

CS6301

MACHINE LEARNING

PROJECT ON **Non-Invasive**

Anemia detection

*(From ACM April 2022 Paper: iNAP: A Hybrid Approach for
Non-Invasive Anemia-Polycythemia Detection in the IoMT)*

Abstract

The project presents a novel, self-sufficient model called iNAP to address the shortcomings of anemia detection. The proposed model captures fingernail images using a smartphone camera and automatically extracts the fingernails as the regions of interest. A novel algorithm extracts the dominant color by analyzing color spectroscopy of the extracted portions and accurately predicts blood hemoglobin level. A less than a certain threshold value is categorized as anemia while a greater than that is non-anemic. The model incorporates machine learning and image processing techniques allowing easy smartphone implementation. Furthermore, a novel brightness adjustment algorithm is developed, allowing robustness to a wide illumination range and the type of device used. The proposed IoMT framework allows virtual consultations between physicians and patients, as well as provides overall public health information. The model thereby establishes itself as an authentic and acceptable replacement for invasive and clinically based hemoglobin tests by leveraging the feature of self-anemia diagnosis.

Introduction

Hematology is the study of blood and its related diseases. In medical terms this field is inclusive of blood disorders such as anemia, sickle-cell anemia, hemophilia, leukemia, and lymphoma, among others. Hemoglobin is a protein in the Red Blood Cells (RBCs) tasked with delivering oxygen to different organs and tissues in the body and transporting carbon dioxide back to the lungs. A healthy level of hemoglobin is therefore primarily required for the proper functioning of the body. However, increased red blood cell destruction due to pancreatic diseases, severe blood loss due to a significant injury or trauma, or an iron-deficient diet for a prolonged period, causes the hemoglobin level to drop below an optimal level. Clinically this condition is termed “Anemia”. Therefore, we can define anemia as the dearth of blood hemoglobin (Hb).

To overcome the shortcomings of traditional invasive techniques of anemia detection, researchers shifted to a new category that does not involve any injection mechanism. They collectively constitute the “non-invasive” method of blood hemoglobin measurement. The significant advantage of these methods is their operational flexibility. With the advent of non-invasive methods, blood hemoglobin estimation became accessible through economical electronic gadgets, which implied that most of the population from low-middle income countries or rural areas could be effectively screened for anemia.

