BiliNet: Serum Bilirubin Prediction for Neonates using Segmentation-Guided Neural Networks

Om Shah
NJSHS '24

Subset of research peer-reviewed and presented at the I**EEE-MIT Undergraduate Research Technology Conference**

What is Neonatal Jaundice?

Formation of bilirubin

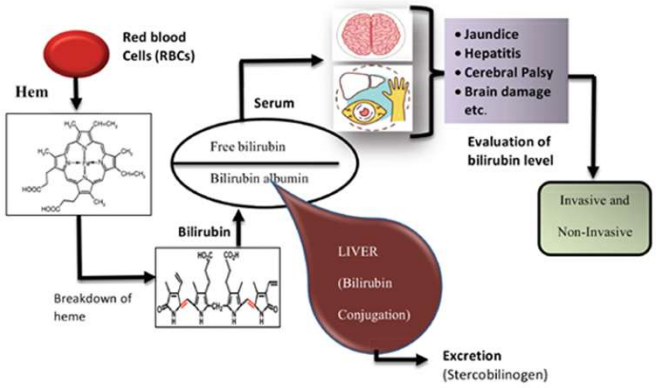


Figure 1: (Kumari et al. 2023)

Disease Characteristics

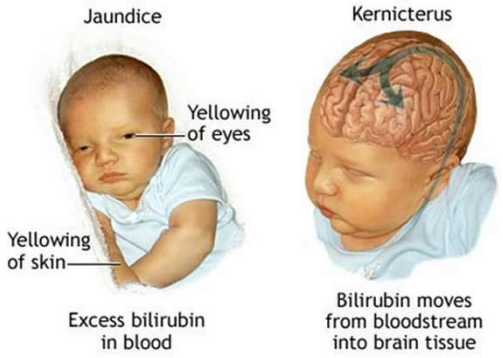


Figure 2: (Pediatric Society of Ghana)

Disease Progression

Bilirubin level indicates severity of jaundice

Mild — **Yellow pigmentation in face and eyes**
Jaundice will likely resolve by itself.

Moderate — **Yellow pigmentation spreads to chest**
Behavioral changes occur; phototherapy necessary.

Severe — **Life-threatening disorders**
Neurotoxicity, permanent brain damage, hearing loss, cerebral palsy, kernicterus; blood transfusion required.

Fatal — **Loss of life**

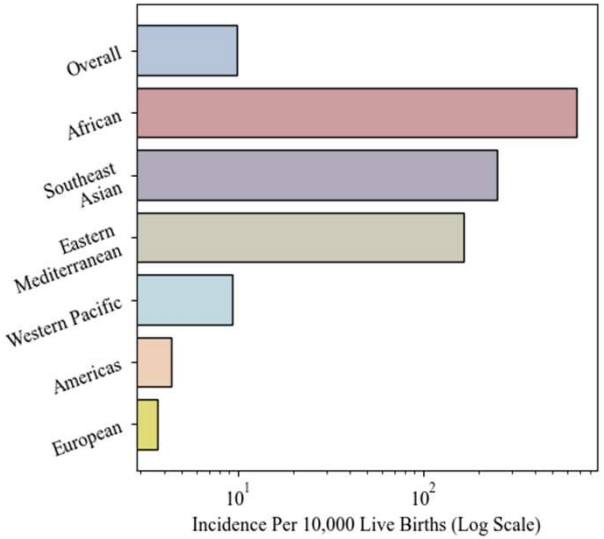
Global Disparities

Skin Tone Differentiation

(Peanuts)

- Difficult to diagnose jaundice in darker-skinned neonates.
- Existing medical technology based on disease characteristics in Caucasian babies

Health Access Divides



~40x greater incidence rate in Africa + South Asia

Figure 3: Geographic rates (Slusher 2017)

Lack of medical infrastructure
Skin tone bias
Low funding

50% of full term babies
80% of preterm babies

→ **70% severe cases in developing countries** (1.1 million annual cases)

→ **75,000 cases of kernicterus**

→ **114,000 deaths** (16-35%)

(NIH, Tessema 2024)

Limitations of Current Approaches

Drawbacks of existing approaches directly correlate with disproportionate incidence rates in developing countries*

Visual Assessment

Unreliable

difficult to assess severity

- Drastically more difficult for darker tones
- Requires frequent observation

Laboratory Blood Test

Time Intensive

takes between 6-24 hours

- Requires existing medical infrastructure
- Requires medical professionals

Transcutaneous bilirubinometer

Expensive

costs between $3,000-$7,000

- Inaccurate outside operating range
- Worse performance on preterm infants

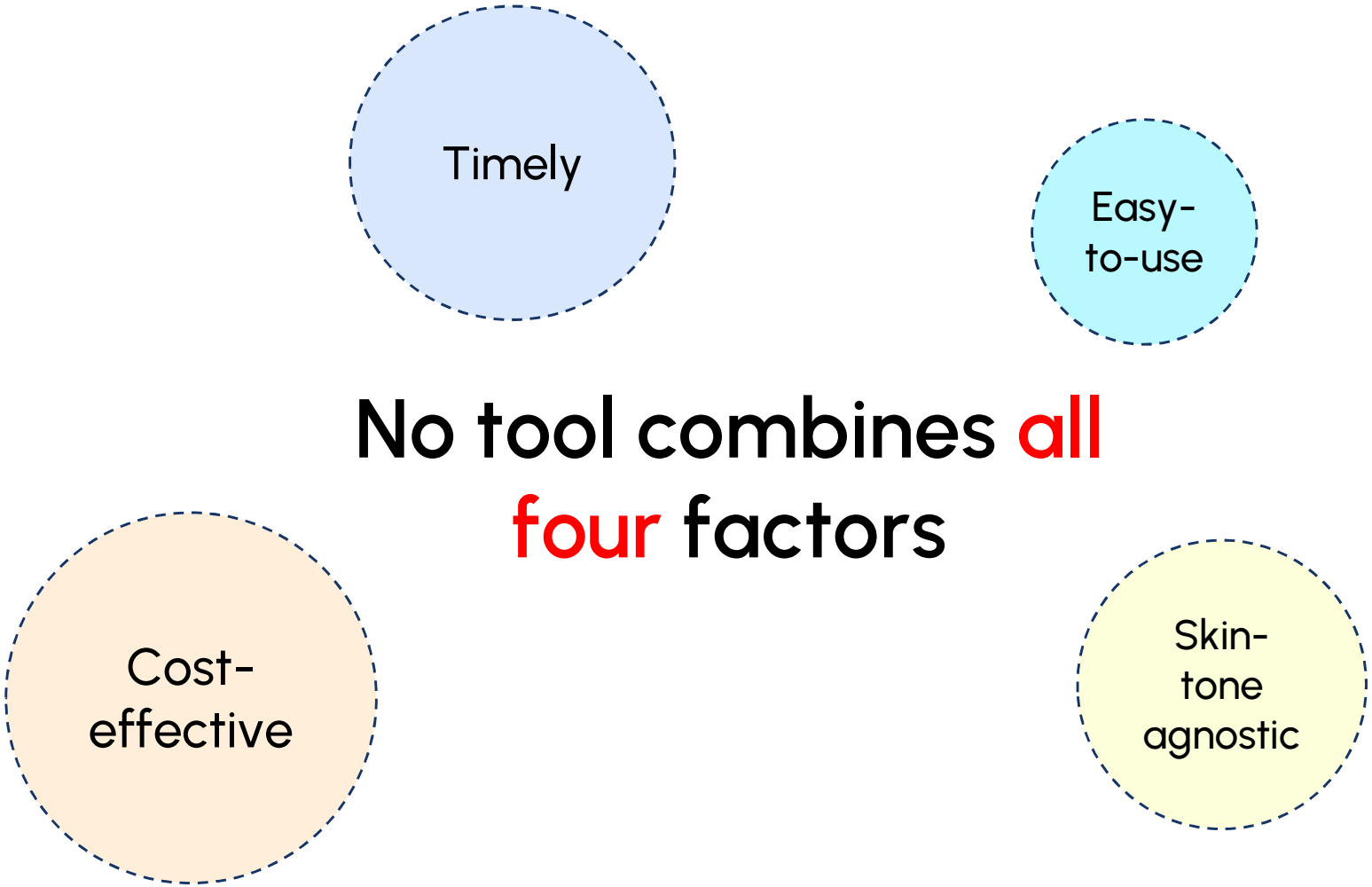
Computer Vision

Difficult to Use

skin/eye features inconsistent

- Fails on darker skin tones
- Eye movements are difficult to track in clinical settings

*Diala 2023, Olusanya 2018, Asefa 2020


Timely
Easy-to-use
No tool combines all four factors
Cost-effective
Skin-tone agnostic

Research Objectives

Guiding Question: Can we predict bilirubin levels accurately, inexpensively, and without bias?

Engineering Goals

1. **Performance:** implement a tool that predicts bilirubin volume with at least 80% correlation to laboratory measurements
2. **Accessible:** utilize a data source that does not bias against darker skin tones.
3. **Scalable & Easy to Use:** an end-to-end system that provide bilirubin volume in under 1 minute.
4. **Low-Cost:** costs less than $5 per test.
5. **Clinical Efficacy:** bilirubin measurements enable severity diagnosis and phototherapy classification with 80% accuracy

Hypothesis

I hypothesized that a machine learning model would be able to learn relationships useful for bilirubin prediction at low financial and temporal cost.

