University of California, Irvine

CHOC

Tarek El-Hajjaoui, Micah Fadrigo, Sheldon Gu, Cecilia Nguyen

Juvenile dermatomyositis (JDM)

- Rare autoimmune disease
- Weak muscle, skin rash …
- No cure → **help alleviate symptoms**

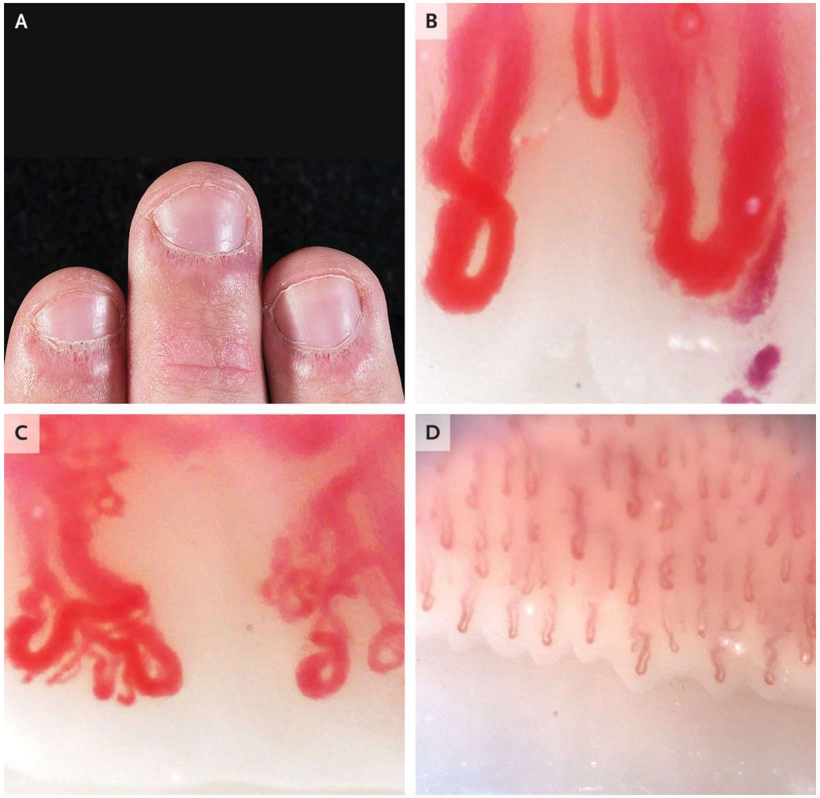
Diagnosis

Labs & Biomarker

- Expensive
- Difficult to identify

NFC

- Easily obtained
- Indicate disease activity



Nailfold Capillaroscopy (NFC)

Project Description

GOAL: simple, quick & inexpensive pre-screening on JDM

Specifically: differentiate JDM patients from healthy control groups

Data Preprocessing

- Load Images
- Interpolate Images
- HOG Feature Engineering

Machine Learning Models

- SVM
- CNN

Optimizing Data & Models

- HOG Features
- Support Vector Machines
- Convolutional Neural Network
- Image augmentation