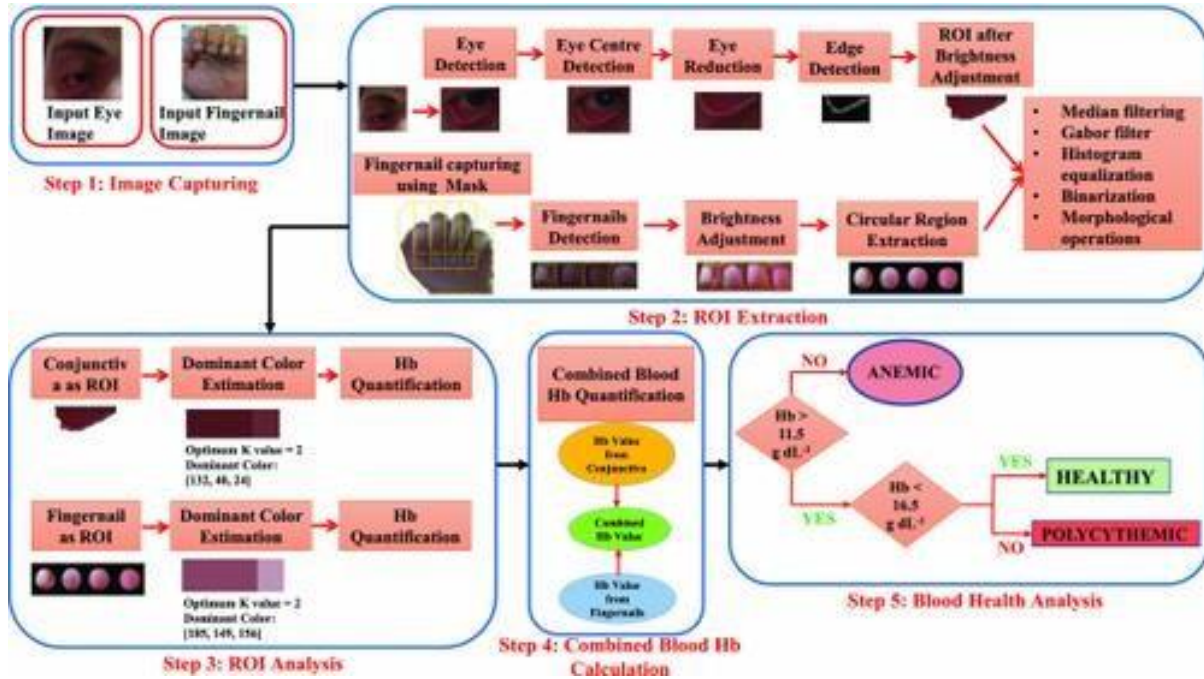
Problem Statement

Around the globe, 1.62 billion people are anemic, which corresponds to 25% of the entire world population, and it is the low and middle-income populations which make up the majority. Worldwide among the female, male, and children population, anemia is highly prevalent among women and children (47.4%) and relatively less prevalent in adult males (12.7%) [24]. Data from the World Health Organization (WHO) regional estimates indicate that the highest number of anemia affected individuals are in Southeast Asia (about 315 million individuals). In India, 21.7% of the male and 53.24% of the female population are clinically termed as anemic [9]. Anemia is predominant in pregnant women, malnourished children, and older adults (more than 60 years of age). Therefore, anemia constitutes a significant threat to public health, making its fast and efficient detection primarily important. Traditionally blood hemoglobin is measured by the syringe injection and blood extraction method. Techniques based on similar concepts collectively constitute the “invasive” method of blood hemoglobin measurement. However, they are excruciating, protracted, dicey, and involves the expertise of trained physicians. All these factors make the invasive method unobtainable and unfeasible for peripheral people living especially in low-middle income countries (LMC), resulting in unchecked blood hemoglobin levels among most of their population.

Overall Objective:

This project is aimed towards contributing to the non-invasive blood hemoglobin measurement domain. We have proposed a model that automatically extracts multiple regions of interest for a highly accurate prediction. The fingernail images are captured as inputs. Then, employing image processing, a prediction of the patient's blood hemoglobin level is made. The prediction is numerically quantified, similar to the Gold Standard method of hemoglobin measurement. Based on the standard thresholds mentioned, participants have been grouped into anemic, and healthy categories. The data set used for the study is appropriately balanced with equal proportions of all three categories and a wide age range. Finally, using the obtained predictions and the IoMT platform, we will implement the feature of virtual consultations and provide accurate statistical information regarding the surveyed area's overall public health that can be utilized by the health administrators.

System Model (Block Diagram):



Code:

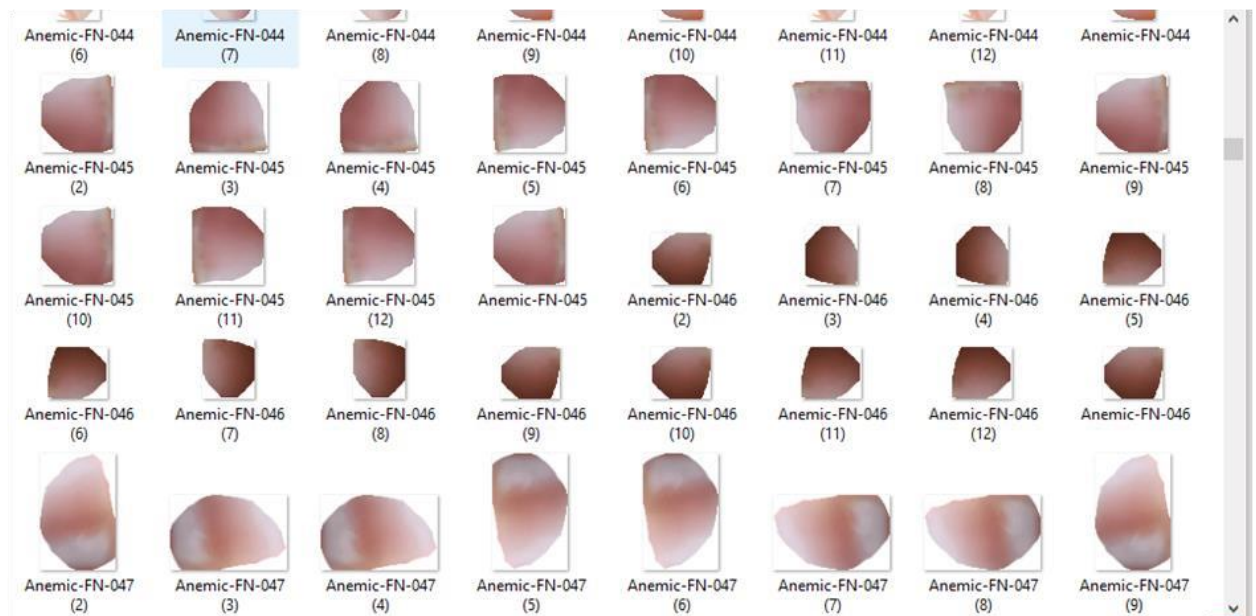
We have used Python as the programming language to code for our project.

Data pre-processing

- Getting the dataset

Source link for Fingernails dataset:






<https://data.mendeley.com/datasets/2xx4j3kjg2>



- Importing the Libraries

```
import tensorflow
import keras
import os
from skimage import io
import numpy as np
import pandas as pd
import csv
from google.colab.patches import cv2_imshow
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import math
```

- Importing the datasets

	Fingernails	4261 of 4261	
	Fingernails	Upload canceled	
	Detection of anaemia using conjunctival i...		

All files are first uploaded to google drive

The drive is then mounted in google colab, where it can then be accessed in the code.

