE	12 May 2022
Feature Extraction + ML	16 May 2022
Segmentation + ML	24 May 2022
Semi-supervised learning approach	27 May 2022
Transfer learning approach	29 May 2022
Fine tuning + Results compilation	1 June 2022

Table 3: The table shows our plan for completing the future parts of the work

References

[1] R. Maglietta, M. E. Griseta, F. Clemente, A. Guarini, G. Dimauro, Advances in the Automated Diagnosis of Anemia: a shared Eyes-defy-anemia dataset and a novel non-invasive system with machine learning, under submission

[2] Sevani, N., and G. B. V. Persulessy. "Detection anemia based on conjunctiva pallor level using k-means algorithm." IOP Conference Series: Materials Science and Engineering. Vol. 420. No. 1. IOP Publishing, 2018.

[3] Chen, Yi-Ming, Shaou-Gang Miaou, and Hongyu Bian. "Examining palpebral conjunctiva for anemia assessment with image processing methods." Computer methods and programs in biomedicine 137 (2016): 125-135.

[4] Tamir, Azwad, et al. "Detection of anemia from image of the anterior conjunctiva of the eye by image processing and thresholding." 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). IEEE, 2017.

[5] Chen, Yi-Ming, and Shaou-Gang Miaou. "A Kalman filtering and nonlinear penalty regression approach for noninvasive anemia detection with palpebral conjunctiva images." Journal of healthcare engineering 2017 (2017).

[6] Kasiviswanathan, Sivachandar, et al. "Semantic Segmentation of Conjunctiva Region for Non-Invasive Anemia Detection Applications." Electronics 9.8 (2020): 1309.

[7] Sevani, N., and G. B. V. Persulessy. "Detection anemia based on conjunctiva pallor level using k-means algorithm." IOP Conference Series: Materials Science and Engineering. Vol. 420. No. 1. IOP Publishing, 2018.

[8] Dimauro G, Simone L. Novel Biased Normalized Cuts Approach for the Automatic Segmentation of the Conjunctiva. Electronics. 2020; 9(6):997. https://doi.org/10.3390/electronics9060997

[9] Dimauro G, Caivano D, Di Pilato P, Dipalma A, Camporeale MG. A Systematic Mapping Study on Research in Anemia Assessment with Non-Invasive Devices. Applied Sciences. 2020; 10(14):4804. https://doi.org/10.3390/app10144804

[10] Jain, Prakhar, Shubham Bauskar, and Manasi Gyanchandani. "Neural network based non-invasive method to detect anemia from images of eye conjunctiva." International Journal of Imaging Systems and Technology 30.1 (2020): 112-125.