A Predictive and Machine Learning Approach to Non-Invasive Anemia Diagnosis

Ashley Chompre, Connie Xu, Erin Dollard, Grace Wei, Helen Xu, Isabelle Xiong, Jennifer Teo, Jay Yoon, Karik Pejavara, Khushmeet Chandi, Kunal Kapoor, Kyle Ferguson, Larry Jiang, Sadab Mannan, Olivia Fan, Rayan Malik, Selena Halabi, Sophie Wu, Sunggun Lee, Tingnan Hu
Dominic Garrity, Dr. Adam Wax, Dr. Nirmish Shah

DukeHealth

Duke UNIVERSITY | BASS CONNECTIONS

BACKGROUND

- Anemia, a condition characterized by impaired tissue oxygenation, weakness, fatigue, and decreased cognitive capacity, affects about ⅛ of the world's population.
- Regions in Sub-Saharan Africa and South Asia are most affected by the disease, with the highest prevalence among children under five years of age and maternal populations.
- Patient hemoglobin (Hb) levels are the primary indicator of clinically diagnosed anemia.
- Currently, hemoglobin is measured using invasive techniques or automated hematology tools, which may be expensive and not readily available in low-resource settings.

PRIOR RESEARCH

- Numerous studies have demonstrated the potential of machine learning to automate visual diagnosis of diseases and conditions in healthcare settings.[1, 2]
- Mannino et al. proved that non-invasive diagnosis of anemia was possible using only a smartphone app and nail bed images, achieving an accuracy of ±2.4 g /dL and a sensitivity of 97% (95% CI, 89–100%) when compared with CBC hemoglobin levels.[3]

QUESTION & HYPOTHESIS

Research Questions: Can predictive models and machine learning accurately estimate a patient's hemoglobin concentration from images of their finger nail bed photos? How can a mobile app be designed to be user-friendly and accessible to patients and physicians?

Hypothesis: Predictive models and machine learning can be used to develop a non-invasive and accurate tool for anemia diagnosis using smartphone photos of the nail bed.

METHODS

- **Data Collection:**
  - Patients in this study were recruited from the Duke Sickle Cell Center and Duke Hematology Clinic.
  - Patient characteristics and images of the hand were obtained for each participant, along with a color calibrated metric alongside the hand.
- Data analysis was conducted in Python (Python 3.11.2)
- Datasets were used to train predictive models that leverage statistical analysis and machine learning to predict hemoglobin concentration from image data.
- A cross-platform mobile app is being developed as an accessible resource for testing in field settings.

METHODS (CONTINUED)