All the images in the fingernail dataset can be accessed as well, as shown from this example


```
myImg = io.imread("drive/MyDrive/Fingernails/Fingernails/Anemic-FN-056.png")
cv2_imshow(myImg)
```



```
myImg = cv2.imread("drive/MyDrive/Fingernails/Fingernails/Non-Anrmic-FN-195.png")
#print(myImg)
cv2_imshow(myImg)
print(myImg)
cv2_imshow(cv2.cvtColor(myImg, cv2.COLOR_BGR2RGB))
```



Handling the data

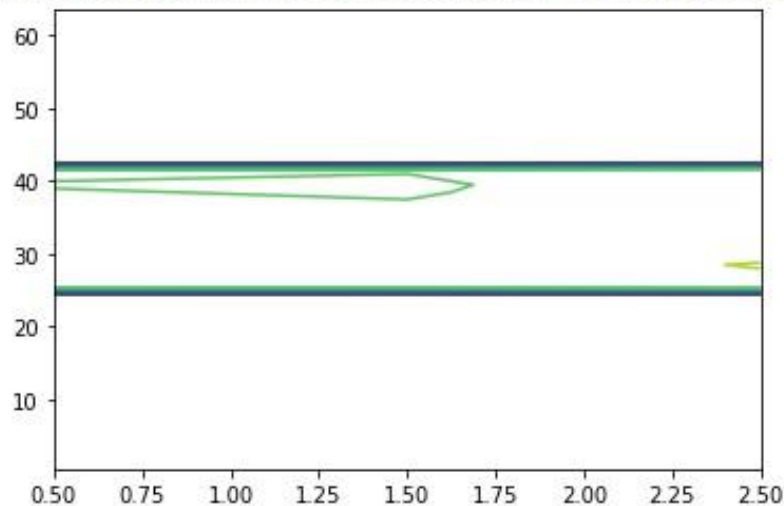
```
def load_data(data_dir):
    images=[]
    dc=[]
    anemic=[]
    i=j=0
    for pics in os.listdir(data_dir):
        if pics[0:3]!='Non':
            #if i==100:
            #    continue
            anemic.append(1)
            #i+=1
        else:
            #if j==100:
            #    continue
            anemic.append(0)
            #j+=1
        print(i,j)
        img=cv2.imread(os.path.join(data_dir,pics),1)
        img=AdaptBrightness(img)
        d = DominantColors(img)
        clrs=d.dominantColors()
        dc.append(d.MAXFREQ)
        img=cv2.resize(img, (200,200), interpolation = cv2.INTER_AREA)
        images.append(np.array(img))

    #if i==100 and j==100:
    #    break
    return images,dc,anemic
```

```
imgs,dc,anemic=load_data("drive/MyDrive/Fingernails/Fingernails")
```

```
✓ [18] plt.contour(cv2.cvtColor(myImg, cv2.COLOR_BGR2RGB)[0], origin = "image")  
0s
```

<matplotlib.contour.QuadContourSet at 0x7f4bd75dee50>



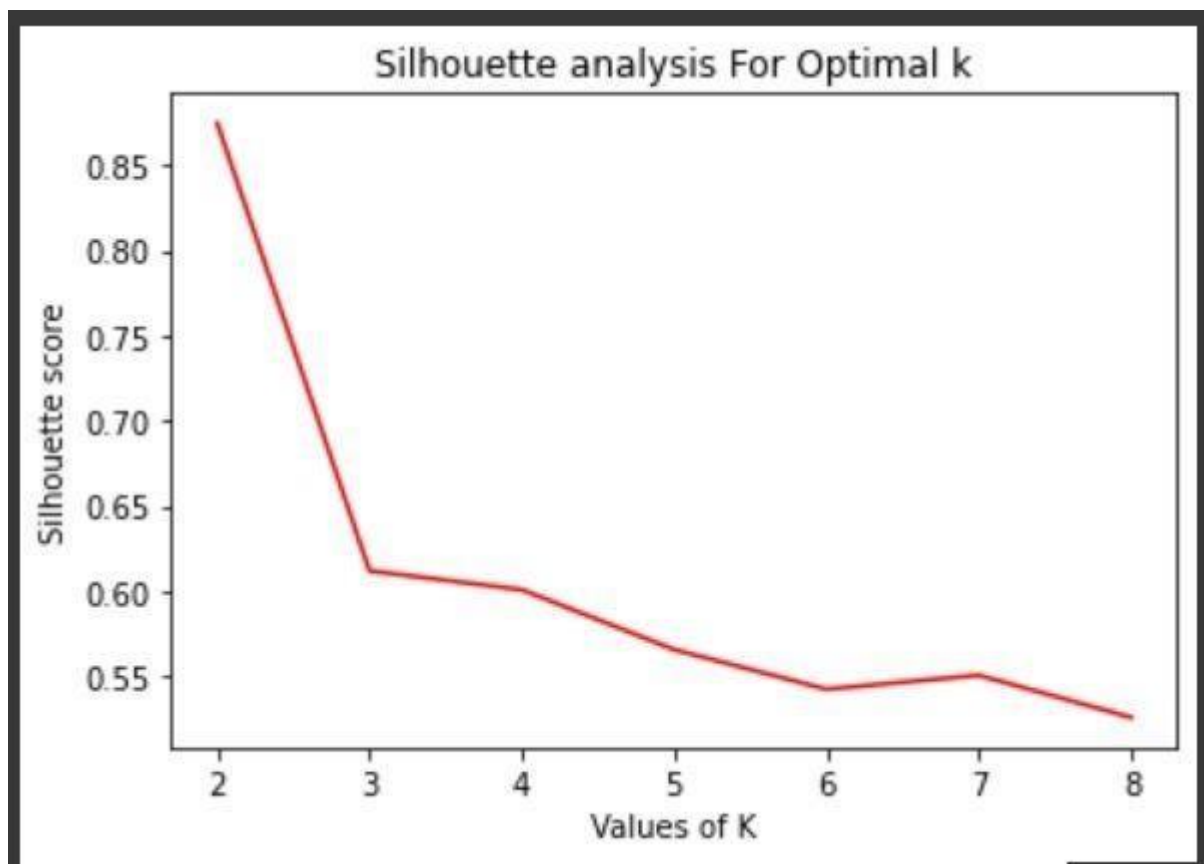
Splitting the data

```
from sklearn.neural_network import MLPClassifier  
from sklearn.datasets import make_classification  
from sklearn.model_selection import train_test_split  
import pandas as pd  
import numpy  
dc1=numpy.array(dc).T  
df1 = pd.DataFrame(dc, columns = ['R','G','B'])  
X_train, X_test, y_train, y_test = train_test_split(df1, anemic, train_size=0.6)
```