Variables

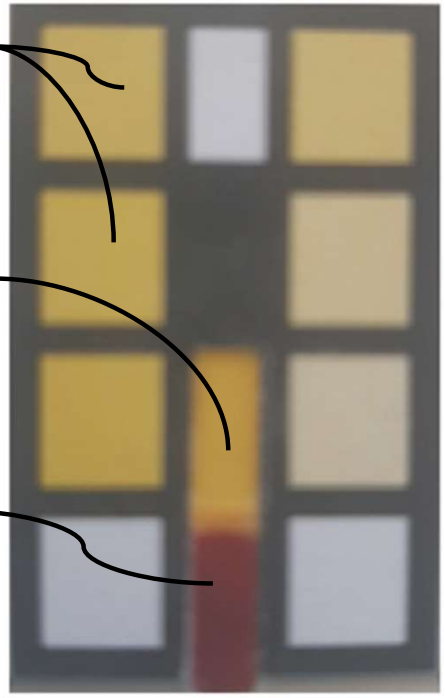
Independent	Dependent
Model input data (source of blood)	Model metrics (IoU, Dice, Pearson R, Spearman R, Sensitivity, Specificity, AUC)

Controlled
Ground truth data (filter box masks, lab bilirubin measurements)

BiliNet Methodology

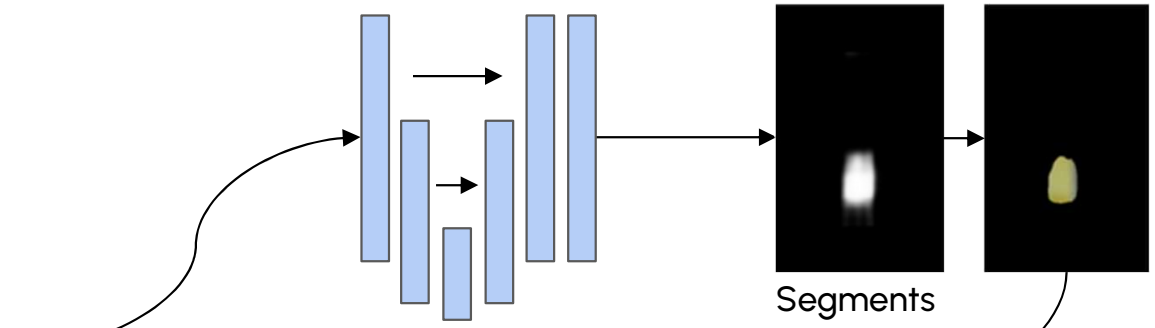
Step 1
Plasma separates from blood on test strip.

Plasma Data

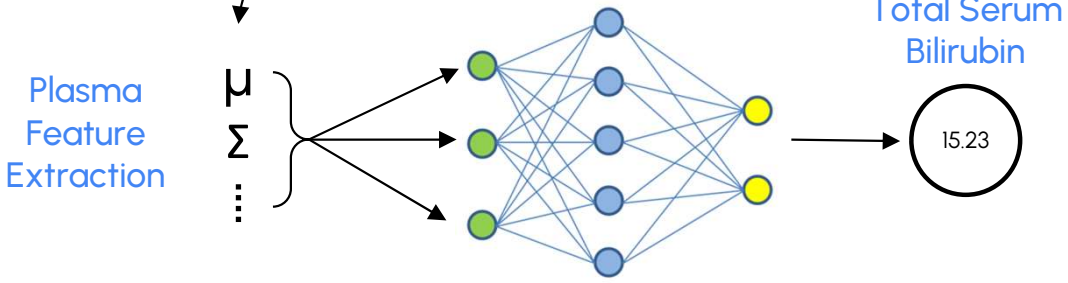


Nitrocellulose Test Strip**

Segmentation Model



Bilirubin Prediction Model



** From Seattle University medical-grade jaundice research dataset

BiliNet Methodology

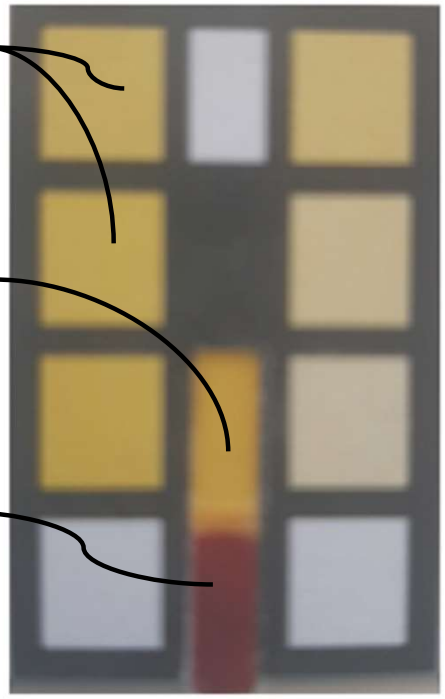
Step 1
Plasma separates from blood on test strip.

Plasma Data

Calibration Boxes

Extracted plasma bilirubin (filter box)

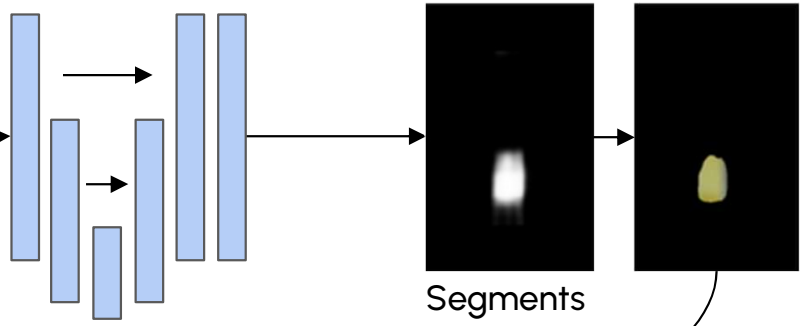
Blood Sample



Nitrocellulose Test Strip**

Segmentation Model

Step 2
Machine learning model finds plasma region.



Segments

Segmentation

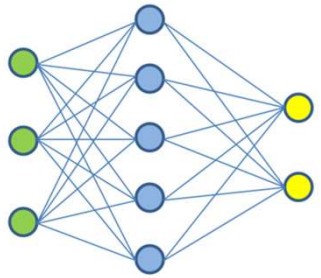
Regression

Bilirubin Prediction Model

Plasma Feature Extraction

μ

Σ



Total Serum Bilirubin

15.23

** From Seattle University medical-grade jaundice research dataset

BiliNet Methodology

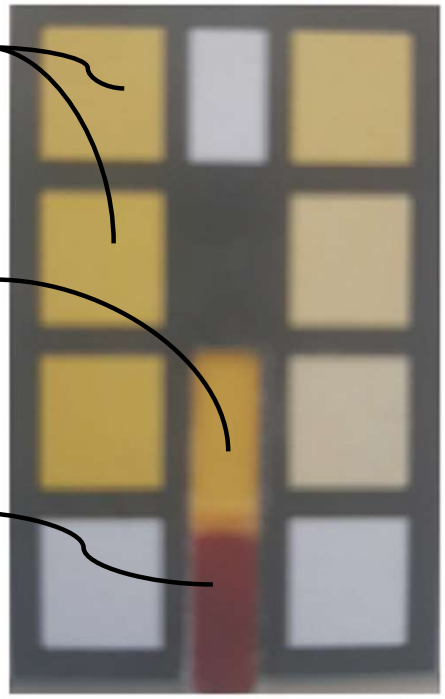
Step 1
Plasma separates from blood on test strip.

Plasma Data

Calibration Boxes

Extracted plasma bilirubin (filter box)

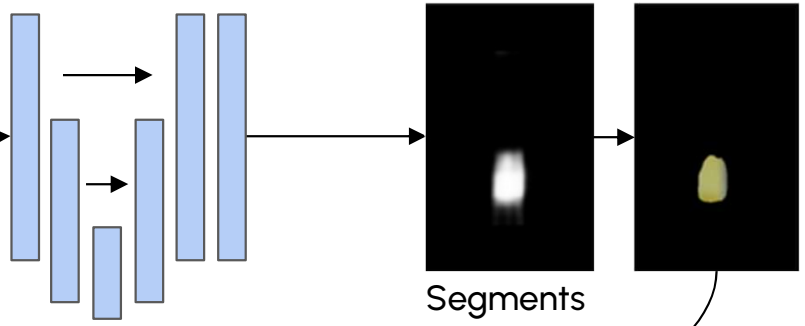
Blood Sample



Nitrocellulose Test Strip**

Segmentation Model

Step 2
Machine learning model finds plasma region.



Segments

Segmentation

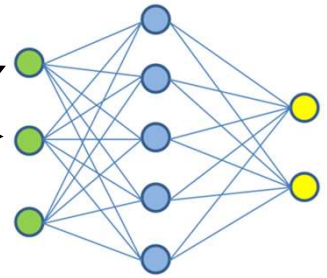
Regression

Step 3
Second model uses color features to predict TSB.

Bilirubin Prediction Model

Plasma Feature Extraction

μ
Σ
⋮



Total Serum Bilirubin

15.23

** From Seattle University medical-grade jaundice research dataset

BiliNet Methodology

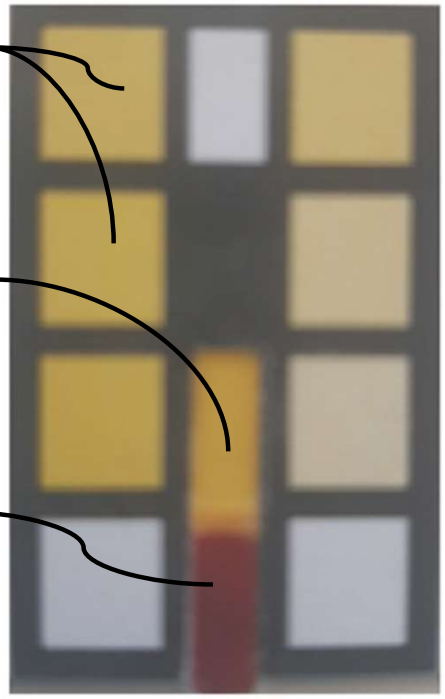
Step 1
Plasma separates from blood on test strip.

