

Skin-Lesion Classification on DERM7PT

A Multimodal Machine Learning Approach to Early Melanoma Detection

Course: F21DL – Data Mining & Machine Learning

H00505014 Anam Ayyub

H00498947 Ayesha Syed Mohammad

H00519518 Nandini Petli

H00363811 Prianshu Rajput

H00521656 Yogesh Lachman Wadhwani

GitHub Repo:[https://github.com/anamayyub/GROUP_8]

Introduction & Literature Review

Prevelance

- Skin cancer is a common cancer.
- Early detection saves lives.
- Difficult to diagnose clinically.

Literature Review

- Kawahara et al.:
[<https://www.cs.sfu.ca/~hamarneh/ecopy/ibhi2018.pdf>]
- Arshad et al.[<https://rdcu.be/eJ0Fe>]
- Nawaz et al. [<https://rdcu.be/eJ0Fe>]

Derm7PT

Combines visual dermoscopic features with clinical metadata to leverage complementary diagnostic information sources.

Source: <https://derm.cs.sfu.ca/>

License: CC BY-NC-ND: Creative Commons Attribution NonCommercial NoDerivatives 4.0 International

Derm7PT includes:

- meta-Data(Patient & lesion)
- Images
- 7Pt checklist
- Expert Diagnosis Labels

Derm7pt is suitable for multimodal learning, as it provides both image and structured metadata inputs.

Exploratory Data Analysis/Data Preparation

Tabular Data [meta.csv)

- Severe Class Imbalance
- Heterogeneous Meta-data

The imbalance and heterogeneity motivated the use of stratified splits, macro-averaged metrics, data augmentation for imaging models, and class weighting and SMOTE for classical machine-learning baselines.

Images

- Clinical Images(1101) 512–768 px (w), 480–532 px (h)]
- Dermoscopic Images(1 missing)

Preprocessing for MetaData and Images

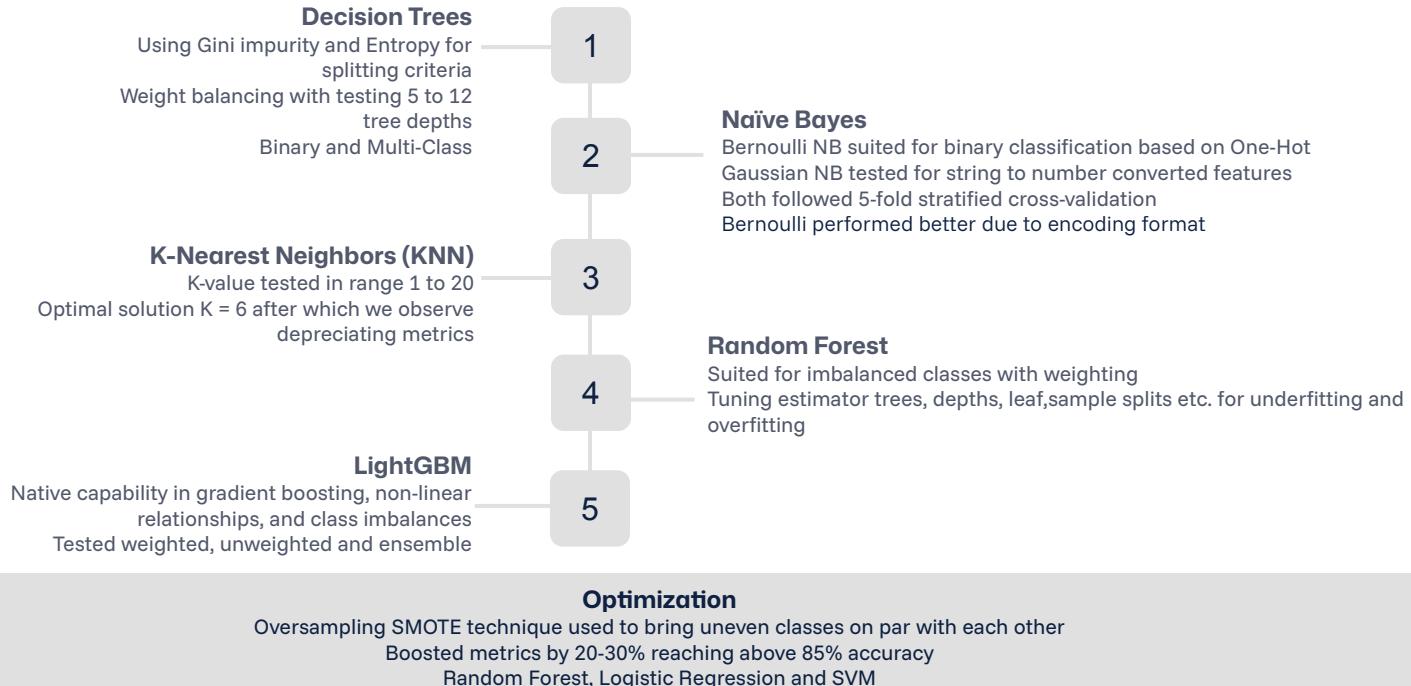
- Dropped Columns (case_num, case_id, notes) (missingness, irrelevance)
- Label Encoding (Blue-whitish-veil, Sex)
- Ordinal Encoding(level-of-diagnostic-difficulty)
- one-hot Encoding (all remaining categorical variables

- Images resized to 224×224 px to standardise input
- Loaded as NumPy arrays(0-255)
- Data augmentation used (random crops, flips, colour jitter) to boost robustness.