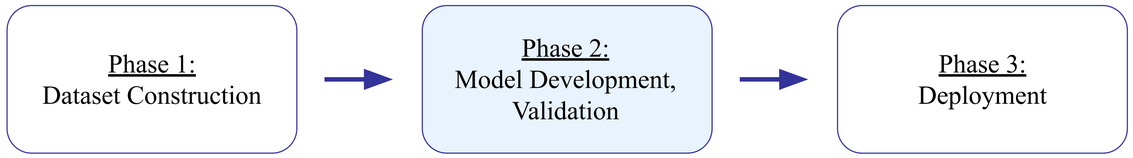


Figure 1. Project Timeline and Goals

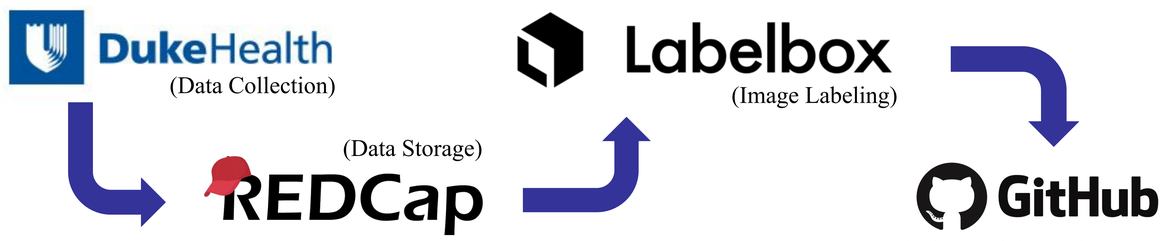
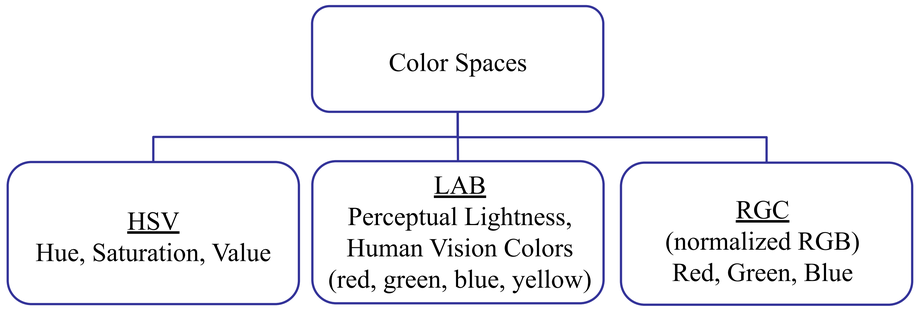


Figure 2. Project Workflow

RESULTS



Color Spaces:
A specific organization of colors or characteristics that describe color in image displays. Color spaces are commonly used in image processing. Various devices support different colors spaces.

Figure 3. Color Space Breakdown

Models:

1. **Multilinear regression (MLR):** statistical approach - originally published by Mannino et al.[3] → performed with k-fold cross validation
2. **Random forest (RF):** gradient boosted decision tree (machine learning) → trained and optimized through grid and random parameter searches

Table 1. Model Performance in Three Color Spaces (R^2 values)

	HSV	LAB	RGC
Multilinear Regression	0.59	0.69	0.64
Random Forest	**0.52**	0.77	**0.89**

- MLR performed most consistently across the three color spaces
- RF exhibited the best fit in RGC color space (R^2 value = 0.89)

CONCLUSIONS

- Results suggest that the random forest model within the RGC color space is the best predictor for hemoglobin concentration
- The MLR model was more precise with its estimates and performed consistently across all color spaces
- These results demonstrate the potential for predictive models, especially machine learning
- Ultimately, these results will be used to improve point-of-care anemia diagnostics by developing a better hematology tool that is non-invasive, accurate, and affordable for low-resource settings.

FUTURE DIRECTIONS

- The project team aims to construct a similar dataset through international institutional collaboration to ensure a diverse patient population for the dataset
- Future work will focus on implementing additional ML models such as a binary classification model or deep learning to improve predictive performance.
- The mobile app will be improved upon so that it may be deployed internationally on mobile devices.

ACKNOWLEDGEMENTS

- Thank you to Dr. Shah and Dr. El-Bitar in Duke University's Medical Center for collaborating with us to collect patient data with which we trained our models.
- Thank you to Dr. Wax for serving as the faculty mentor for this BASS project and providing guidance.
- Thank you to Bass Connections Foundation for making this Bass course possible
- Thank you to the advisors at the Global Alliance for Medical Innovation for the effort over the years to support this project and for being champions of student-driven medical innovation

SUPPORT

- Duke University Medical Center
- Bass Connections
- Global Alliance for Medical Innovation

REFERENCES

1. Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022). Machine-Learning-Based Disease Diagnosis: A Comprehensive Review. Healthcare (Basel, Switzerland), 10(3), 541. https://doi.org/10.3390/healthcare10030541
2. Zhang, Rui et al. "Multiscale Mask R-CNN-Based Lung Tumor Detection Using PET Imaging" Molecular Imaging 18 (2019): https://doi.org/10.1177/1536012119863531.
3. Mannino, R.G., Myers, D.R., Tyburski, E.A. et al. Smartphone app for non-invasive detection of anemia using only patient-sourced photos. Nat Commun 9, 4924 (2018). https://doi.org/10.1038/s41467-018-07262-2