Plasma Data

Calibration Boxes

Extracted plasma bilirubin (filter box)

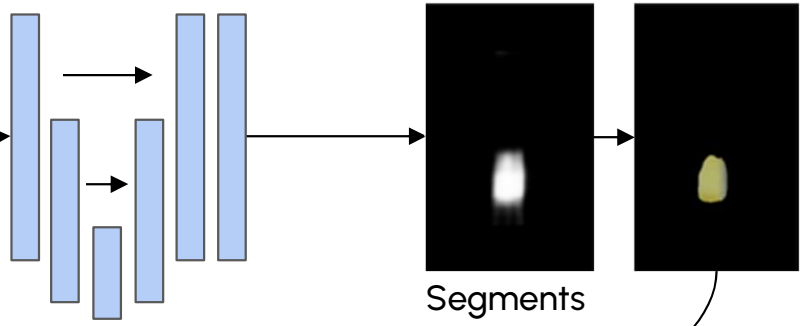
Blood Sample



Nitrocellulose Test Strip**

Segmentation Model

Step 2
Machine learning model finds plasma region.



Segments

Segmentation

Regression

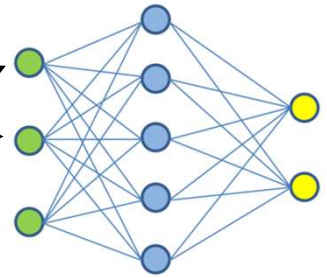
Step 3
Second model uses color features to predict TSB.

Bilirubin Prediction Model

Step 4
Custom-built models are evaluated

Plasma Feature Extraction

μ
Σ
⋮

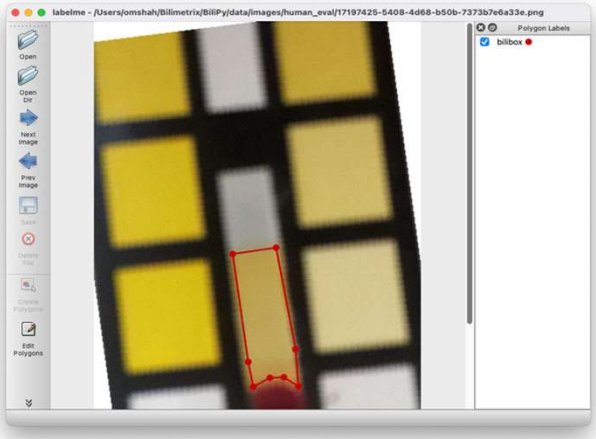


Total Serum Bilirubin

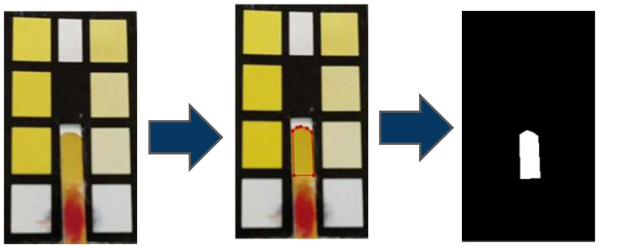
15.23

** From Seattle University medical-grade jaundice research dataset

Data Annotation Framework

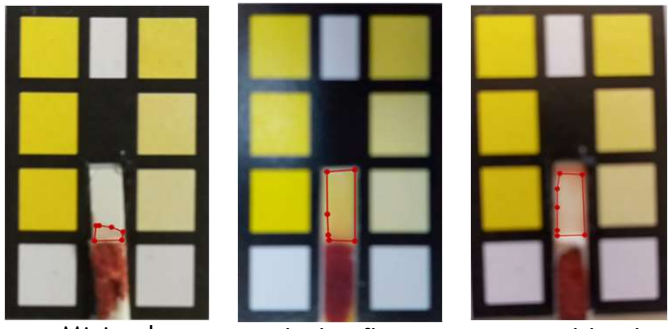


819 test strip images annotated

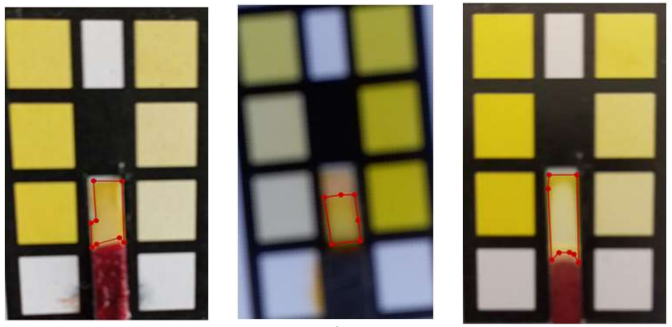


Original Image | Polygon Annotation | Binary Mask

Goal: build high quality dataset



Minimal bilirubin flow | Bilirubin flow gradient | Major blood interference



Splotchy bilirubin flow | Significant blur and rotation | Bilirubin collection near edges

Annotation Goals

Exclude Blood Interference

Minimize Boundary Uncertainty

Include Complete Bilirubin Flow Gradients

Segmentation Model

Goal: find filter box and its RGB values

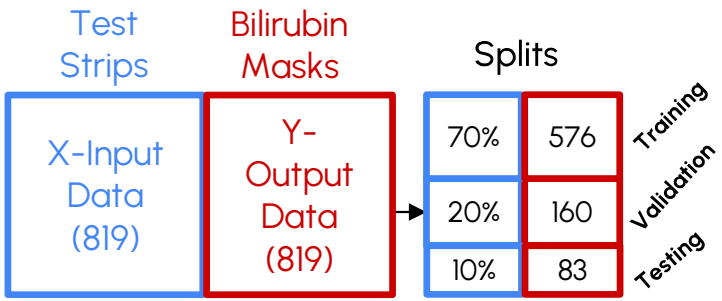
Data Preprocessing

Image Augmentation

Random transformations reflect real-world irregularities

Normal | Rotated + scaled | Brightness contrast | All transforms

Splitting X & Y into datasets



Segmentation Model

Goal: find filter box and its RGB values

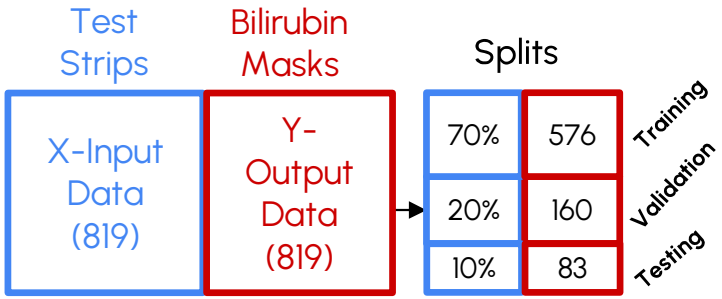
Data Preprocessing

Image Augmentation

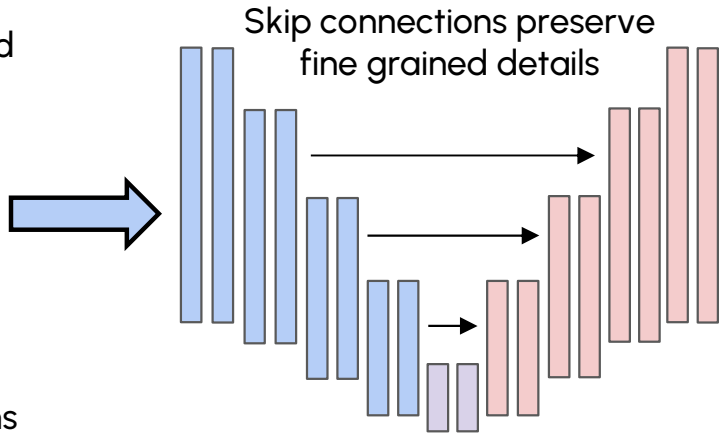
Random transformations reflect real-world irregularities

Normal | Rotated + scaled | Brightness contrast | All transforms

Splitting X & Y into datasets



U-Net: Encoder-Decoder



Downsampling finds "what" is in the image (feature extractor)

Upsampling translates high level features → specific segmentation

Parameters: 31,031,745
Adam optimizer: minimizing log loss
Compared to: DeepLabv3

Segmentation Model

Goal: find filter box and its RGB values

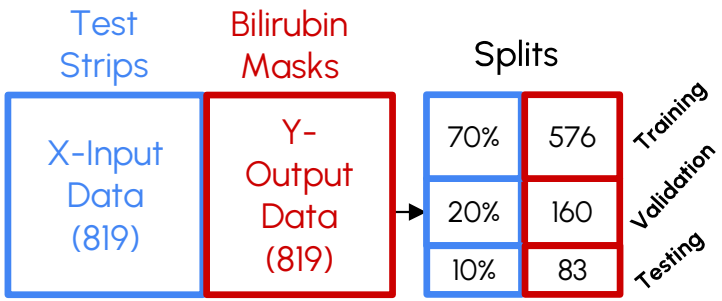
Data Preprocessing

Image Augmentation

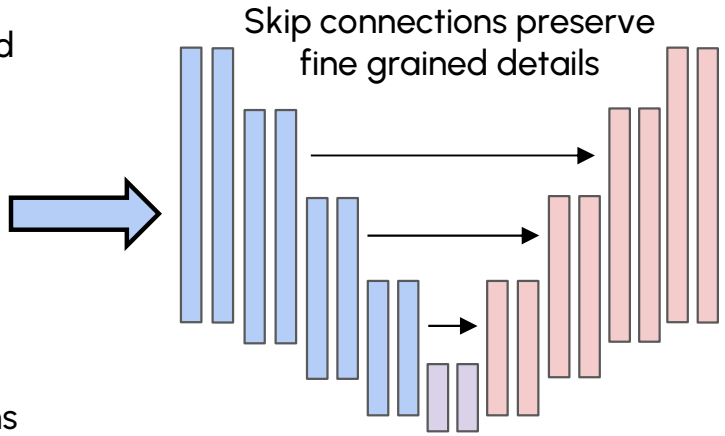
Random transformations reflect real-world irregularities