The final metadata matrix consisted of 65 encoded features.

*All the relevant graphs are available in
[01_EDA.ipynb](#)

Baseline ML Models



Conclusion - Baseline Models

Model	Accuracy	Macro F1	Weighted F1	Precision	Recall	Other
BernoulliNB	0.5924	0.44	0.60	0.44	0.45	Std = 0.0237
GaussianNB	0.2226	0.25	0.20	0.29	0.36	Std = 0.0339
KNN (k=6)	0.6100	0.5763	0.58	0.5918	0.6080	—
Random Forest	0.5400	0.41	0.56	0.43	0.45	—
LightGBM (unweighted)	0.6127	0.3658	0.5871	0.6072	0.6127	AUC = 0.8967
LightGBM (weighted)	0.5291	0.3635	0.5602	0.6231	0.5291	AUC = 0.8989
Ensemble (DT+LGBM)	0.4658	0.3027	0.4998	0.5930	0.4658	AUC = 0.8963

Random Forest:

Original Data - Accuracy: 0.6533, F1: 0.6366

SMOTE Data - Accuracy: 0.9698, F1: 0.9677

Accuracy improvement: +0.3166

F1-Score improvement: +0.3311

Resulting metrics and evaluation show that SMOTE optimized Random Forest give us the best performance which exceed or outweigh the imbalanced weighting techniques. As comparing to models like the KNN, LightGBM and imbalanced Random Forest metrics, they perform to a decent point without optimization which is something worth considering as it reaches at a similar point for the metrics detailed in the research papers. Despite the weighting, optimization works best and improves the overall result immensely as to without optimization.

*All the relevant graphs are available in
[02_baselines.ipynb](#)

Image Classification with CNN & ResNet50

Image Classification Backbones:

- CNN: Lightweight, hierarchical feature learning
- ResNet50: 50-Layer with skip connections, superior pattern recognition

Multimodal Fusion - Core Innovation:

```
code: combined = Concatenate()([image_features, tabular_features])
      z = Dense(128, activation='relu')(combined)
      outputs = Dense(n_classes, activation='softmax')(z)
```

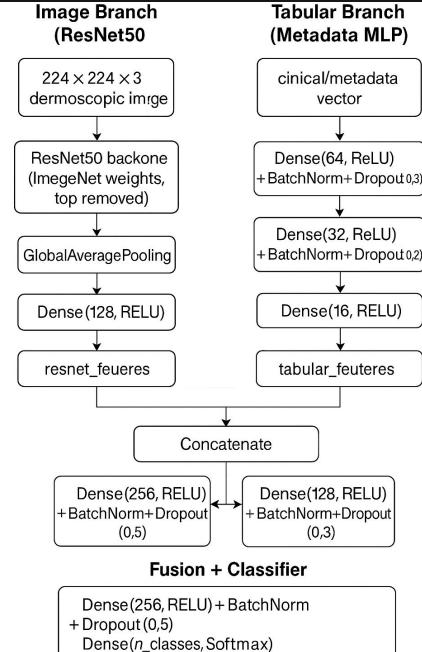
- integrates visual patterns with clinical data
- Enables context-aware diagnosis

Training Strategy:

- Class weighting for Imbalance
- Early stopping at optimal epochs
- Multimodal: Epoch 16 | ResNet50: Epoch 3

*All the relevant graphs are available in [Image_Implementation.ipynb](#)

Architecture for Fusion(Image and Tabular Data)



Performance & Key Insights

Model Comparison:

- Multimodal CNN accuracy: 48.3%
- Multimodal CNN AUC: 0.637
- Multimodal ResNet50: 38.9% accuracy, 0.531 AUC

Critical Findings:

- Fusion provides better clinical balance vs single modality(shown in the table below)
- SMOTE essential for realistic deployment
- Trade-off: Accuracy vs AUC for clinical utility

	Multimodal CNN	Multimodal ResNet50	CNN-image Only	ResNet50-Image Only	Tabular-Only
Accuracy	0.482758621	0.389162562	0.103448276	0.472906404	0.177339901
Precision	0.415758163	0.355229176	0.260986197	0.325297908	0.266255742
Recall	0.482758621	0.389162562	0.103448276	0.472906404	0.177339901
F1-Score	0.436834024	0.370383275	0.027414493	0.384835682	0.191121541
AUC-ROC	0.636546782	0.530973713	0.488538338	0.512146654	0.414272573