Results & Discussion

- Overall Results
- Challenges
- Future Research

Data Description

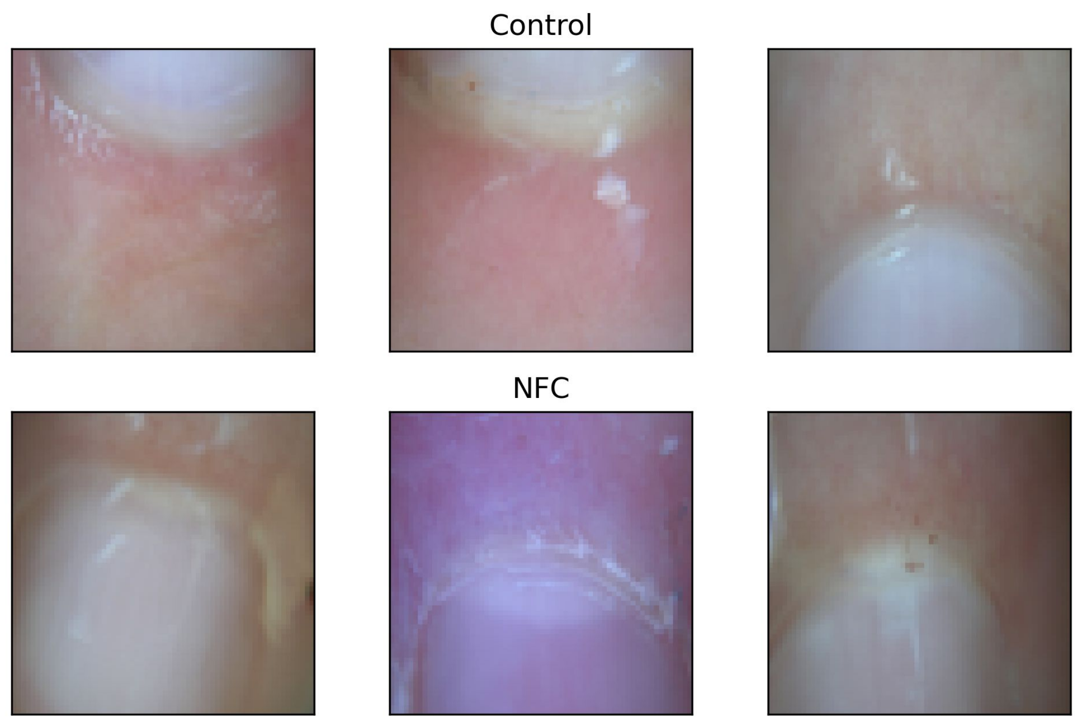
<PatientID + Finger>

Image Level:

- JDM: **1120** images
- Control: **321** images

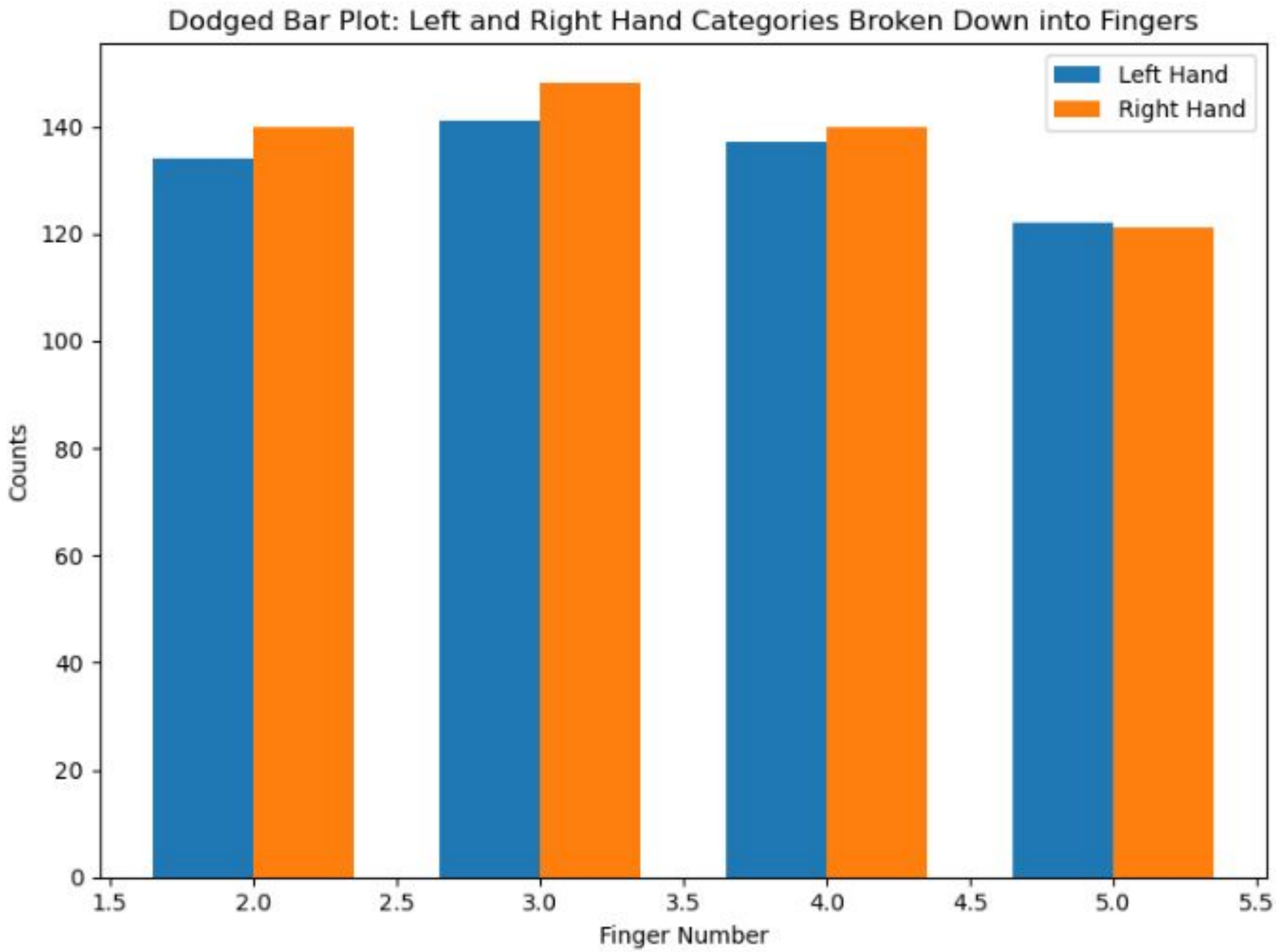
Patient Level:

- JDM: **111** patients
- Control: **31** patients

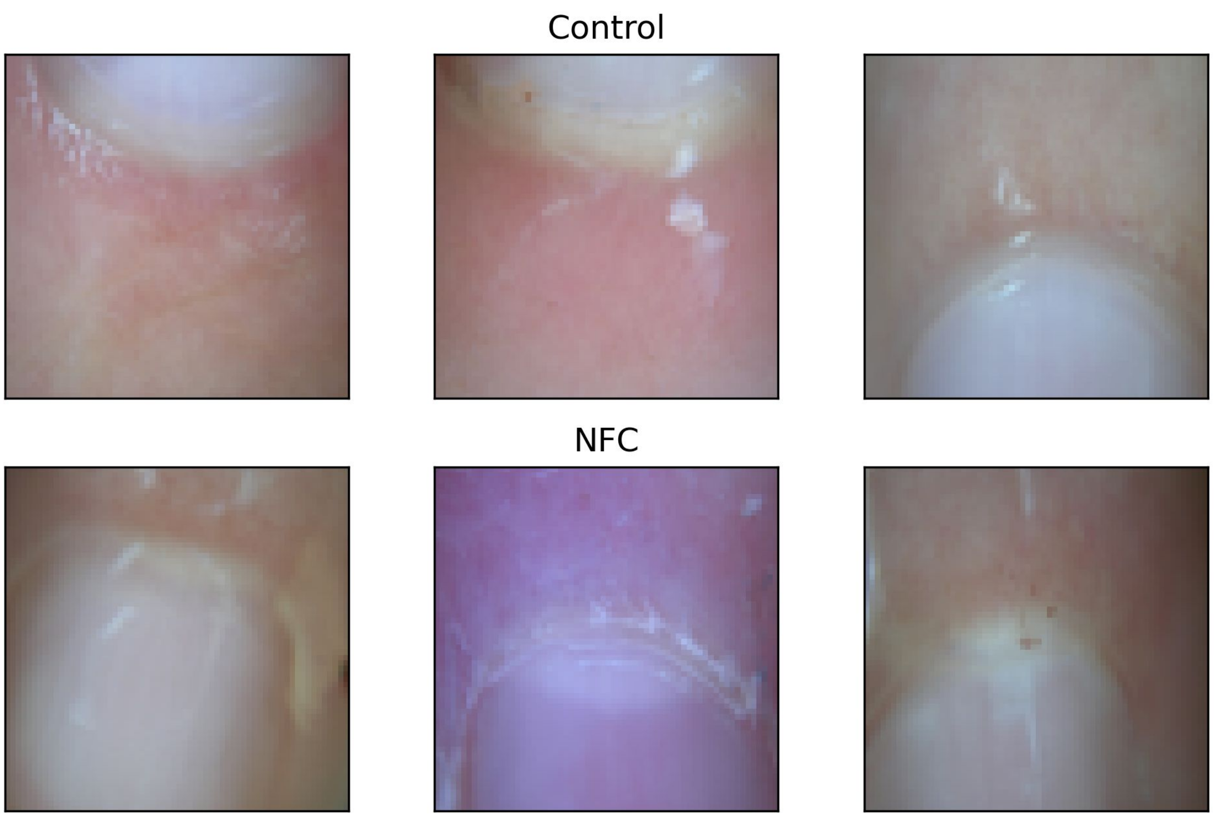


Response variable: JDM & Control

EDA


Dodged Bar Plot: Left and Right Hand Categories Broken Down into Fingers
Left Hand
Right Hand
Counts
Finger Number
0
20
40
60
80
100
120
140
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5