Normal | Rotated + scaled | Brightness contrast | All transforms

Splitting X & Y into datasets



U-Net: Encoder-Decoder



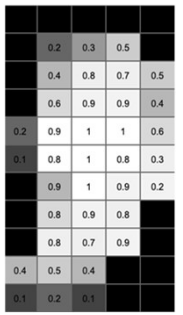
Downsampling finds "what" is in the image (feature extractor)

Upsampling translates high level features → specific segmentation

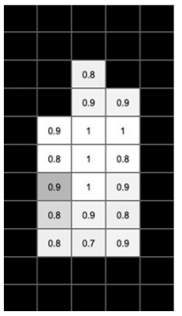
Parameters: 31,031,745
Adam optimizer: minimizing log loss
Compared to: DeepLabv3

Post Processing

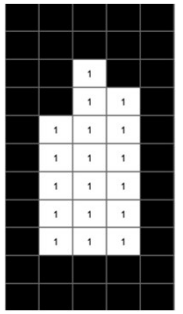
1. Raw



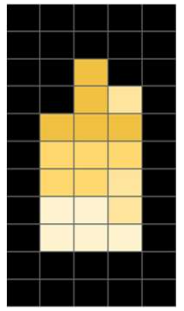
2. Threshold



3. Binarize



4. Stack



$$\text{prediction}_{i,j} \in [0, 1]$$

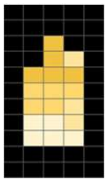
$$M_{\text{binarybitwise}} = \{\text{filter}_{i,j} \mid \text{prediction}_{i,j} > 0.8\}$$

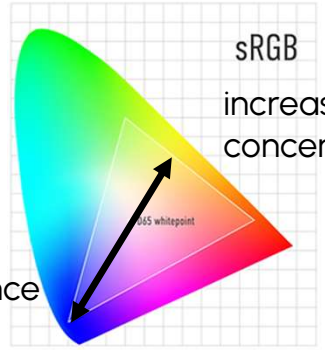
threshold

Mask Analysis and Feature Engineering

Goal: build features

Bilirubin concentration tracks yellow-blue axis

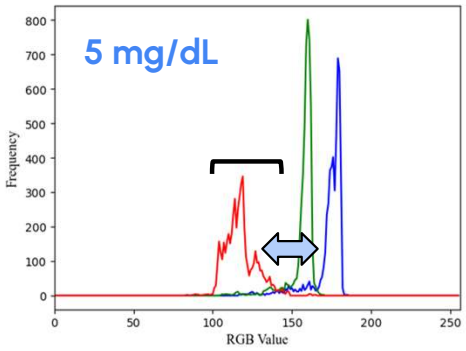


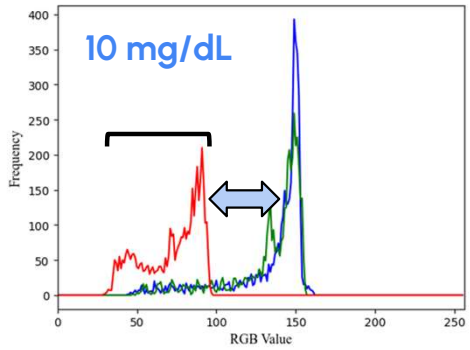


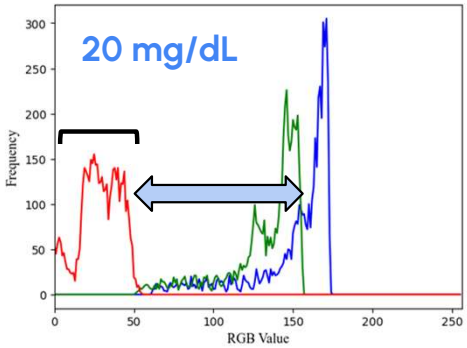
blue is reference point

(Unravel.com)

Bilirubin intervals demonstrates yellow-blue relationship + non-linear traits



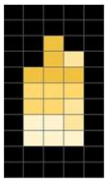


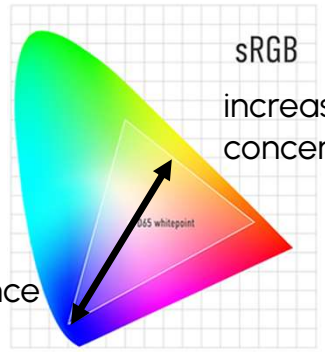


Mask Analysis and Feature Engineering

Goal: build features

Bilirubin concentration tracks yellow-blue axis

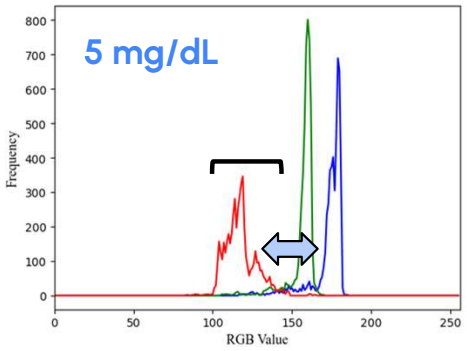


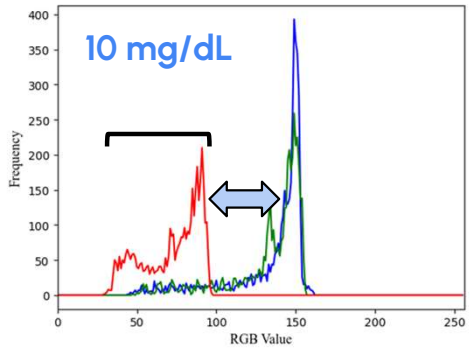


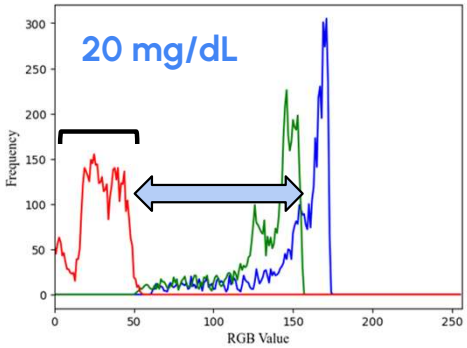
blue is reference point

(Unravel.com)

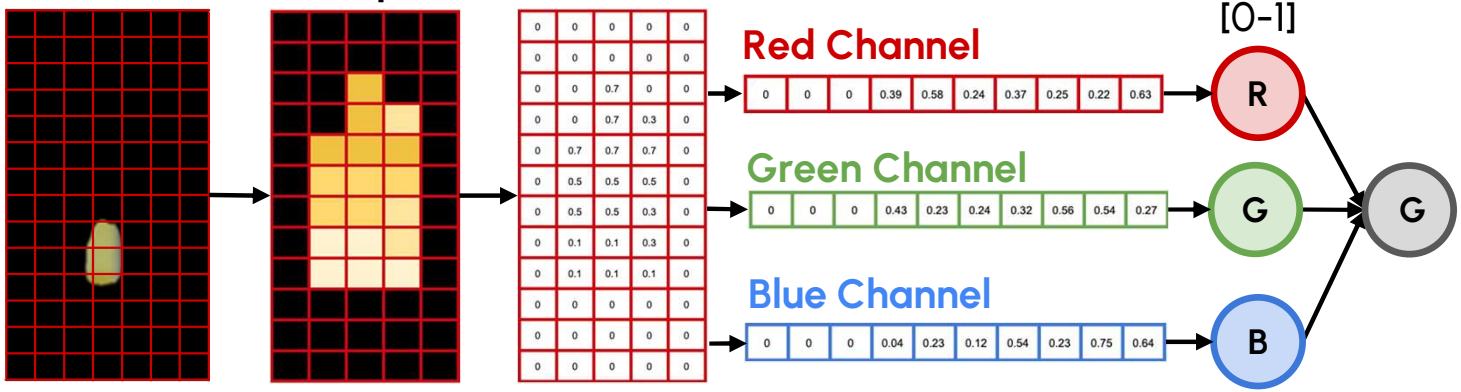
Bilirubin intervals demonstrates yellow-blue relationship + non-linear traits







