Comparative DCT-feature classification of TUH breast images

Yusuf Qwareeq

Electrical and Computer Engineering Department
Temple University

May 5, 2025

Introduction
Task Overview

- Image classification task on TUH DPATH Breast data patches.
- Data provided as three-channel 32×32 DCT coefficients.
- Original data has nine classes.
- Task: Classify into a six-class structure based on scoring rules.
- Key Metric: Minimize 90/10 weighted-error rate.

LightGBM: why a baseline?

Motivation

- Establish a ground-truth yardstick before GPU-heavy NN runs.
- Raw DCT coefficients are naturally tabular $\rightarrow$ LightGBM suitability.
- Provides a cost-effective environment to test FE and imbalance handling.

LightGBM: data preparation

Phase 0

- We began with the original `train.csv` and `dev.csv` datasets.
- These files initially contained data categorized into nine distinct classes.

Adaptation

- Based on the task's scoring rules, we excluded data from classes 1, 4, and 7.
- The remaining six classes (0, 2, 3, 5, 6, 8) became our focus, and we assigned them simpler IDs from 0 to 5.

- This process yielded the final dataset, structured specifically for training and evaluating our six-class problem.

LightGBM: tuning strategy

Phase 1

Approach

- Split the six-class dataset into `train` + `validation` pool (85%) and a held-out `test` set (15%).
- Used Optuna with `optuna-distributed` for multi-threaded tuning on the `train` + `validation` pool.
- Evaluation within tuning: three-fold *stratified* cross-validation.
- Objective: Minimize average 90/10 weighted-error across CV folds.

LightGBM: tuning phases and results

Phase 1

Phase 1a – feature engineering

- Tuned DCT block sizes and PCA variance (200 trials).
- Kept the three DC coefficients.
- Best: $k = (3, 4, 2)$, $p \approx 99.95\%$.
- **Score ↓ from 62.92% to 50.03%.**

Phase 1b – model parameters (500 trials)

- Tuned `num_leaves`, `max_depth`, etc.
- Best: `num_leaves`=24, `max_depth`=4, ...
- **Score ↓ to 43.15%.**

Phase 1c – class weights (300 trials)

- Tuned individual w_k for each of the six remapped classes.
- Best: $w^\star \approx [4.1, 2.6, 36.9, 36.5, 19.8, 1.8]$.
- **Score ↓ to 42.44%.**

LightGBM: repeated unbiased testing

Phase 2

Process

- Repeated 10 times with different random `train/validation/test` splits.
- Preprocessing fitted on the `train` split, applied to all splits.
- Model trained on combined `train` + `validation` data (using best tuned HPs).
- Evaluated only once on the held-out `test` set for each repetition.
- Collected 90/10 weighted-error for each repetition's `test` set.
- Resulted in a mean weighted-error of $\mathbf{41.87\%} \pm \mathbf{1.56\%}$.

LightGBM: prediction

Phase 3

Process

- Use `train` for fitting and `dev` for early stopping.
- Feature pipeline (DC picks + per-channel $k \times k$ DCT + PCA) is fitted on `train` and reused everywhere.
- Train the tuned model; stop early if the `dev` weighted-error does not improve for 50 rounds.
- Apply the same pipeline to the original `train`, `dev`, and `eval` CSV files and write predictions.
- Resulted in a weighted-error of **17.41%** on the `training` set and **45.06%** on the `validation` set.

LightGBM: take-aways

- Optuna tuning significantly reduced the 90/10 weighted-error from $\approx 63\%$ to $\approx 42\%$ (-21% relative reduction).
- Model still shows high error rates on specific classes.
- Provides a solid classical baseline performance for comparison with neural network approaches.

Neural networks: why leave "tabular" land?

Motivation

- Inverse-DCT of the 32×32 blocks restores real spatial context – CNNs/ViTs can exploit that.
- Potential to learn texture cues that hand-engineered DCT-subsets miss.
- Goal: beat the baseline or, at worst, offer a complementary view for an ensemble.

Neural networks: many roads traveled

The “graveyard”

Architectures prototyped

- Plain CNNs (ResNet-18/34/50) on 256×256 IDCT images – under-fit, best `dev` WE $\approx 50\%$.
- Frequency-domain CNN directly on 32×32 DCT cubes – quick but plateaued at `dev` WE $\approx 55\%$.
- Dual-stream hybrid (DCT CNN + IDCT CNN, late fusion) – compute heavy, no `dev` gain.
- Tiny/Small ViTs (ViT-Ti, ViT-S) – faster but accuracy similar to ResNet.
- ViT-B/16 + layer-wise LR decay $\rightarrow$ clearly strongest; baseline of $\approx 33\%$.