Data Issue Beyond Control



- **Unclear** (reflected lights, too dark)
- **Skin tone** varies

Data Issue Beyond Control

Obstructed Images

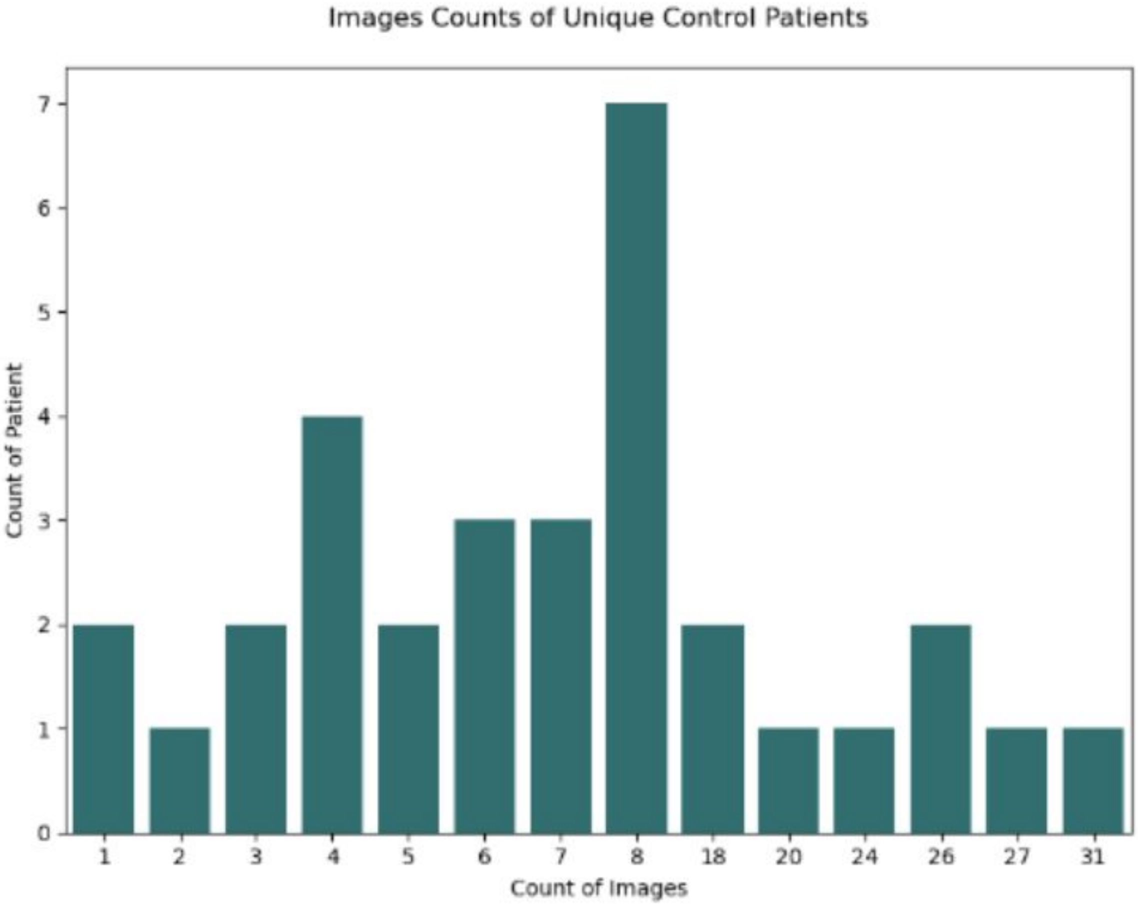
https://www.stylecraze.com/articles/8-simple-nail-art-desi
gns/

https://laurenbbeauty.com/blogs/blog/how-to-remove-nail-
polish-from-skin-around-nails