Mask to Features Pipeline

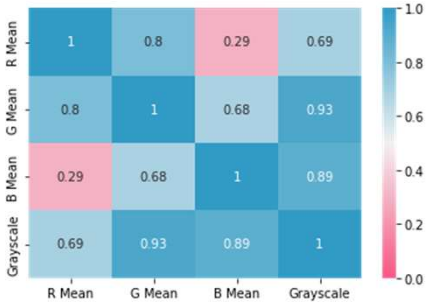


512x320 pixels

Zoomed-in

RGB Matrix

Heatmap depicts non-linearity between input features



Dimensionality reduction from ~2500 pixels to 4 features

Bilirubin Prediction Model

Goal: predict total serum bilirubin

Features

Red Channel Mean
- Specific intensity of yellow

Green Channel Mean
- Specific intensity of yellow

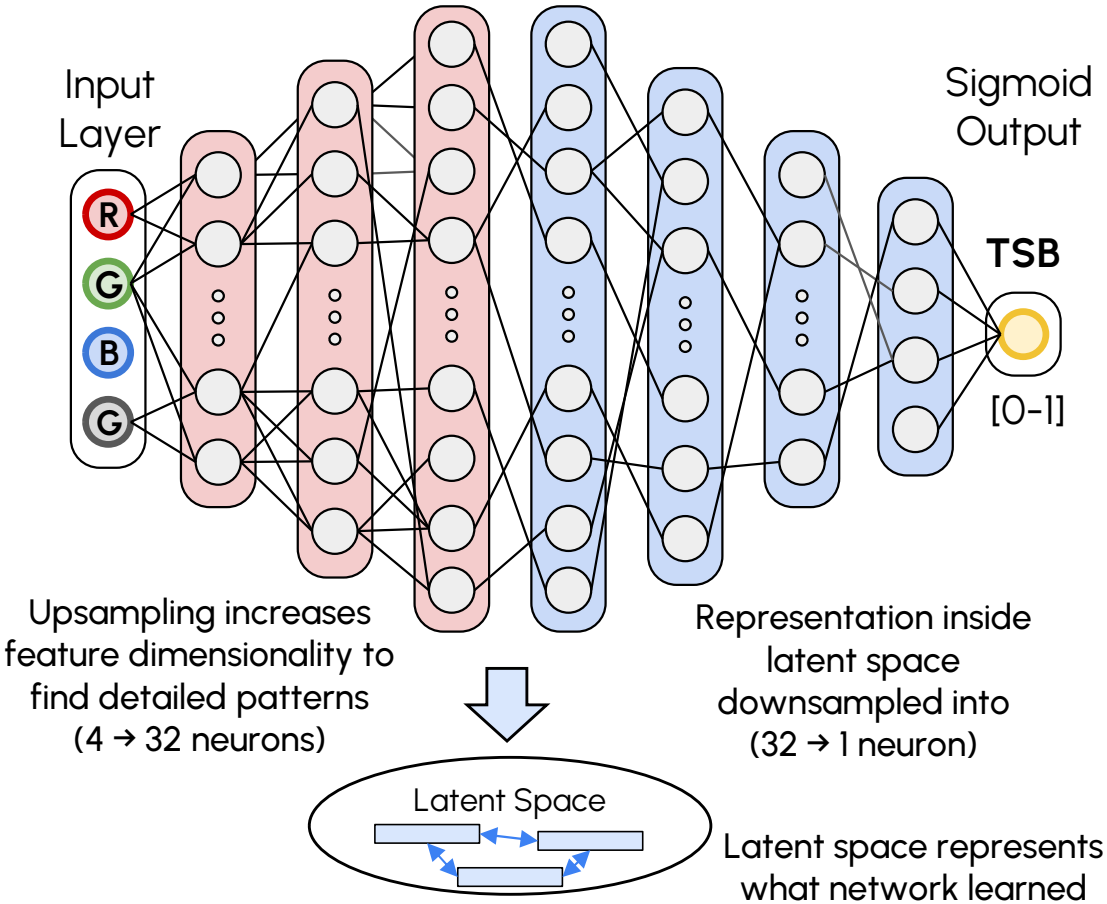
Blue Channel Mean
- Reference point + ambient lighting

Greyscale
- Overall intensity + luminance

Parameters: 2047
Adam optimizer: minimizing MSE
Initializer: glorot normal facilitates reliable convergence

Compared to: random forest, ridge regression (L2), support vector regression

Neural Network Maps Features to TSB



Model Tuning and Optimization

Goal: maximize model performance

What we learned...

Segmentation Task

18 experiments

Activation: ReLU
Batch Size: 16
Epochs: 40
Learning Rate: 0.0001

ReLU: induces optimal non-linearity into model; may result in dead neurons

Learning Rate: slow to learn and does not overfit on data

Regression Task

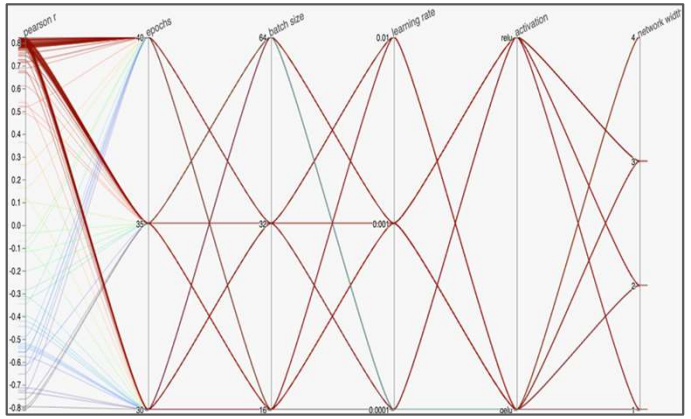
216 experiments

Activation: ReLU
Batch Size: 64
Epochs: 35
Learning Rate: 0.01
Network Width: 1

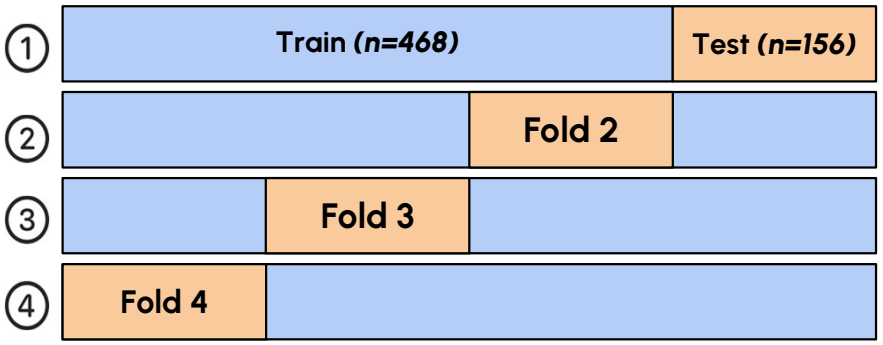
Complexity: narrow layers reduces overfitting

Depth: fewer layers and low feature dimensions means model learns quick and fast

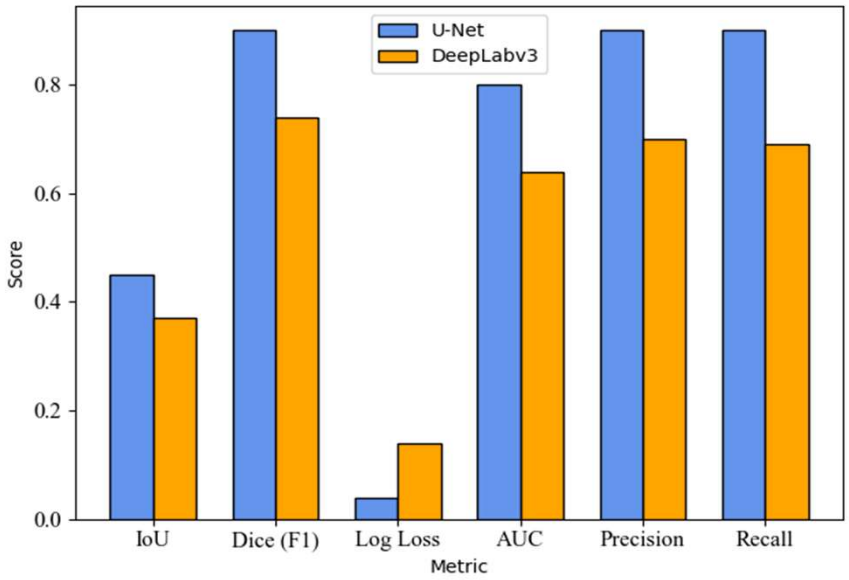
Parallel plot shows neurons per layer and training speed are most critical for regression



Cross validation ensures model is not memorizing labels



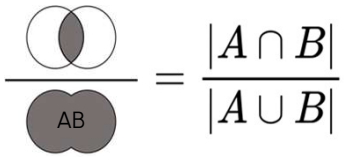
Results: Segmentation Model Selection



U-Net outperforms DeepLabv3 on all metrics due to better neuron distribution

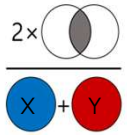
Metrics	Range	U-Net	DeepLabv3
IoU	0 to 1	0.45	0.37
Dice (F1)	0 to 1	0.9	0.74
Log Loss	0 to 1	0.04	0.14
Area under Curve (AUC)	0 to 1	0.8	0.64
Precision	0 to 1	0.9	0.7
Recall	0 to 1	0.9	0.69

Key Metrics



$$= \frac{|A \cap B|}{|A \cup B|}$$

Intersection over Union: performance on continuous filter box region *(p<0.05)*



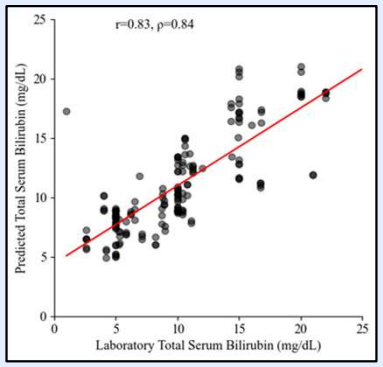
$$= \frac{2\,|X \cap Y|}{|X| + |Y|}$$

Sørensen–Dice coefficient (F1): average pixel-wise binary classification performance *(p<0.05)*

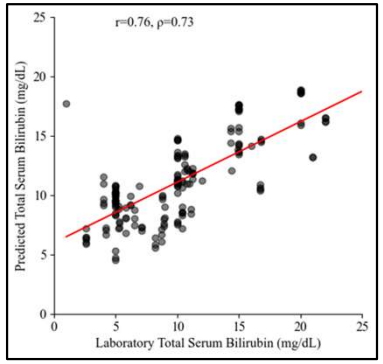
Secondary Metrics

Log Loss: measures confidence on pixel-wise output probabilities (pre-binarization) *(p<0.05)*
Area under (ROC) Curve: relationship between precision and recall
Precision: measures confidence on pixel-wise output probabilities
Recall: measures performance solely on filter box region

