భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Glucose Prediction from Laser Speckle Image Patterns

~ Sailee Prashant Allyadwar

1. Project Overview

Objective

Develop a machine learning model that predicts blood glucose levels (mg/dL) from images of biological samples (e.g., test strips, skin speckle patterns).

Approach

- Extract texture features (GLCM, LBP) from grayscale images.
- Train regression models (XGBoost, Random Forest, etc.) to predict glucose levels.
- Optimize hyperparameters for best performance.

Key Challenges

- Limited dataset size (~300 images).
- High prediction errors (MAE: 65–87 mg/dL).
- Weak correlation (R^2: 0.14–0.49).

---

2. Dataset

Structure

text
Copy
Download

```
F:\8.Glucometer Setup IITH\
├── train_data/
│   ├── 100 mg/dL/  # Folder per concentration
│   │   ├── image1.jpg
│   │   └── ...
│   └── 200 mg/dL/
│       └── ...
└── test_data/
    └── ...
```

- Classes: Glucose concentrations (e.g., 100, 150, 200 mg/dL).
- Images: Grayscale, resized to 256×256 px.

Preprocessing

1. Validation
   - Check for corrupt/blank images.
   - Remove uniform-color images.
2. Normalization
   - Resize to 256×256.
   - Contrast normalization (`cv2.NORM_MINMAX`).

---

3. Feature Extraction

Methods

Feature Type	Description	Parameters
GLCM	Gray-level co-occurrence matrix	Distances: [1, 3, 5], Angles: [0°, 45°, 90°, 135°]
LBP	Local Binary Patterns	Radius: [1, 3, 5], Points: 8×radius
Speckle Statistics	Mean, median, std, skewness, kurtosis	--
Histogram	Intensity distribution (16 bins)	Range: [0, 256]

---

1. Gray-Level Co-Occurrence Matrix (GLCM) Features

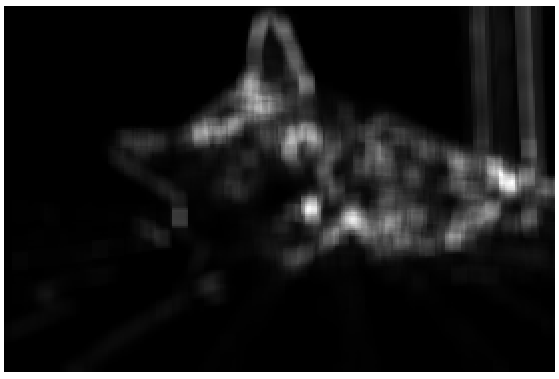
GLCM analyzes texture by calculating how often pairs of pixel intensities occur in an image at a given distance and angle.

Key Parameters

- Distance (d): 1, 3, 5 pixels
  Defines how far apart pixel pairs should be.
- Angles (θ): 0°, 45°, 90°, 135°
  Directions in which pixel pairs are analyzed.
- Properties Calculated:
  1. Contrast: Measures local intensity variations.
     - High contrast = Sharp edges/textures.
     - Formula: $$\sum_{i,j}(i-j)^2 \cdot P(i,j)$$
  2. Dissimilarity: Similar to contrast but linear (less sensitive to outliers).
     - Formula: $$\sum_{i,j}|i-j| \cdot P(i,j)$$
  3. Homogeneity: Measures closeness of GLCM values to the diagonal.
     - High homogeneity = Smooth textures.
     - Formula: $$\sum_{i,j}\frac{P(i,j)}{1+|i-j|}$$
  4. Energy (Angular Second Moment): Measures uniformity of pixel pairs.
     - High energy = Very similar pixels.
     - Formula: $$\sum_{i,j}P(i,j)^2$$
  5. Correlation: Measures linear dependency between pixels.
     - Formula: $$\sum_{i,j}\frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j}$$

Why GLCM?

- Captures texture patterns (e.g., smooth vs. grainy regions).
- Effective for biological samples where glucose affects speckle patterns.