UCI

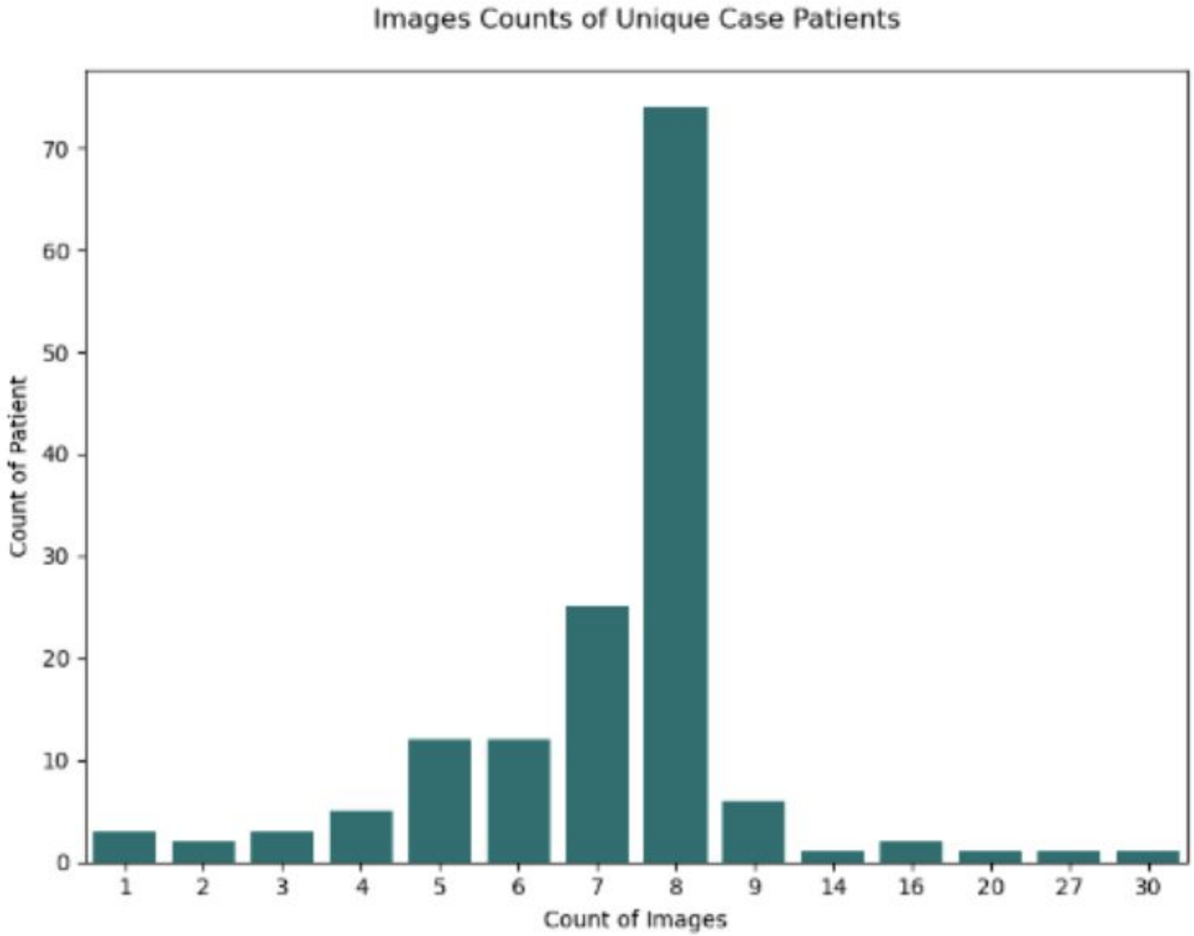
Data Issue Beyond Control

Imbalance: Not all patients represented equally between case/control and within.



Data Issue Beyond Control

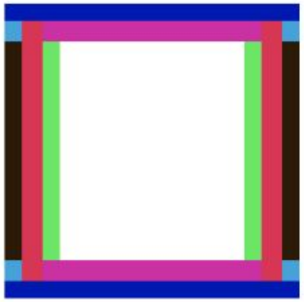
Imbalance: Not all patients represented equally between case/control and within.



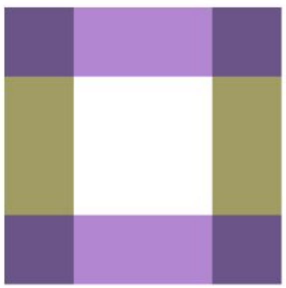
Data Issue - Solution

- **Large image size** → Interpolation
- **Absence of feature** → Histogram of Oriented Gradients (HOG)

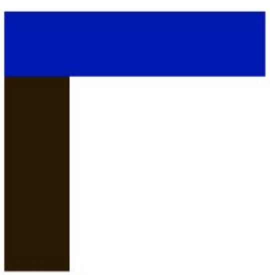
Data Preprocessing



initial image
[16px x 16px]



PIL.Image.resize
[4px x 4px]



tf.image.resize_bicubic
[4px x 4px]

CHOC: Manually examined and corrected orientation of NFCs.

Our Steps:

1. **Downscale** to size: 128, 64, 32
2. **Scale** input pixels between **(-1, 1)**
3. **Vectorize** images
4. **HOG** transformation
5. 10-Fold Stratified **Cross-Validation**

Histogram of Oriented Gradients (HOG)

