# Classification of Histopathological Lung Cell Cancers Images using Deep Learning

## Project Report

Sahil Khose[1] Ankita Ghosh[2]

[1]3rd year, ICT Department
Manipal Institute of Technology

[2]3rd year, CSE Department
Manipal Institute of Technology

24 April 2021

# Table of Contents

# Scope of the problem

According to the American Cancer Society and the Cancer Statistics Center, over 150,000 patients with lung cancer succumb to the disease each year.

Classification of lung cancer type is a key diagnostic process because the available treatment options, including conventional chemotherapy and, more recently, targeted therapies, differ for LUAD and LUSC

# Classical Methods

- Combination of Second-Order Edge Detection:OLGA (Optimum Laplacian of Gaussian Assimilator) and Principle Component Analysis for Counting of Cancer Cell Nuclei in Tissue Sections. (Loukas et al, 2003)
- Conversion of RGB colorspace to HSV colorspace using color thresholding algorithm and generation of cell-web and cell-map using positional values for automated classification of inflamation in colon. (Ficsor et al, 2008)
- Automated segmentation of tissue images through variable color intensities and superposition. (Cataldo et al, 2010)
- Automated classification of breast cancer morphology using SVM classifiers. (Ojansivu et al, 2013)
- Tumor type prediction using image tiling, PCA clustering and deep feature extraction. (Barker et al, 2015)

# Architecture

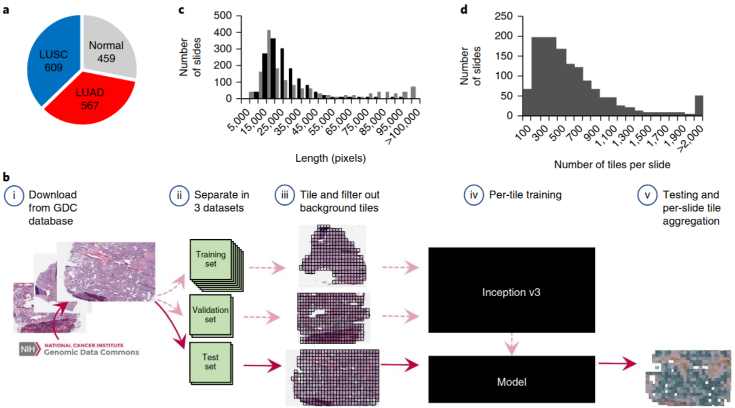



**Fig. 1 | Data and strategy. a**, Number of whole-slide images per class. **b**, Strategy for training. (**b**, **i**), Images of lung cancer tissues were first downloaded from the Genomic Data Commons database; (**b**, **ii**), slides were then separated into a training (70%), a validation (15%) and a test set (15%); (**b**, **iii**), slides were tiled by nonoverlapping 512- × 512-pixel windows, omitting those with over 50% background; (**b**, **iv**), the Inception v3 architecture was used and partially or fully retrained using the training and validation tiles; (**b**, **v**), classifications were performed on tiles from an independent test set, and the results were finally aggregated per slide to extract the heatmaps and the AUC statistics. **c**, Size distribution of the images widths (gray) and heights (black). **d**, Distribution of the number of tiles per slide.

Image from Coudray, N. et al.

# Data Pre-processing

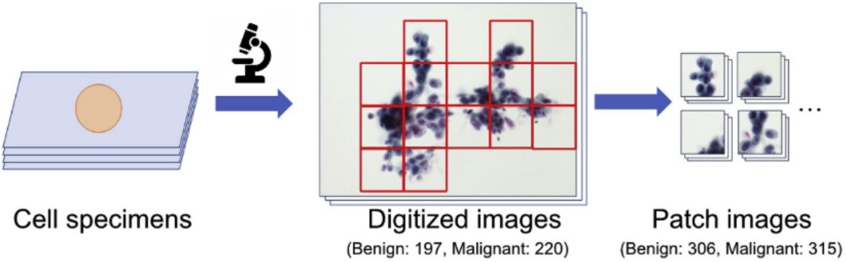



**Fig. 1.** Extraction of patches from a microscopic image.

Image from Teramoto, A. et al.

- tiles are cropped from the slides of the dataset
- tissue parts are extracted
- background is cleaned using thresholding before labeling the cropped images

# Data pre-processing

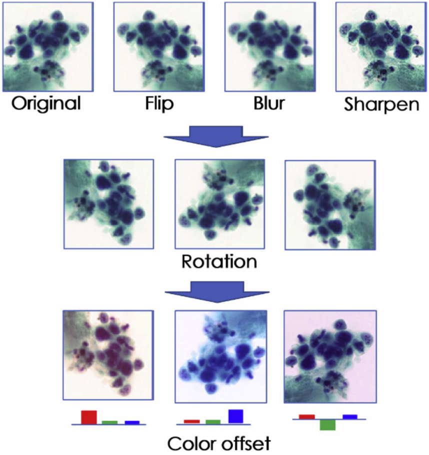



**Fig. 2.** Data augmentation methods to improve DCNN training.

Image from Teramoto, A. et al.

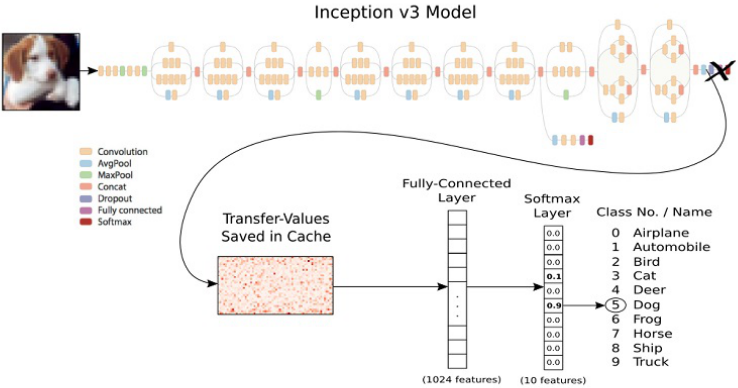
**Data Augmentation**

- rotation
- shifting
- morphological transforms
- colour offset
- use of filters

# Deep Learning Model

Three types of classification approaches:

- patch-based classification
- case-based classification
- distribution of malignant cells in microscopic images



# Evaluation Techniques

- **Sensitivity** $= \frac{TruePositive}{TruePositive+FalseNegative}$
- **Specificity** $= \frac{TrueNegative}{TrueNegative+FalsePositive}$
- **Accuracy** $= \frac{TruePositive+TrueNegative}{TruePositive+FalsePositive+TrueNegative+FalseNegative}$
- **AUC (Area Under The Curve)**: AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes.

# Concluding Discussion

Considering the advancements in the Deep Learning architectures in the recent years, the dataset can be trained on superior models (better accuracies, fewer parameters) to obtain improved results.

Data augmentation techniques can be used to increase the number of data points available for training larger Deep Learning models.

Further, the problem statement can be expanded to tasks such as segmentation, detection and localization.