Silhouette Analysis:

Silhouette analysis can be used to study the separation distance between the resulting clusters. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters and thus provides a way to assess parameters like number of clusters visually.

Prior to using K Means Clustering, silhouette score is used to find the optimal k value for each image.



K means Clustering:

```
[ ] import cv2
    from sklearn.cluster import KMeans
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
    import sklearn
    from sklearn.cluster import KMeans
    from sklearn.metrics import silhouette_score
    import matplotlib.pyplot as plt
    from mpl_toolkits.mplot3d import Axes3D

class DominantColors:

    CLUSTERS = None
    IMAGE = None
    COLORS = None
    LABELS = None
    MAXFREQ = None

    def __init__(self, image, clusters=3):
        self.CLUSTERS = clusters
        self.IMAGE = image

    def dominantColors(self):
        #img = cv2.imread(self.IMAGE)
        img = cv2.cvtColor(self.IMAGE, cv2.COLOR_BGR2RGB)
        img = img.reshape((img.shape[0] * img.shape[1], 3))
        self.IMAGE = img
        range_n_clusters = [2, 3, 4, 5, 6, 7, 8]
        silhouette_avg = []
        for num_clusters in range_n_clusters:

            # initialise kmeans
            kmeans = KMeans(n_clusters=num_clusters)
```

```

[ ]      kmeans.fit(img)
         cluster_labels = kmeans.labels_

         # silhouette score
         silhouette_avg.append(silhouette_score(img, cluster_labels))
k=range_n_clusters[silhouette_avg.index(max(silhouette_avg))]
self.CLUSTERS=k
kmeans = KMeans(n_clusters = k)
kmeans.fit(img)
self.COLORS = kmeans.cluster_centers_
self.LABELS = kmeans.labels_
self.calcmaxfreq()
return self.COLORS.astype(int)
def calcmaxfreq(self):
    #labels form 0 to no. of clusters
    numLabels = np.arange(0, self.CLUSTERS+1)

    #create frequency count tables
    (hist, _) = np.histogram(self.LABELS, bins = numLabels)
    hist = hist.astype("float")
    hist /= hist.sum()

    #appending frequencies to cluster centers
    colors = self.COLORS
    #descending order sorting as per frequency count
    colors = colors[(-hist).argsort()]
    hist = hist[(-hist).argsort()]
    if colors[0].astype('int')[0] == 0 and colors[0].astype('int')[1] == 0 and colors[0].astype('int')[2] == 0:
        self.MAXFREQ=colors[1]
    else:
        self.MAXFREQ=colors[0]
def plotHistogram(self):

    #labels form 0 to no. of clusters
    numLabels = np.arange(0, self.CLUSTERS+1)