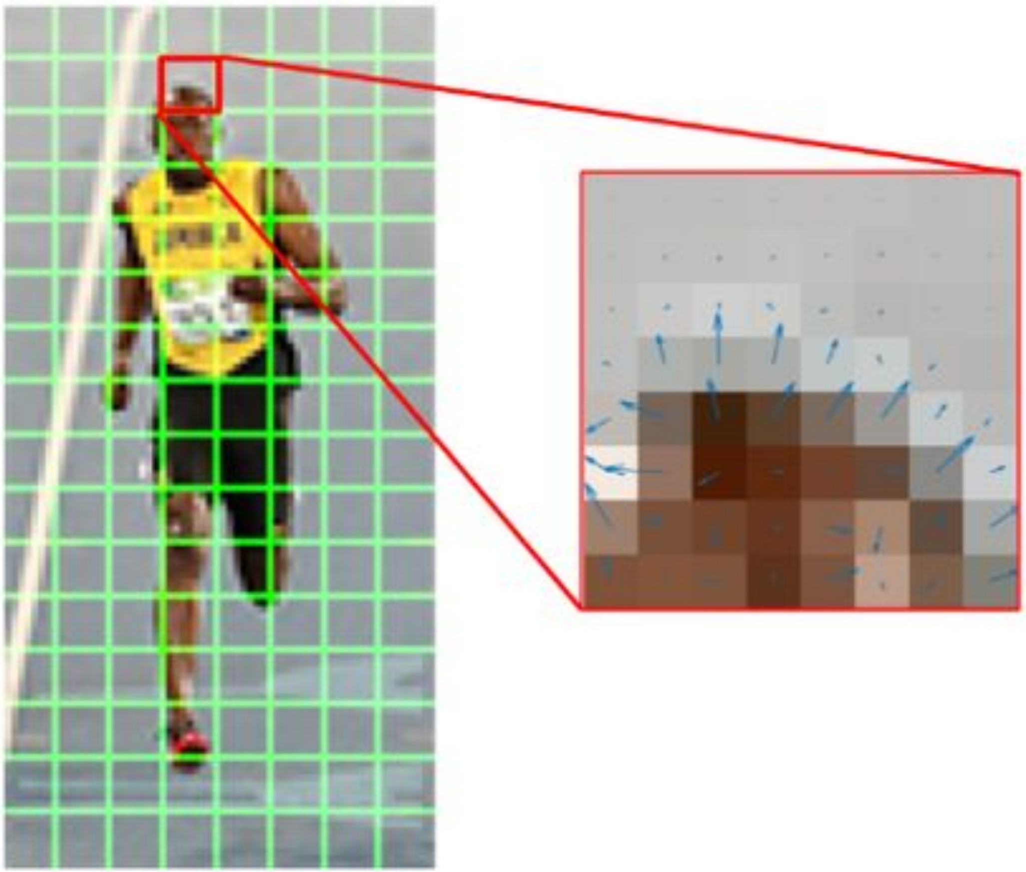
What is HOG

- Computer vision feature descriptor technique.
- Distribution of edge orientations.

Why is this useful

- Learn structural and spatial patterns of images.
- Reduces noise of images (for classification or object detection tasks).
- Generally preferred over vectorized images.

HOG Example



2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

Gradient Magnitude

80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

Gradient Direction

Center : The RGB patch and gradients represented using arrows.
Right : The gradients in the same patch represented as numbers

Data Preprocessing

- Load Images
- Interpolate Images
- HOG Feature Engineering

Machine Learning Models

- SVM
- CNN

Optimizing Data & Models

- HOG Features
- Support Vector Machines
- Convolutional Neural Network
- Image augmentation

Results & Discussion

- Overall Results
- Challenges
- Future Research

Convolution Neural Network

- **Widely used** for computer vision tasks

- Standard Architectures:
  - Batch normalization is **sensitive to large variation in the data**
  - **Uninterpretable**

- CHOC developed NFC-Net = **lightweight** CNN = 3 layers
  - Working on explainability

	Accuracy	Precision	Recall	F1 Score	ROC AUC	Specificity
NFC-Net	0.91	0.95	0.85	0.897	0.93	0.9

Overall Accuracy

Class 1 Accuracy

Class 0 Accuracy

Why Pursue Simpler Models?

- ★ Baseline Measurement & Reference
  - Are simple models able to achieve similar scores to NFC-Net?
- ★ Quicker Deployment to Mobile Devices
  - **Automate** clinical analyses of NFC
  - **Accelerate** JDM data collection & research
- ★ Robustness
  - Deals with high-level of noise

Why only focus on SVM for Simple Models

Logistic Regression + Lasso

❌ Assumptions violated

Random Forest

❌ Poor explainability
❌ Long training times

SVM

✅ Despite default hyper-parameters, SVM had better scores.
✅ RBF Kernel sensitivity to hyper-parameters.
✅ HOG + SVM is proven in Computer Vision tasks.
✅ Model explainability with Support Vectors.

Data Preprocessing

- Load Images
- Interpolate Images
- HOG Feature Engineering

Machine Learning Models

- SVM
- CNN

Optimizing Data & Models

- HOG Features
- Support Vector Machines
- Convolutional Neural Network
- Image augmentation

Results & Discussion

- Overall Results
- Challenges
- Future Research

HOG & SVM Tuning

Leveraged Scikit-Learn library to create a tuning framework:

HOG parameters → SVM parameters → Stratified CV

- Orientations
- Pixels per cell
- Cells per block

- C
- Class weight

- Refit based on ROC
- Calculate scores

Optimal Parameter Results

Achieved significantly better results:

ROC AUC	Class 0 Acc.	Precision	Accuracy	Recall	F1
0.920	0.726	0.923	0.896	0.945	0.934

Contradicts previous presentation's results.

➢ Highlights the importance of hyperparameter tuning.

