

# CPS 4801 AI Final Project: Skin Cancer Image Classification

Amir Zaman  
Dept. of Computer Science  
Kean University  
Union, United States  
zamanam@kean.edu

Ayan Desai  
Dept. of Computer Science  
Kean University  
Union, United States  
desaiay@kean.edu

Chigozie Ofodike  
Dept. of Computer Science  
Kean University  
Union, United States  
ofodikech@kean.edu

**Abstract**—This paper endeavors to cultivate expertise in AI programming, specifically within the Python PyTorch library domain, with a primary focus on crafting a discriminative model to discern benign from malignant skin cancer images. Employing classification methodologies, we delve into the implementation intricacies of decision tree and deep learning architectures. Despite decision trees offering expeditious and straightforward implementations, they yield models with diminished confidence and efficacy in image classification tasks. In contrast, the realm of deep learning, albeit posing greater complexity in implementation, facilitates nuanced tuning, culminating in superior performance outcomes. However, the models developed within this study do not yet meet the stringent criteria for practical deployment in real-world scenarios.

## I. INTRODUCTION

The importance of categorizing medical images, especially in skin cancer diagnosis, cannot be emphasized enough. It speeds up diagnostic procedures, ensuring quick and accurate prognosis, potentially reducing negative outcomes. Developing highly precise and effective image classification models is essential for various diagnostic methods.

Previous research has looked into combining segmentation and classification approaches in skin cancer detection. Some studies have shown improved performance by moving away from traditional supervised learning methods and embracing unsupervised techniques. The current challenge revolves around effectively sorting skin abnormalities while navigating through different model architectures, all while mastering Python coding techniques to improve computational efficiency [1].

This study aims to examine the effectiveness of two supervised learning approaches and compare their performance trends. The implications of the results go beyond academic discussion, serving as a foundational resource for aspiring AI practitioners starting their journey with Python and providing practical insights into medical image classification complexities.

## II. METHODS

We employ two different models for our analysis. The first model utilizes decision trees, requiring initial preprocessing steps such as image loading and flattening. Subsequently, we

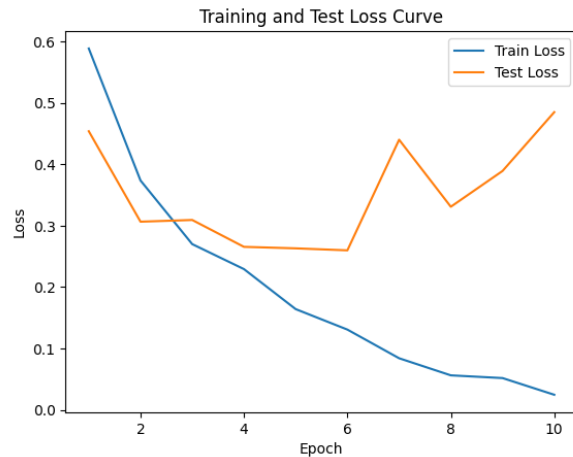


Fig. 1. Deep learning training loss curve over 10 epochs

conduct a thorough evaluation using precision, recall, and entropy metrics to assess the effectiveness of the model [2].

In contrast, the second approach utilizes deep learning and capitalizes on a pre-trained ResNet-50 architecture. Data ingestion is facilitated through a customized loader, and model evaluation mirrors that of the decision tree model [2].

For computational support, we utilize Google Colab, providing a highly scalable environment, coupled with a T4 GPU setup. We leverage various Python libraries including Torch, Timm, NumPy, Matplotlib, tqdm, PIL, and Scikit-learn, each playing a crucial role in both model development and evaluation processes.

## III. EXPERIMENT RESULTS

### A. Decision tree metrics:

Confusion Matrix:

[[271 70]

[ 88 231]]

Precision: 0.7674418604651163

Recall: 0.7241379310344828

Entropy: 0.9944220705380191

### B. Deep learning metrics:

Confusion Matrix:

[[305 14]

[ 75 266]]

Precision: 0.95

Recall: 0.7800586510263929

Average entropy: 0.1231112971547104

## IV. DISCUSSION

### A. Conclusive Remarks:

The decision tree instantiation exhibits commendable precision and recall metrics, albeit marred by elevated entropy levels, suggestive of inherent inconsistency within predictions. This can be caused by the feature space of the decision tree being so vast. Each pixel in the input images is counted as one feature so the deeper a tree has to go for a decision, the more uncertain that decision becomes. Conversely, the deep learning paradigm espouses heightened precision metrics, albeit at the expense of diminished recall, underpinned by a significantly lower entropy profile, denoting heightened predictive certitude. Therein lies the impetus for continued refinement and optimization of both models.

### B. Challenges Encountered:

Navigating through debugging quagmires, notably entailing rectifying disparate class counts impacting entropy computations, constituted a formidable challenge. Furthermore, computational constraints endemic to free Colab notebooks, typified by a finite 15GB GPU RAM allocation, engendered limitations on model training intensity, precipitating suboptimal performance outcomes.

### C. Future Endeavors:

Prospective endeavors are poised to pivot towards establishing a dedicated computational milieu, thereby obviating resource constraints inherent to cloud-based deployments. Exploratory forays into alternative model architectures, notably decision tree ensembles, coupled with the assimilation of advanced learning techniques like grouped model learning, augur well for enhancing performance metrics in subsequent iterations.

## REFERENCES

- [1] Barata, C., Rotemberg, V., Codella, N.C.F. *et al.* A reinforcement learning model for AI-based decision support in skin cancer. *Nat Med* 29, 1941–1946 (2023). <https://doi.org/10.1038/s41591-023-02475-5><https://doi.org/10.1038/s41591-023-02475-5>
- [2] Kwiatkowski, S. (2018, October 6). *Entropy is a measure of uncertainty*. Medium. <https://towardsdatascience.com/entropy-is-a-measure-of-uncertainty-e2c000301c2c><https://towardsdatascience.com/entropy-is-a-measure-of-uncertainty-e2c000301c2c>