```

```
[ ]    #create frequency count tables
      (hist, _) = np.histogram(self.LABELS, bins = numLabels)
      hist = hist.astype("float")
      hist /= hist.sum()

      #appending frequencies to cluster centers
      colors = self.COLORS
      #descending order sorting as per frequency count
      colors = colors[(-hist).argsort()]
      hist = hist[(-hist).argsort()]
      #creating empty chart
      chart = np.zeros((50, 500, 3), np.uint8)
      start = 0
      if colors[0].astype('int')[0] == 0 and colors[0].astype('int')[1] == 0 and colors[0].astype('int')[2] == 0:
          self.MAXFREQ=colors[1]
      else:
          self.MAXFREQ=colors[0]
      #creating color rectangles
      for i in range(self.CLUSTERS):
          end = start + hist[i] * 500

          #getting rgb values
          r = colors[i][0]
          g = colors[i][1]
          b = colors[i][2]

          #using cv2.rectangle to plot colors
          cv2.rectangle(chart, (int(start), 0), (int(end), 50), (r,g,b), -1)
          start = end

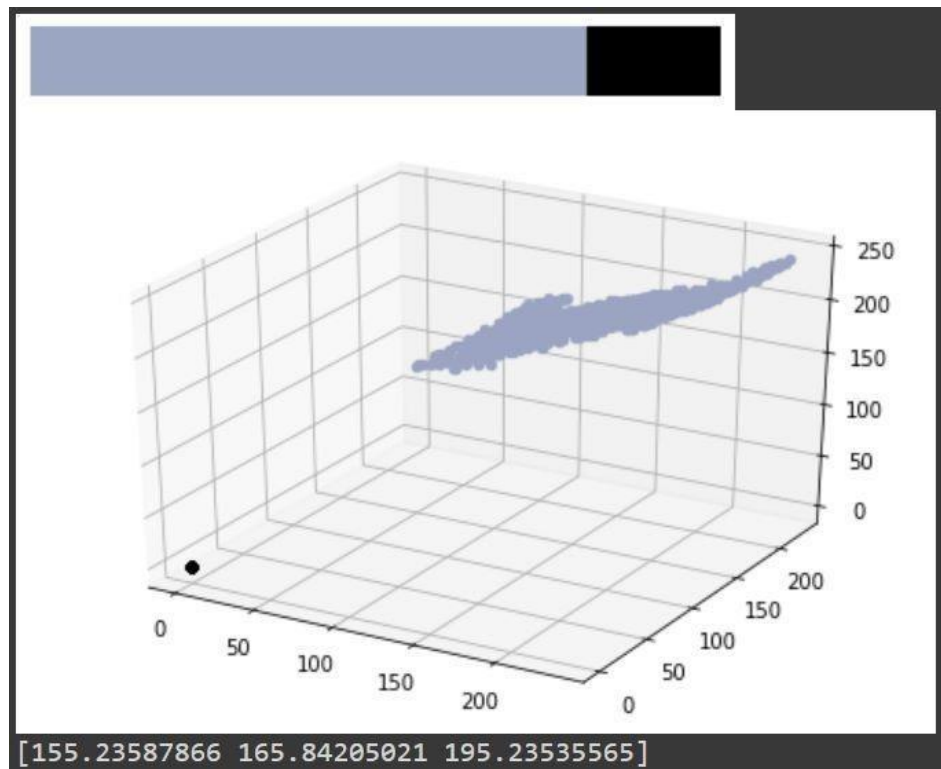
      #display chart
      plt.figure()
      plt.axis("off")
      plt.imshow(chart)
      plt.show()
```

```
def rgb_to_hex(self, rgb):
    return '#%02x%02x%02x' % (int(rgb[0]), int(rgb[1]), int(rgb[2]))

def plotClusters(self):
    #plotting
    fig = plt.figure()
    ax = Axes3D(fig)
    for label, pix in zip(self.LABELS, self.IMAGE):
        ax.scatter(pix[0], pix[1], pix[2], color = self.rgb_to_hex(self.COLORS[label]))
    plt.show()
```

```
dc = DominantColors(myImg)
colors = dc.dominantColors()
dc.plotHistogram()
dc.plotClusters()
print(dc.MAXFREQ)
```

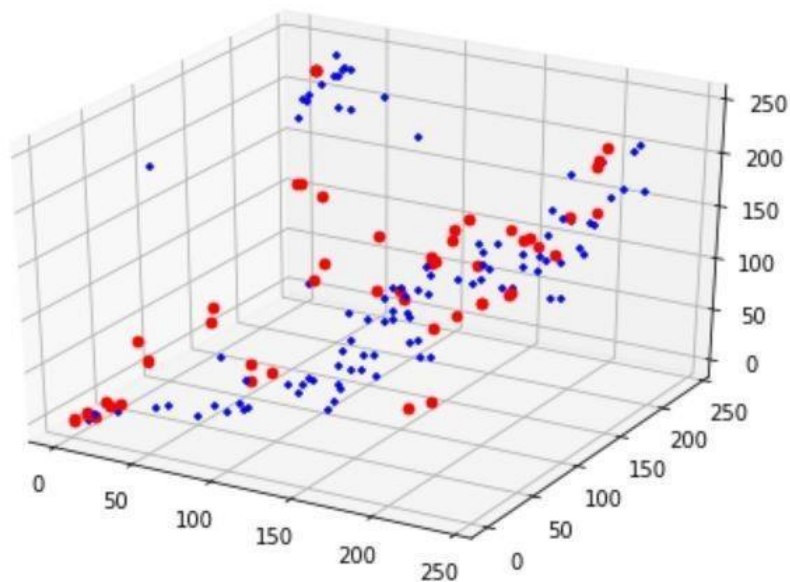
Output:



For a sample of 200:

```
import numpy as np
import matplotlib.pyplot as plt

from mpl_toolkits import mplot3d
fig = plt.figure()
ax = Axes3D(fig)
dc1=dc[0:100]
dc2=dc[100:200]
for i in dc1:
    ax.scatter3D(i[0],i[1],i[2],marker='+',c='b')
for i in dc2:
    ax.scatter3D(i[0],i[1],i[2],c='r')
plt.show()
```



Performance Measures

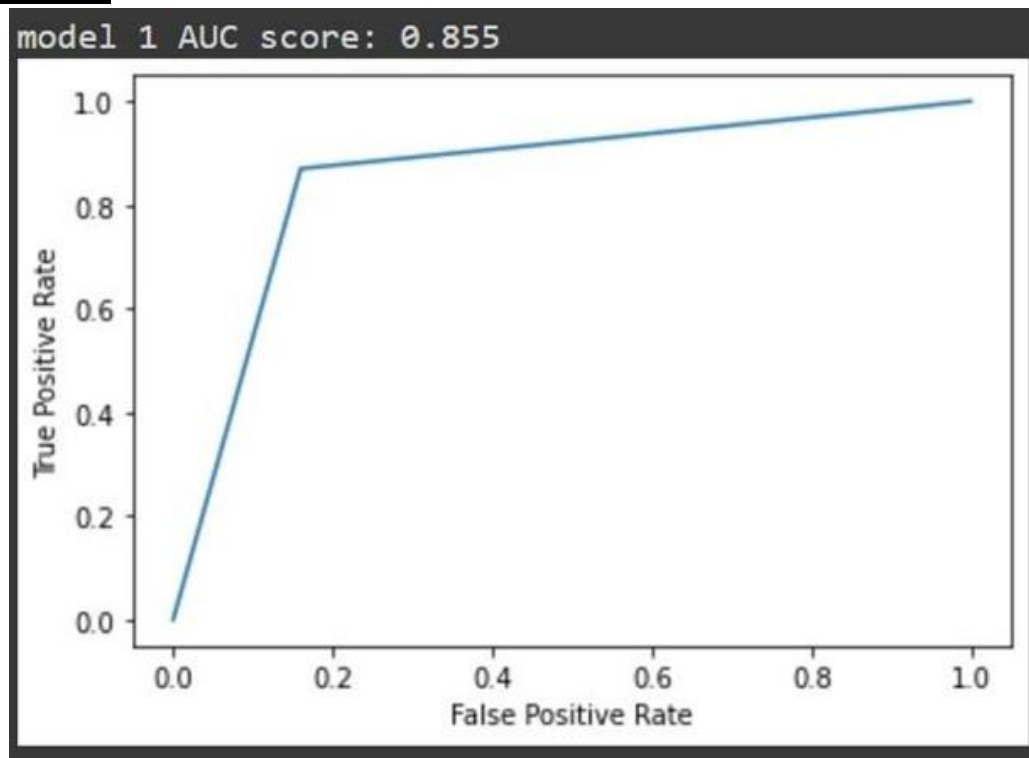
```
ypred=clf1.predict(X_test)
import matplotlib.pyplot as plt
import numpy
from sklearn import metrics

actual = numpy.random.binomial(1,.9,size = 1000)
predicted = numpy.random.binomial(1,.9,size = 1000)

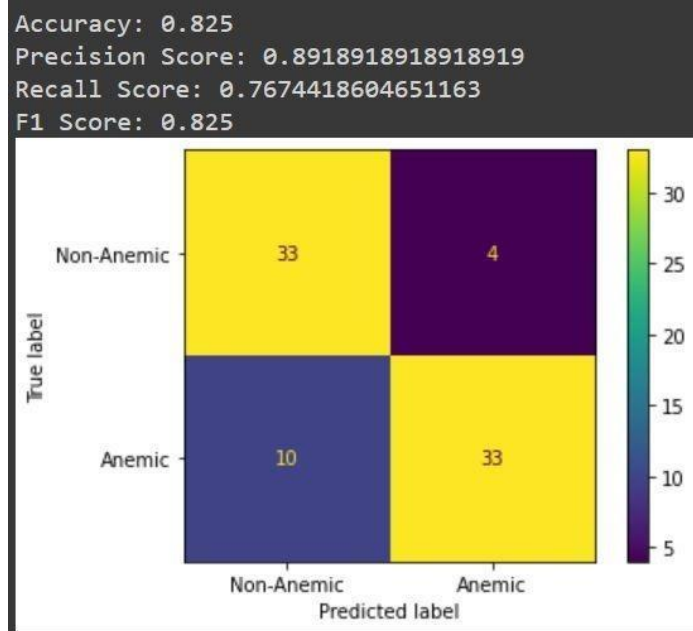
confusion_matrix = metrics.confusion_matrix(y_test, ypred)

cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = confusion_matrix, display_labels = ["Non-Anemic", "Anemic"])
print("Accuracy:",metrics.accuracy_score(y_test, ypred))
## Precision
print("Precision Score:",metrics.precision_score(y_test, ypred))
## Recall
print("Recall Score:",metrics.recall_score(y_test, ypred))
## F1 Score
print("F1 Score:",metrics.f1_score(y_test, ypred))
cm_display.plot()
plt.show()
```