SVM Comparison

Previous Best - Linear SVM [32x32] (Vectorized)					
ROC AUC	Class 0 Acc.	Precision	Accuracy	Recall	F1
0.756	0.624	0.892	0.830	0.890	0.890

New Best - RBF SVM [32x32] (HOG)					
ROC AUC	Class 0 Acc.	Precision	Accuracy	Recall	F1
0.920	0.726	0.923	0.896	0.945	0.934

SVM Improvements

Grid Search on different image sizes

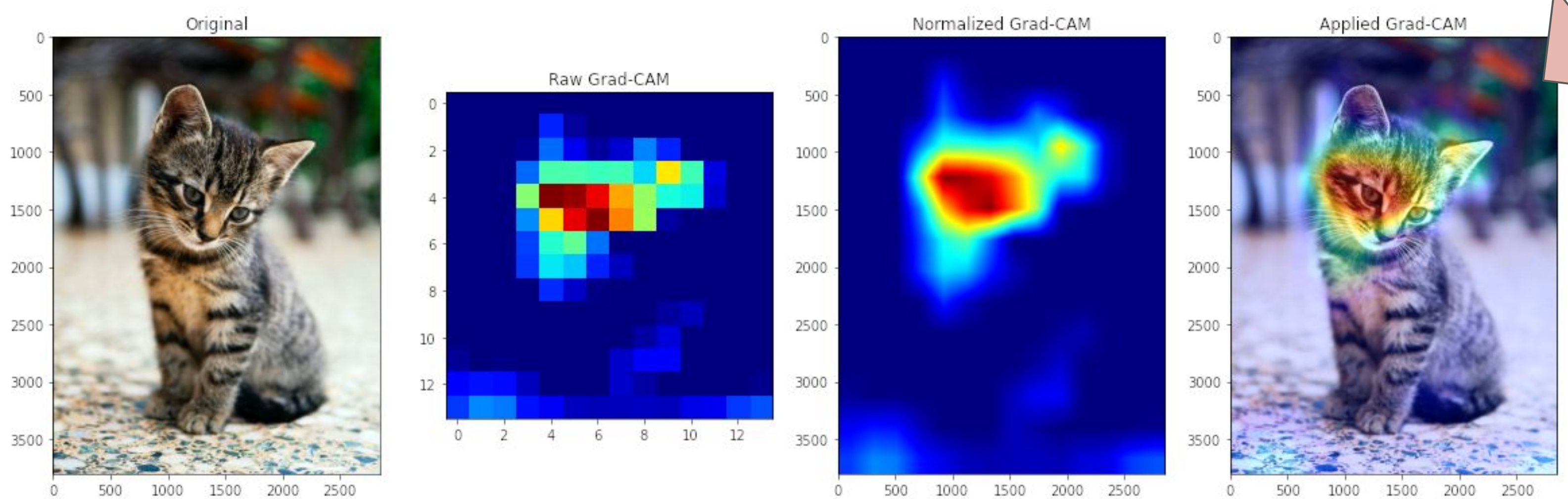
+

Model Explainability

Model Explainability

- *How* does the model arrive at its predictions?
- Why do we need interpretable models?
  - Build **trust** & **user confidence**
  - Develop **ethical** AI systems
- **Trade-off** between **performance** & **interpretability**
- **Gradient-based methods** have been developed