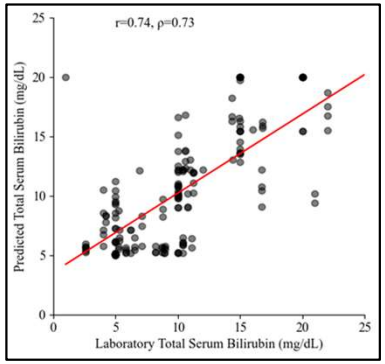
Results: Bilirubin Prediction Model Selection



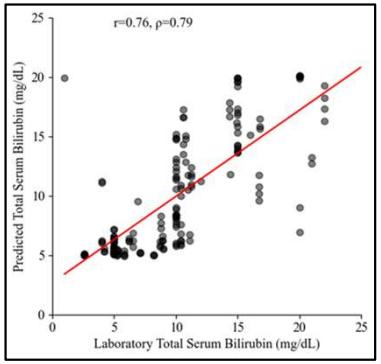
Neural Network
Pearson R: **0.83**
Spearman Rank: **0.84**



Ridge (L2)
Pearson R: 0.76
Spearman Rank: 0.73



Random Forest
Pearson R: 0.74
Spearman Rank: 0.73



Support Vector
Pearson R: 0.76
Spearman Rank: 0.79

- Pearson correlation finds strict linear relationship + is sensitive to outliers

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

- Spearman rank finds non-linear relationships as well + indicates missing features

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Pearson: 83%; Spearman: 84%
Neural network outperforms a variety of regression architectures and exhibits strict linear relationship

Results: Comparison to Gold Standard Blood Tests

R-Squared

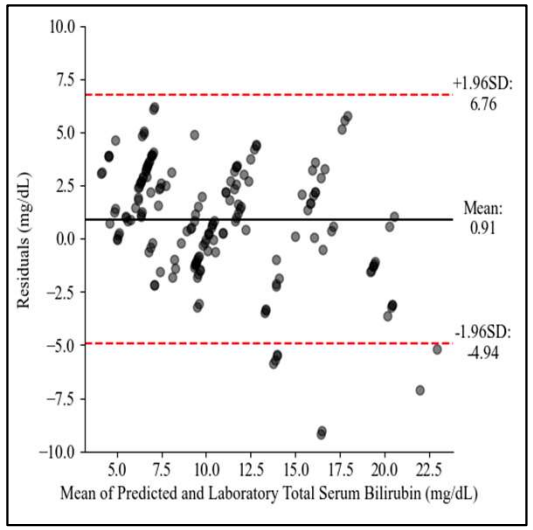
Measures strength of linear relationship between x and y

- Sum of residuals over total variance
- Can be compared to other predictive models

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \overline{y})^2}$$

Model has R-Squared of **69%**, indicating high strength of correlation

Residual TSB Errors



Bland-Altman plot shows consistent error spread + **few outliers**

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - \hat{Y}_i| \qquad RMSE = \sqrt{\sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)^2}{n}}$$

	Metrics	
Correlation	Pearson R	0.83
	Spearman Rank	0.84
	R-Squared	0.69
Error	Mean Absolute Error (mg/dL)	2.38
	Root Mean Squared Error (mg/dL)	3.12
Scale	Total Data Points (training + testing)	624 (468+156)
	Bilirubin Prediction Speed (sec)	10

All metrics cross-validated using four folds (train: 468, test: 156)

Model has strong predictive correlation and low error (2.38 mg/dL) on a large testing dataset

Results: Qualitative Analysis

Input Test Strip
Variety of lighting, scale, and rotation conditions

Ground Truth Mask
General polygon shape across all masks

Predicted Mask (raw)
Predictions reflect environmental + blood irregularities

Post-processed Mask
Final mask contains maximal bilirubin flow

Predicted TSB

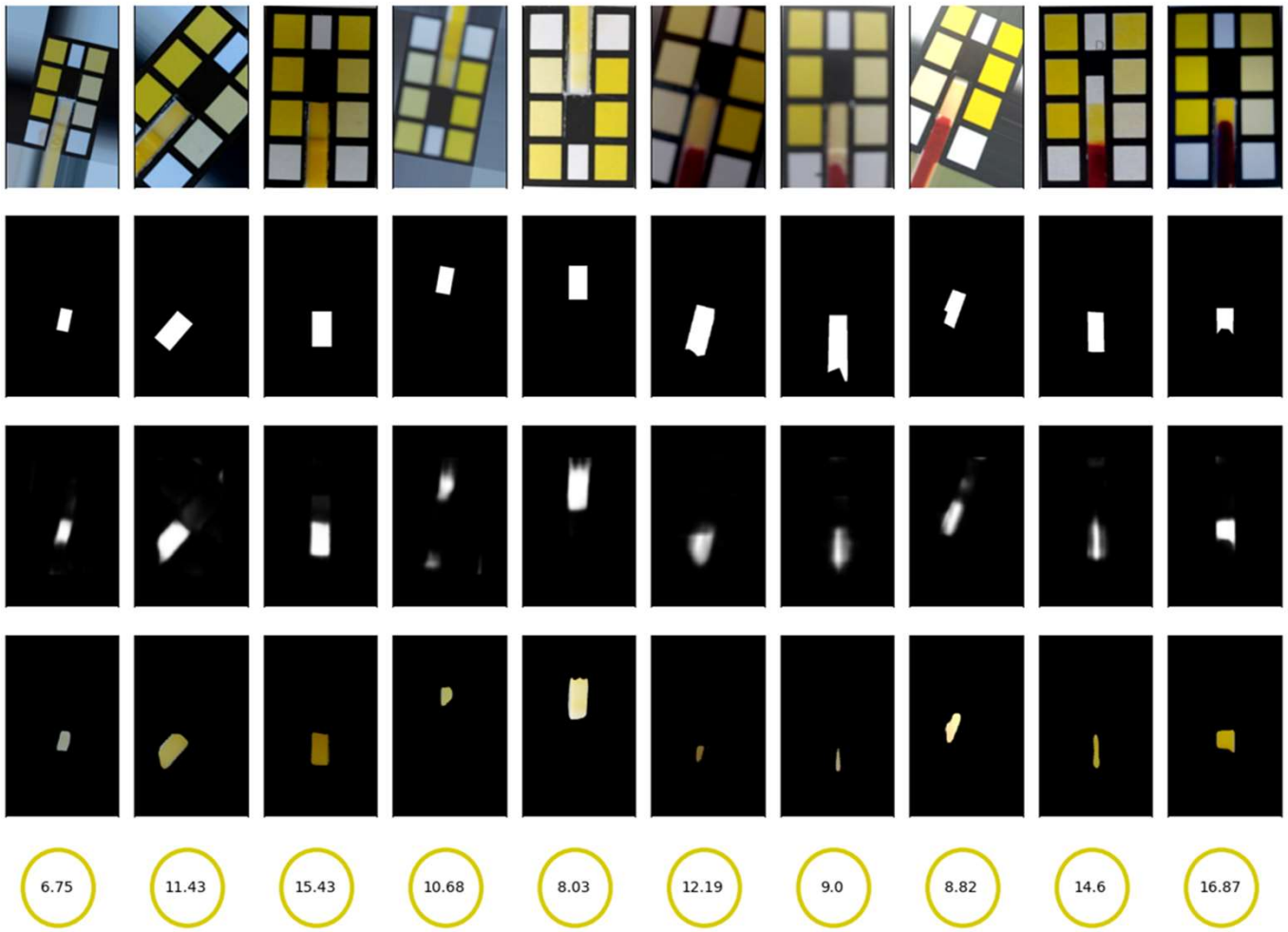
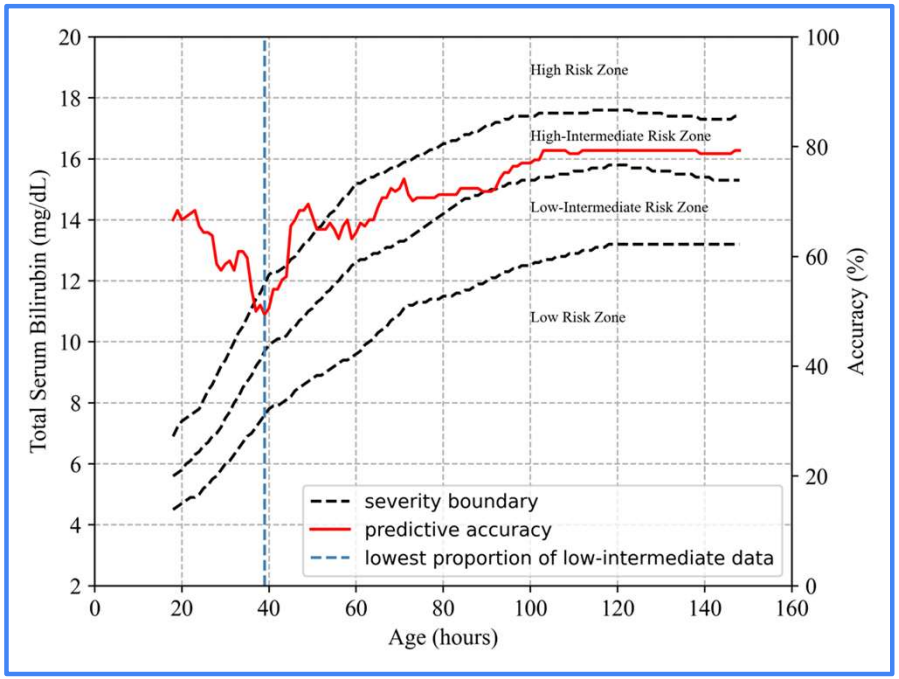


Figure 14: Segmentation performance in variety of test strip scenarios

Clinical Efficacy: Severity Diagnosis

Question: Can BiliNet accurately classify the severity of jaundice?