ROC Curve:



Confusion Matrix



Conclusion:

For the project on Non-Invasive Anemia Detection based on ACM April 2022 Paper (iNAP: A Hybrid Approach for Non-Invasive Anemia-Polycythemia Detection in the IoMT), we have developed a model for Anemia detection in patients. The project involves the use of Machine Learning algorithms like K-Means Clustering and MultiLayer Perceptron (MLP) to classify individuals as either anemic or non-anemic. The fingernails image dataset has been used to obtain the results. Algorithms have been used to find out the dominant color from each image of the dataset and based on a certain threshold, KMeans and MLP algorithms have been applied to classify individuals as either anemic or non-anemic. The model has an accuracy of 0.825(82.5%), a precision score of 0.8918918918918919, a recall score of 0.7674418604651163 and a F1 score of 0.825. Thus, the project finds its application in the medical field where is applied to various scopes. Since it is a non-invasive method of anemia detection in patients, it is most preferable in the medical domain to detect anemia in patients.

References:

1. Ghosal, S., Das, D., Udutalapally, V. and Wasnik, P.N., 2022. iNAP: A Hybrid Approach for NonInvasive Anemia-Polycythemia Detection in the IoMT. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(3), pp.1-28.
2. Adams, C., 2019. Ergonomic lighting levels by room for residential spaces. *Available at <https://www.thoughtco.com/lighting-levels-by-room-1206643>*.
3. Ajmal, A.A., Shankarnath, S., Athif, M. and Jayatunga, E.H., 2019, November. Noninvasive screening tool to detect anemia. In *2019 IEEE Healthcare Innovations and Point of Care Technologies, (HI-POCT)* (pp. 67-70). IEEE.
4. AlZahir, S. and Donker, H., 2010. A novel regression based model for detecting anemia using color microscopic blood images. *Journal of Software Engineering and Applications*, 3(8), p.756.
5. Bevilacqua, V., Dimauro, G., Marino, F., Brunetti, A., Cassano, F., Di Maio, A., Nasca, E., Trotta, G.F., Girardi, F., Ostuni, A. and Guarini, A., 2016, May. A novel approach to evaluate blood parameters using computer vision techniques. In *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (pp. 1-6). IEEE.
6. Biswas, S.K., Chatterjee, S., Bandyopadhyay, S., Kar, S., Som, N.K., Saha, S. and Chakraborty, S., 2021. Smartphone-enabled paper-based hemoglobin sensor for extreme point-of-care diagnostics. *ACS sensors*, 6(3), pp.1077-1085.
7. Calinski, T., 1974. A dendrite method for cluster analysis. *Communication in statistics*, 3, pp.1-27.
8. Collings, S., Thompson, O., Hirst, E., Goossens, L., George, A. and Weinkove, R., 2016. Non-invasive detection of anaemia using digital photographs of the conjunctiva. *PloS one*, 11(4), p.e0153286.
9. Davies, D.L. and Bouldin, D.W., 1979. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), pp.224-227.
10. Didzun, O., De Neve, J.W., Awasthi, A., Dubey, M., Theilmann, M., Bärnighausen, T., Vollmer, S. and Geldsetzer, P., 2019. Anaemia among men in India: a nationally representative cross-sectional study. *The Lancet Global Health*, 7(12), pp.e1685e1694.
11. Dimauro, G., Baldari, L., Caivano, D., Colucci, G. and Girardi, F., 2018, June. Automatic segmentation of relevant sections of the conjunctiva for non-invasive anemia detection. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)* (pp. 1-5). IEEE.
12. Dimauro, G., Caivano, D. and Girardi, F., 2018. A new method and a non-invasive device to estimate anemia based on digital images of the conjunctiva. *IEEE Access*, 6, pp.46968-46975.
13. Dunn, J.C., 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.
14. Elsalamony, H.A., 2017. Anaemia cells detection based on shape signature using neural networks. *Measurement*, 104, pp.50-59.
15. Filonov, S., 2019. Tracking your eyes with Python. *Available at <https://medium.com/@stepanfilonov/tracking-your-eyes-with-python-3952e66194a6>*.