Original: Output: 

---

2. Local Binary Patterns (LBP)

LBP encodes local texture by comparing each pixel to its neighbors.

Key Parameters

- Radius (R): 1, 3, 5 pixels
  Defines neighborhood size.
- Points (P): 8 × radius
  Number of neighboring pixels to compare.
- Method: 'uniform'
  Reduces feature dimensions by grouping similar patterns.

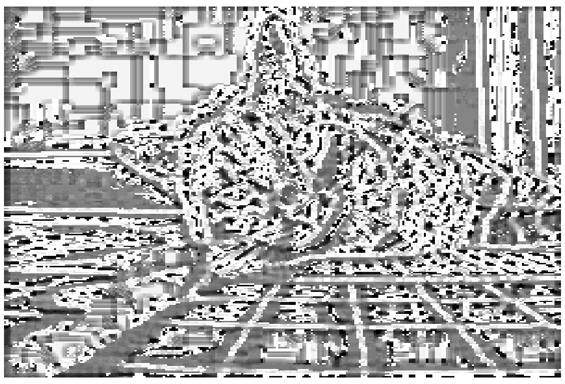
How It Works

$$\text{LBP} = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p$$

1. For each pixel, compare intensity with neighbors:
2. where $s(x)$= 1 if $x \geq 0$, else 0.
3. Compute histogram of LBP codes (bins = 0 to P+2).
4. Normalize histogram to sum to 1.

Why LBP?

- Rotation-invariant (uniform patterns).
- Captures micro-textures (e.g., speckle noise).

Original : Output: 

---

3. Speckle Statistics

Measures intensity distribution properties.

Features Extracted

1. Speckle Contrast

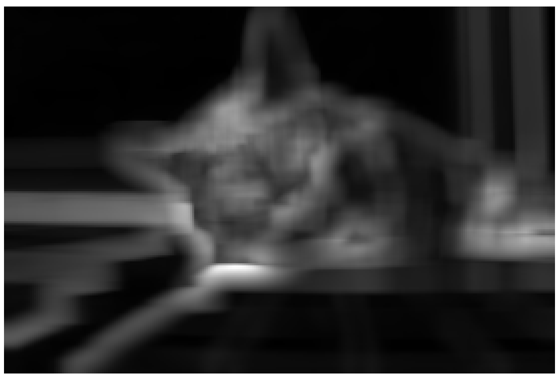
$$\frac{\sigma}{\mu} \quad \text{(Standard deviation / Mean)}$$

   - High contrast = More noise (e.g., higher glucose variability).
2. Intensity Mean/Median
   - Average/median pixel value.
3. Standard Deviation
   - Measures intensity spread.

4. Skewness
    - Asymmetry of intensity distribution.
    - Positive skew = More dark pixels.
5. Kurtosis
    - "Peakedness" of intensity distribution.
    - High kurtosis = Sharp intensity peaks.

Why Speckle Stats?

- Glucose changes may alter light scattering properties.

Original : Output: 

---

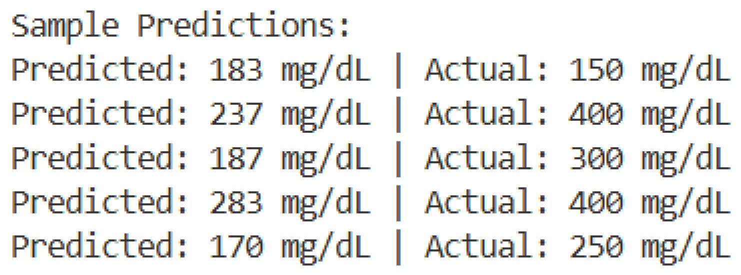
4. Models & Training

Algorithms Tested

Model	Tuning Method	Best MAE (mg/dL)	R²
XGBoost	Random Search	65.88 (Train)	0.4896