BiliNet performance on Bhutani Nomogram

- Bhutani Nomogram: mg/dL cutoffs by age for risk of severe hyperbilirubinemia
- Used between 18-144 hours after birth
- Difference between risk zones not discernable to the eye

A: Consistent when there is enough data

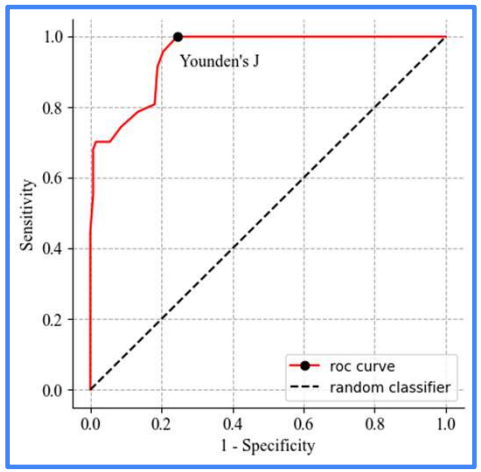
Clinical Efficacy: Phototherapy Diagnosis

Question: Can BiliNet accurately classify if the baby needs phototherapy 24 hours after birth?

Confusion Matrix

Bilirubin test outcome	**Neonates with jaundice**	
	Condition Positive	Condition Negative
Test outcome positive	True positive (TP) = 47	False positive (FP) = 31
Test outcome negative	False negative (FN) = 0	True negative (TN) = 96
	Sensitivity = TP / (TP + FN) = 100%	Specificity = TN / (FP + TN) = 75%

ROC Curve



Greater area under curve (AUC) means better classifier

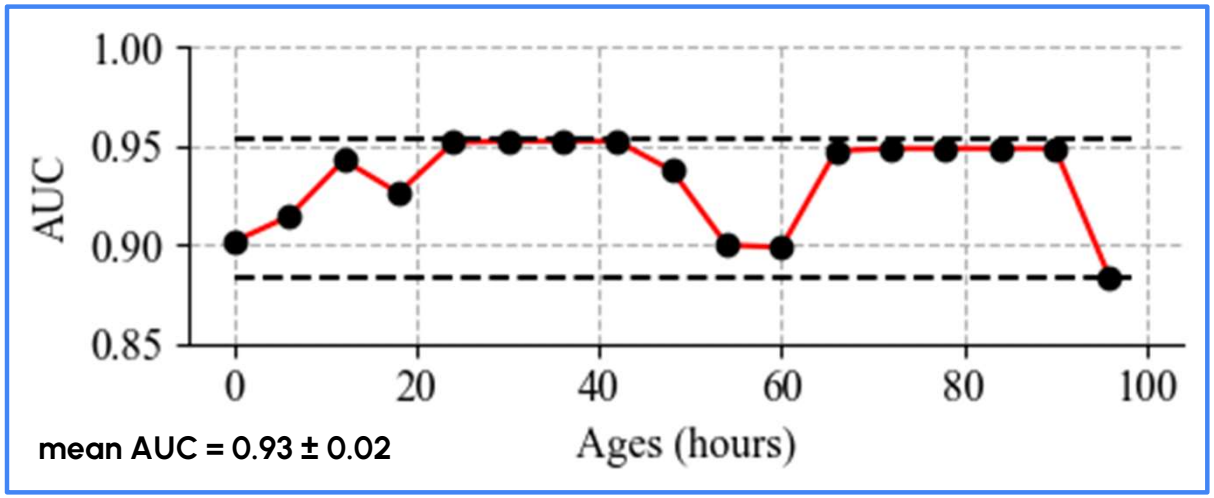
A: Yes, BiliNet correctly classifies neonates with 100% sensitivity and 75% specificity.

Area under curve is 95%

Phototherapy thresholds from National Institute of Health

Clinical Efficacy: Phototherapy Diagnosis

Question: Can BiliNet accurately classify whether the baby needs phototherapy for all 100 hours after birth?

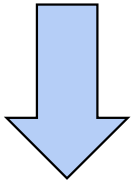


A: Bilinet has 93% average "accuracy" (AUC) for phototherapy classification across all 100 hours

Phototherapy thresholds from National Institute of Health

Overall Performance Relative to Other Approaches

Study	Method	Pearson R	Sensitivity	Specificity	Area under ROC Curve
Engle et al. 2005	TcB: JM-103	0.77	-	-	-
Romangnoli et al. 2012	TcB: BiliChek	0.82	0.99	0.30	0.89
BiliNet	**Blood plasma test strips**	**0.83**	**1.00**	**0.75**	**0.95**
Outlaw et al. 2020	Sclera capture	0.75	1.00	0.54	0.85
Leung et al. 2016	Sclera capture	0.72	1.00	0.50	0.87
Swarna et al. 2018	Sternum capture	0.60	-	-	-



BiliNet outperforms TcB AND sternum/scelera approaches across all metrics

End-to-End System Architecture

Inference Module

On Device

Data Ingestion subsystem

Image Reader

Preprocessing

Trained Model Instances

Segmentation Prediction Service

<10 second inference

Regression Prediction Service

TSB

Error handling and guardrails

Quality Evaluation subsystem

End-User Error Handling subsystem

Test Strip Error: cannot read test strip	Range Error: TSB out of range (0 - 30 mg/dL)
Camera Access Error: cannot access camera	Error: test failed; try again

Training Pipeline

Data Pipeline

Blob Storage

Validate

Standardize

Curate

Feature Engineering

RGB Feature Extraction

NumPy

pandas

matplotlib

Dual Model Training and Deployment

Unit Testing

Validate

Train

Code

Train time: 45 min

scikit learn

TensorFlow

Keras

Conclusion

1. **Performance:**
   - Segmentation model **enables reading of blood plasma test strips in 10 seconds** compared to 24-hour laboratory blood tests → does not require medical infrastructure
   - **Adapts to test strip misalignment**, blood leakage, and irregular lighting conditions
   - Extracted blood features effectively capture yellow-blue axis of bilirubin concentration
   - **Stronger 83% correlation with gold standard** compared to other solutions → regression model found non-linear relationships in extracted features
   - Predictions have **2.38 mg/dL error** and **low rate of outliers**

Conclusion

1. **Performance:**
   - Segmentation model **enables reading of blood plasma test strips in 10 seconds** compared to 24-hour laboratory blood tests → does not require medical infrastructure
   - **Adapts to test strip misalignment,** blood leakage, and irregular lighting conditions
   - Extracted blood features effectively capture yellow-blue axis of bilirubin concentration
   - **Stronger 83% correlation with gold standard** compared to other solutions → regression model found non-linear relationships in extracted features
   - Predictions have **2.38 mg/dL error** and **low rate of outliers**

2. **Accessible, Scalable, and Easy-to-Use:**
   - Use of blood means BiliNet completely **eliminates skin tone bias**
   - **Mobile application scales globally**; accessible in high incidence rate regions such as sub-Saharan Africa
   - **Minimal participation from the newborn** which is critical in fast-paced clinical environments

Conclusion

1. **Performance:**
   - Segmentation model **enables reading of blood plasma test strips in 10 seconds** compared to 24-hour laboratory blood tests → does not require medical infrastructure
   - **Adapts to test strip misalignment,** blood leakage, and irregular lighting conditions
   - Extracted blood features effectively capture yellow-blue axis of bilirubin concentration
   - **Stronger 83% correlation with gold standard** compared to other solutions → regression model found non-linear relationships in extracted features
   - Predictions have **2.38 mg/dL error** and **low rate of outliers**

2. **Accessible, Scalable, and Easy-to-Use:**
   - Use of blood means BiliNet completely **eliminates skin tone bias**
   - **Mobile application scales globally**; accessible in high incidence rate regions such as sub-Saharan Africa
   - **Minimal participation from the newborn** which is critical in fast-paced clinical environments

3. **Low-Cost:**
   - Complete system is cost effective, **reducing barrier for care from $7,000 for TcB to $1 for test strip**

Conclusion

1. **Performance:**
   - Segmentation model **enables reading of blood plasma test strips in 10 seconds** compared to 24-hour laboratory blood tests → does not require medical infrastructure
   - **Adapts to test strip misalignment,** blood leakage, and irregular lighting conditions
   - Extracted blood features effectively capture yellow-blue axis of bilirubin concentration
   - **Stronger 83% correlation with gold standard** compared to other solutions → regression model found non-linear relationships in extracted features
   - Predictions have **2.38 mg/dL error** and **low rate of outliers**

2. **Accessible, Scalable, and Easy-to-Use:**
   - Use of blood means BiliNet completely **eliminates skin tone bias**
   - **Mobile application scales globally**; accessible in high incidence rate regions such as sub-Saharan Africa
   - **Minimal participation from the newborn** which is critical in fast-paced clinical environments