16. Ghosal, S., Das, D., Udutalapally, V., Talukder, A.K. and Misra, S., 2020. sHEMO: Smartphone spectroscopy for blood hemoglobin level monitoring in smart anemicare. *IEEE Sensors Journal*, 21(6), pp.8520-8529.
17. Halder, A., Sarkar, P.K., Pal, P., Chakrabarti, S., Chakrabarti, P., Bhattacharyya, D., Chakraborty, R. and Pal, S.K., 2017. Digital camera-based spectrometry for the development of point-of-care anemia detection on ultra-low volume whole blood sample. *IEEE Sensors Journal*, 17(21), pp.7149-7156.
18. Hasan, M.K., Sakib, N., Field, J., Love, R.R. and Ahamed, S.I., 2019. A novel technique of noninvasive hemoglobin level measurement using hsv value of fingertip image. *arXiv preprint arXiv:1910.02579*.
19. Indi, T.S. and Gunge, Y.A., 2016. Early stage disease diagnosis system using human nail image processing. *IJ Information Technology and Computer Science*, 7, pp.3035.
20. Joshi, K.D. and Nalwade, P.S., 2013. Modified k-means for better initial cluster centres. *International Journal of Computer Science and Mobile Computing*, 2(7), pp.219-223.
21. Macknet, M.R., Allard, M., Applegate, R.L. and Rook, J., 2010. The accuracy of noninvasive and continuous total hemoglobin measurement by pulse CO-Oximetry in human subjects undergoing hemodilution. *Anesthesia & Analgesia*, 111(6), pp.1424-1426.
22. Mannino, R.G., Myers, D.R., Tyburski, E.A., Caruso, C., Boudreaux, J., Leong, T., Clifford, G.D. and Lam, W.A., 2018. Smartphone app for non-invasive detection of anemia using only patient-sourced photos. *Nature communications*, 9(1), pp.1-10.
23. Mente, R. and Marulkar, S.V., 2017. A review: Fingernail images for disease detection. *Int. J. Eng. Comput. Sci*, 6(11), pp.22830-22835.
24. Mitani, A., Huang, A., Venugopalan, S., Corrado, G.S., Peng, L., Webster, D.R., Hammel, N., Liu, Y. and Varadarajan, A.V., 2020. Detection of anaemia from retinal fundus images via deep learning. *Nature Biomedical Engineering*, 4(1), pp.18-27.
25. World Health Organization, 2008. Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia.
26. Park, S.M., Visbal-Onufrak, M.A., Haque, M.M., Were, M.C., Naanyu, V., Hasan, M.K. and Kim, Y.L., 2020. mHealth spectroscopy of blood hemoglobin with spectral superresolution. *Optica*, 7(6), pp.563-573.
27. Polu, S.K. and Polu, S.K., 2019. IoMT based smart health care monitoring system. *International Journal for Innovative Research in Science & Technology*, 5(11), pp.58-64.
28. Raedler, L.A., 2014. Diagnosis and management of polycythemia vera: proceedings from a multidisciplinary roundtable. *American health & drug benefits*, 7(7 suppl3), p.S36.
29. Ranganathan, H. and Gunasekaran, N., 2006. Simple method for estimation of hemoglobin in human blood using color analysis. *IEEE transactions on Information Technology in Biomedicine*, 10(4), pp.657-662.
30. Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, pp.53-65.
31. Sari, M., Pee, S.D., Martini, E., Herman, S., Bloem, M.W. and Yip, R., 2001. Estimating the prevalence of anaemia: a comparison of three methods. *Bulletin of the World Health Organization*, 79(6), pp.506-511.
32. Sevani, N. and Persuleshy, G.B.V., 2018, September. Detection anemia based on conjunctiva pallor level using k-means algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 420, No. 1, p. 012101). IOP Publishing.
33. Stott, G.J. and Lewis, S.M., 1995. A simple and reliable method for estimating haemoglobin. *Bulletin of the World Health Organization*, 73(3), p.369.

34. Suner, S., Crawford, G., McMurdy, J. and Jay, G., 2007. Non-invasive determination of hemoglobin by digital photography of palpebral conjunctiva. *The Journal of emergency medicine*, 33(2), pp.105-111.
35. Tamir, A., Jahan, C.S., Saif, M.S., Zaman, S.U., Islam, M.M., Khan, A.I., Fattah, S.A. and Shahnaz, C., 2017, December. 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) (pp. 697-701). IEEE.
36. Ghosal, S., Das, D., Udutalapally, V. and Wasnik, P.N., 2022. iNAP: A Hybrid Approach for NonInvasive Anemia-Polycythemia Detection in the IoMT. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(3), pp.1-28.
37. Tibshirani, R., Walther, G. and Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), pp.411-423.
38. van den Bosch, W.A., Leenders, I. and Mulder, P., 1999. Topographic anatomy of the eyelids, and the effects of sex and age. *British journal of ophthalmology*, 83(3), pp.347-352.
39. Viola, P. and Jones, M.J., 2004. Robust real-time face detection. *International journal of computer vision*, 57(2), pp.137-154.
40. Von Schenck, H., Falkensson, M. and Lundberg, B., 1986. Evaluation of "HemoCue," a new device for determining hemoglobin. *Clinical chemistry*, 32(3), pp.526-529.
41. Wang, E.J., Li, W., Hawkins, D., Gernsheimer, T., Norby-Slycord, C. and Patel, S.N., 2016, September. HemaApp: noninvasive blood screening of hemoglobin using smartphone cameras. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 593-604).