ViT-B/16: tuning strategy

Phase 1

Search space (coarse and fine; 30 trials each)

- Learning rate: $3 \times 10^{-5} - 5 \times 10^{-4}$.
- Weight decay: $10^{-3} - 10^{-1}$.
- Layer-wise decay: $0.5 - 0.85$.
- Drop-path: $0 - 0.20$.

Early stopping in the objective

- Patience $= 8$ epochs on a held-out 15% `validation` subset.
- Pruned with Optuna's `MedianPruner`.

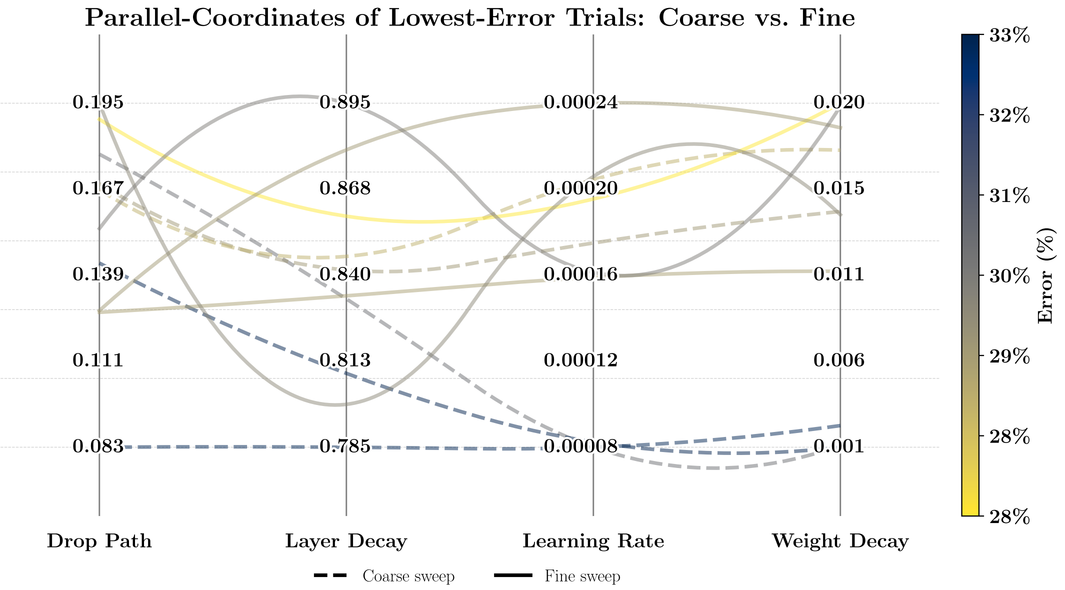
Best trial (coarse sweep)

Score $\downarrow$ from 33.05% to 28.87%.

Best trial (fine sweep)

$$\begin{aligned} \texttt{lr} &= 1.98 \times 10^{-4} \\ \texttt{weight\_decay} &= 2.03 \times 10^{-2} \\ \texttt{layer\_decay} &= 0.859 \\ \texttt{drop\_path} &= 0.190 \end{aligned}$$

Score $\downarrow$ from 28.87% to 27.56%.



Parallel-Coordinates of Lowest-Error Trials: Coarse vs. Fine
0.195
0.167
0.139
0.111
0.083
0.895
0.868
0.840
0.813
0.785
0.00024
0.00020
0.00016
0.00012
0.00008
0.020
0.015
0.011
0.006
0.001
Drop Path
Layer Decay
Learning Rate
Weight Decay
Coarse sweep
Fine sweep
33%
32%
31%
30%
29%
28%
28%
Error (%)

ViT-B/16: repeated unbiased testing

Phase 2

Process

- Repeated 10 times with different random `train`/`validation`/`test` splits.
- Early-stopped on the 15% slice, evaluated once on the fixed independent test set.
- Model trained on combined `train` + `validation` data (using best tuned HPs).
- Evaluated only once on the held-out `test` set for each repetition.
- Resulted in a mean weighted-error of $\mathbf{31.85\%} \pm \mathbf{1.45\%}$.

ViT-B/16: prediction

Phase 3

Process

- Reconstruct 256×256 RGB patches with IDCT, scale to 224×224, ImageNet normalization.
- Train on `train` data, stop if `dev` WE no longer improves for two epochs.
- Fine-tune for one epoch on `train` + `dev` with LR $\times\, 0.1$.
- Save best checkpoint and predict `train`, `dev`, and `eval`.
- Resulted in a weighted-error of **14.84%** on the `training` set and **28.80%** on the `validation` set.

Neural networks: take-aways

- Vision Transformer beats all CNN variants we tried and surpasses the tuned LightGBM baseline.
- Data volume (12k images) is small for ViT; model starts to memorize noise even with heavy regularization.