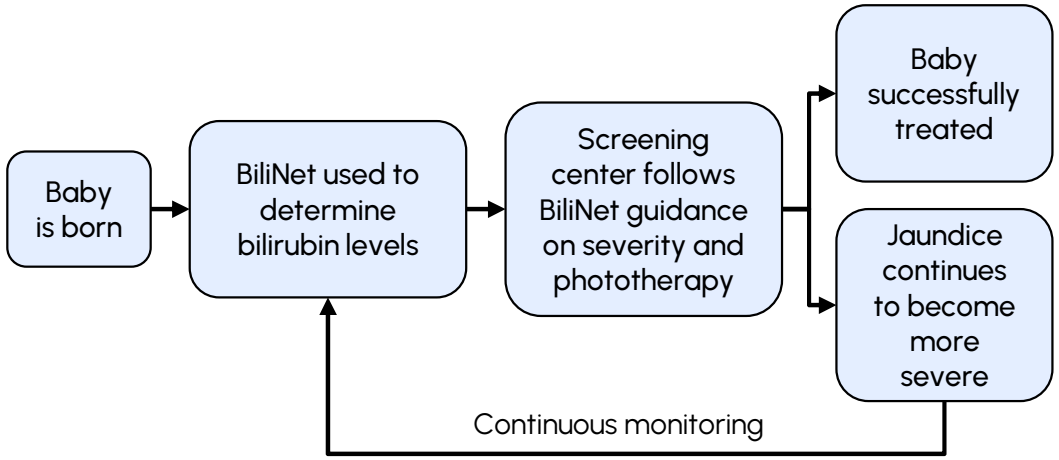
3. **Low-Cost:**
   - Complete system is cost effective, **reducing barrier for care from $7,000 for TcB to $1 for test strip**

4. **Clinical Efficacy:**
   - 95% accuracy for phototherapy classification means **BiliNet is viable as a tool for universal screening operations IRL**

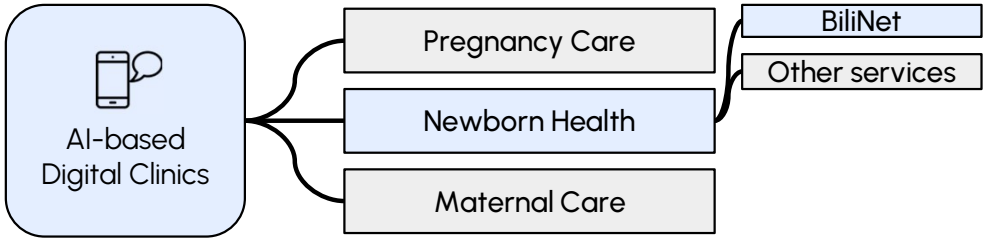
Hypothesis was proved: machine learning model was able to find relationships in extracted features useful for bilirubin prediction at low financial and temporal cost.

Discussion and Future Work

Real World Use of BiliNet



Integration with Digital Health Service



Limitations

- Accuracy tapers at high TSB → **can be solved through Bayesian inference**

Clinical Trials

- **Pursue stringent clinical trials** with the Doctors for You NGO in **India**
- Scale: **9 hospitals, 4 states, 500 births per month**
- + partner universities in Nigeria and Egypt

Digital Health Services

- **Integrate BiliNet with AI SMS services** for contextual maternal and neonatal health support in low-resource settings
- eg. Kenyan-based Jacaranda Health (2.4 million mothers connected)

References

[1] C. G. Scrafford, L. C. Mullany, J. Katz, S. K. Khatry, S. C. LeClerq, G. L. Darmstadt, and J. M. Tielsch, "Incidence of and risk factors for neonatal jaundice among newborns in southern Nepal," Tropical Medicine & International Health, vol. 18, no. 11, pp. 1317–1328, Nov. 2013.

[2] M. Johnston and R. R. Ravindran, "Jaundice," Surgery (Oxford), vol. 41, no. 6, pp. 334–341, Jun. 2023.

[3] A. W. Khan, P. Bhatt, P. Y. Yagnik, M. Ayensu, N. A. Adjetey, A. A. Agyekum, N. S. Bhatt, K. Donda, and F. M. Dapaah-Siakwan, "Trends in Hospitalization for Neonatal Jaundice and Kernicterus in the United States, 2006-2017," Pediatrics, vol. 147, no. 3 Meeting Abstract, pp 744–745, Mar. 2021.

[4] G. G. Asefa, T. G. Gebrewahid, H. Nuguse, M. W. Gebremichael, M. Birhane, K. Zereabruk, T. M. Zemicheal, A. Hailay, W. A. Abrha, S. A. Hadera, A. G. Hailu, B. H. Beyene, E. A. Dagnazgi, F. G. Tekulu, and F. Welay, "Determinants of Neonatal Jaundice among Neonates Admitted to Neonatal Intensive Care Unit in Public General Hospitals of Central Zone, Tigray, Northern Ethiopia, 2019: a Case-Control Study," BioMed Research International, vol. 2020, pp. 1–8, Oct. 2020.

[5] T. M. Slusher, T. G. Zamora, D. Appiah, J. U. Stanke, M. A. Strand, B. W. Lee, S. B. Richardson, E. M. Keating, A. M. Siddappa, and B. O. Olusanya, "Burden of severe neonatal jaundice: a systematic review and meta-analysis," BMJ Paediatrics Open, vol. 1, no. 1, p. e000105, Nov. 2017.

[6] M. J. Aminoff, Encyclopedia of the neurological sciences, 2nd ed. Amsterdam: Academic Press, 2014.

[7] M. Donneborg, K. Knudsen, and F. Ebbesen, "Effect of infants' position on serum bilirubin level during conventional phototherapy," ActaPaediatrica, vol. 99, no. 8, pp. 1131–1134, 2010.

[8] A. Wan, S. Mat Daud, S. H. Teh, Y. M. Choo, and F. M. Kutty "Management of neonatal jaundice in primary care," Malaysian Family Physician: The Official Journal of the Academy of Family Physicians of Malaysia, vol. 11, no. 2-3, pp. 16–19, 2016.

[9] K. M. Satrom, Z. L. Farouk, and T. M. Slusher, "Management challenges in the treatment of severe hyperbilirubinemia in low- and middle-income countries: Encouraging advancements, remaining gaps, and future opportunities," Frontiers in Pediatrics, vol. 11, p. 1001141, Feb. 2023.

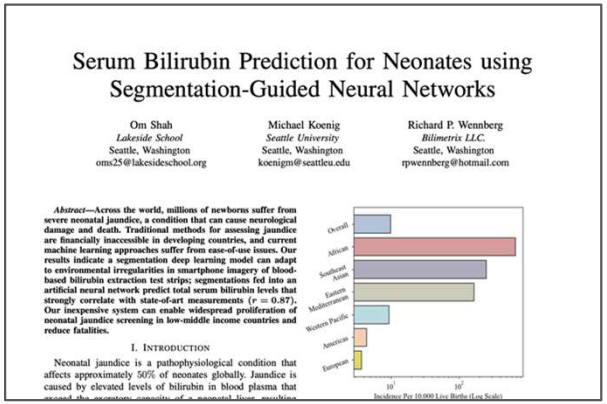
[10] C. I. Okwundu, A. Olowoyeye, O. A. Uthman, J. Smith, C. S. Wiysonge, V. K. Bhutani, M. Fiander, and K. S. Gautham, "Transcutaneous bilirubinometry versus total serum bilirubin measurement for newborns," Cochrane Database of Systematic Reviews, vol. 2023, no. 5, May 2023.

[11] L. I. Kramer, "Advancement of dermal icterus in the jaundiced newborn," American journal of diseases of children (1960), vol. 118, no. 3, p. 454, 1969, place: United States.

[12] S. Maya-Enero, J. Candel-Pau, J. Garcia-Garcia, X. Duran-Jord`a, and M. Lopez-V´ılchez, "Reliability of transcutaneous bilirubin determination based on skin color determined by a neonatal skin color scale of our own," European Journal of Pediatrics, vol. 180, no. 2, pp. 607–616, Feb. 2021.

[13] S. Samiee-Zafarghandy, J. Feberova, K. Williams, A. S. Yasseen, S. L.Perkins, and B. Lemyre, "Influence of skin colour on diagnostic accuracy of the jaundice meter JM 103 in newborns," Archives of Disease in Childhood - Fetal and Neonatal Edition, vol. 99, no. 6, pp. F480–F484, Nov. 2014.

[14] Diala, Udochukwu M et al. "Global Prevalence of Severe Neonatal Jaundice among Hospital Admissions: A Systematic Review and Meta-Analysis." Journal of clinical medicine vol. 12,11 3738. 29 May. 2023.



Serum Bilirubin Prediction for Neonates using Segmentation-Guided Neural Networks

Om Shah
Lakeside School
Seattle, Washington
oms25@lakesideschool.org

Michael Koenig
Seattle University
Seattle, Washington
koenigm@seattleu.edu

Richard P. Wennberg
Bilimetrix LLC.
Seattle, Washington
rpwennberg@hotmail.com

***Abstract*—Across the world, millions of newborns suffer from severe neonatal jaundice, a condition that can cause neurological damage and death. Traditional methods for assessing jaundice are financially inaccessible in developing countries, and current machine learning approaches suffer from ease-of-use issues. Our results indicate a segmentation deep learning model can adapt to environmental irregularities in smartphone imagery of blood-based bilirubin extraction test strips; segmentations fed into an artificial neural network predict total serum bilirubin levels that strongly correlate with state-of-art measurements ($r = 0.87$). Our inexpensive system can enable widespread proliferation of neonatal jaundice screening in low-middle income countries and reduce fatalities.**

I. Introduction

Neonatal jaundice is a pathophysiological condition that affects approximately 50% of neonates globally. Jaundice is caused by elevated levels of bilirubin in blood plasma that

Thank you to Professor Koenig at Seattle University for guiding my research focus, data acquisition, and asking forward-thinking questions

Thank you to Dr. Wennberg at Bilimetrix for guiding final product considerations and assisting with clinical regulations.