Gradient-weighted Class Activation Mapping



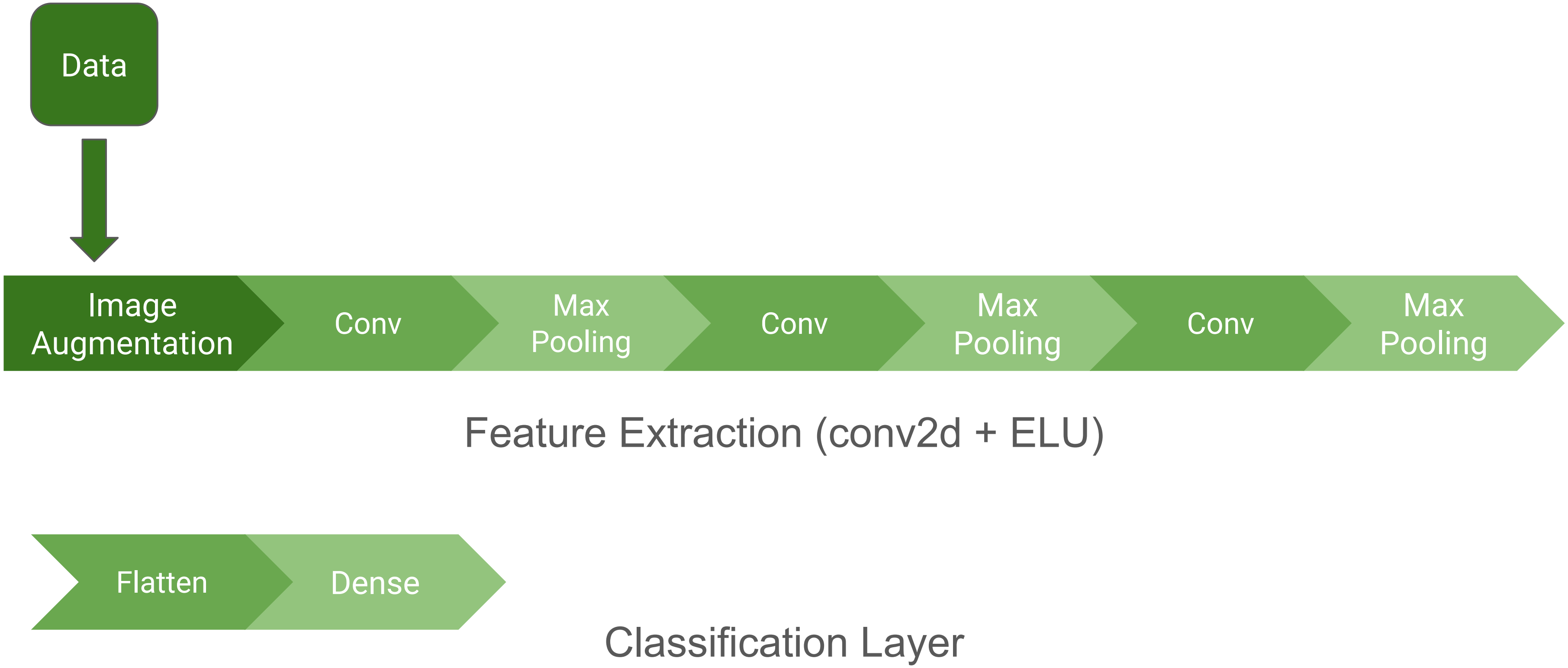
Goal: Create an equivalent of CNN explainability but for SVM model

SVM Explainability

[SVM explainability pics with suboptimal parameters]

[SVM explainability pics with optimal parameters]

Convolution Neural Network



Convolution Neural Network

Image Augmentation:

- random flip
- random zoom
- ~~random rotation~~

With Class weight {0: 1.5, 1: 1}

- **Accuracy**: 0.84

Data Preprocessing

- Load Images
- Interpolate Images
- HOG Feature Engineering

Machine Learning Models

- SVM
- CNN

Optimizing Data & Models

- HOG Features
- Support Vector Machines
- Convolutional Neural Network
- Image augmentation

Results & Discussion

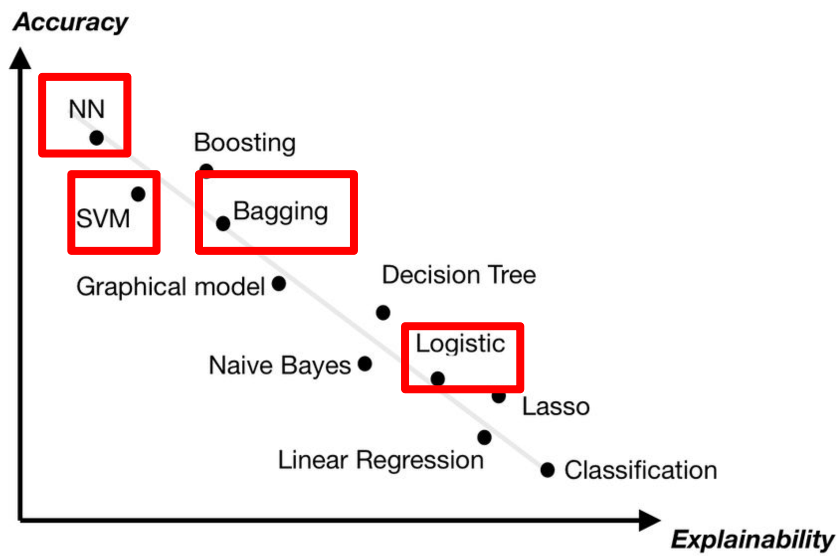
- Overall Results
- Challenges
- Future Research

Overall Results

- Simple Models (SVM) performance can compare to CNN (NFC-Net) Results
  - Interpretability, robustness

	Accuracy	Precision	Recall	F1 Score	ROC-AUC	Specificity
NFC-Net	0.91	0.95	0.85	0.897	.93	0.90
SVM RBF [32*32] (HOG)	0.896	0.923	0.945	0.934	.920	0.726

- Determining Simple Models
  - Trade-off
- Hyperparameter Tuning SVM
  - HOG feature tuning



Challenges & Future Research

Issues and Potential reasons?

- Data
- Computational time
- Limited timeline

What to do in the future to improve?

- Better Image Preprocessing
- Parameters
- Test Interpretability Techniques
- Standardizing Procedure
- Proof-of-Concept

Special thanks to

Dr. Peyman Kassani

Nadine Afari

Louis Ehwerhemuepha

Q & A