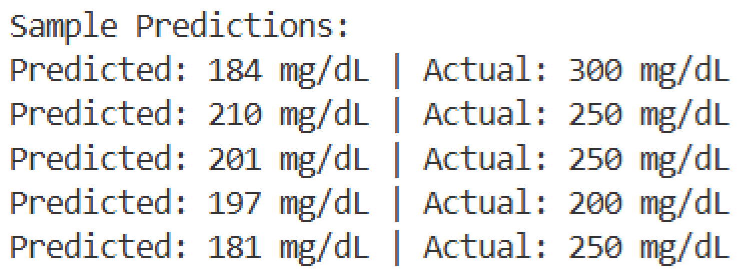
Random Forest	Grid Search	38.67 mg/dL(Train)
		87.96 mg/dL(Validation)
		82.20 mg/dL(Testing)
			0.8527(Testing)
			-0.0301(Validation)
			0.2255(Testing)

Random Forest Output:



```
Sample Predictions:
Predicted: 183 mg/dL | Actual: 150 mg/dL
Predicted: 237 mg/dL | Actual: 400 mg/dL
Predicted: 187 mg/dL | Actual: 300 mg/dL
Predicted: 283 mg/dL | Actual: 400 mg/dL
Predicted: 170 mg/dL | Actual: 250 mg/dL
```

XGBoost Output:



```
Sample Predictions:
Predicted: 184 mg/dL | Actual: 300 mg/dL
Predicted: 210 mg/dL | Actual: 250 mg/dL
Predicted: 201 mg/dL | Actual: 250 mg/dL
Predicted: 197 mg/dL | Actual: 200 mg/dL
Predicted: 181 mg/dL | Actual: 250 mg/dL
```

Hyperparameter Tuning (XGBoost)

python
Download

```
Best Parameters:
 - learning_rate: 0.014
 - max_depth: 3
- n_estimators: 109
- subsample: 0.825
- colsample_bytree: 0.872
```

Evaluation Metrics

Split	MAE (mg/dL)	RMSE (mg/dL)	R²
Training	65.88	79.32	0.4896
Validation	87.31	109.13	0.1388
Test	84.08	102.58	0.1725

---

5. Performance Analysis

Key Issues

1. High Error Rates
   - MAE > 65 mg/dL is clinically unacceptable (target: <15 mg/dL).
   - Poor performance on extremes (e.g., predicted 275 vs. actual 100 mg/dL).
2. Low R²
   - Model explains <50% of variance.
3. Overfitting
   - Training MAE (65.88) << Validation MAE (87.31).

Sample Predictions

Actual (mg/dL)	Predicted (mg/dL)	Error (%)
150	193	+28.7%
100	275	+175%
250	182	-27.2%

---

6. Recommendations for Improvement

Data-Level

1. Increase Dataset Size
   - Collect 1,000+ images across glucose ranges.
2. Augmentation
   - Rotate, flip, adjust contrast/blur to synthetically expand data.
3. Class Balancing
   - Ensure equal samples per glucose level.

Model-Level

1. Try Simpler Models
   - Linear Regression (baseline) or SVR with RBF kernel.
2. Adjust XGBoost
   - Increase `learning_rate` (0.05–0.2).
   - Use early stopping.
3. Feature Engineering
   - Add color features (if RGB images available).
   - Experiment with CNN-based feature extraction.

Evaluation

1. Clinical Accuracy
    - Report % predictions within ±15% of actual values.
2. Error Analysis
    - Visualize residuals vs. glucose levels.

---

7. Code Structure

Key Files

File	Purpose
`python6.py`	Main training/prediction script
`glucose_model.pkl`	Saved XGBoost model
`glucose_scaler.pkl`	Feature scaler

Usage

bash
Copy
Download

```bash
# Train model
python python6.py

# Predict single image
python predict.py --image test_image.png
```

---

8. Future Work

- Integrate deep learning (CNNs).
- Deploy as a mobile/web app.
- Validate with clinical trials.