Detecting Accredited vs Non-Accredited Online Pharmacies using Deep Learning

Haley Egan, Mani Shanmugavel,
Tulsi Ratnam, and Seth Galluzzi

Aim

Use Deep Learning to detect accredited vs non-accredited pharmacy websites, using text and images.

* Github

Motivation

- NABP - 96% online pharmacies are illegitimate.
- Verifying legitimacy is manual/slow.
- Our study is first to train neural networks to detect accredited vs non-accredited pharmaceutical websites.

Research and Literature

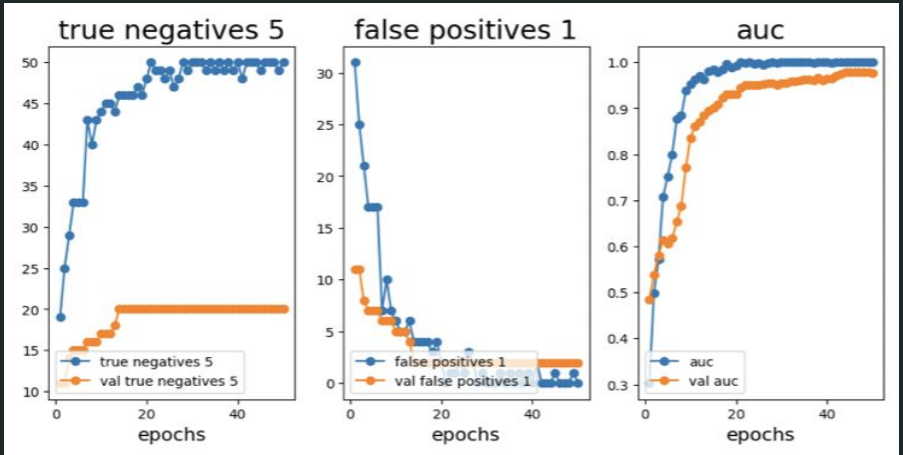
- The NABP website
  - accredited pharmacies & methodology for verifying site legitimacy
- The FDA
  - warning letters to illegitimate websites
- Detecting Fake Websites: The Contribution of Statistical Learning Theory
  - Website text & images with SVM algorithm to classify
- USPS & Verifone
  - validate the zip codes and phone numbers

Data Collection and Processing

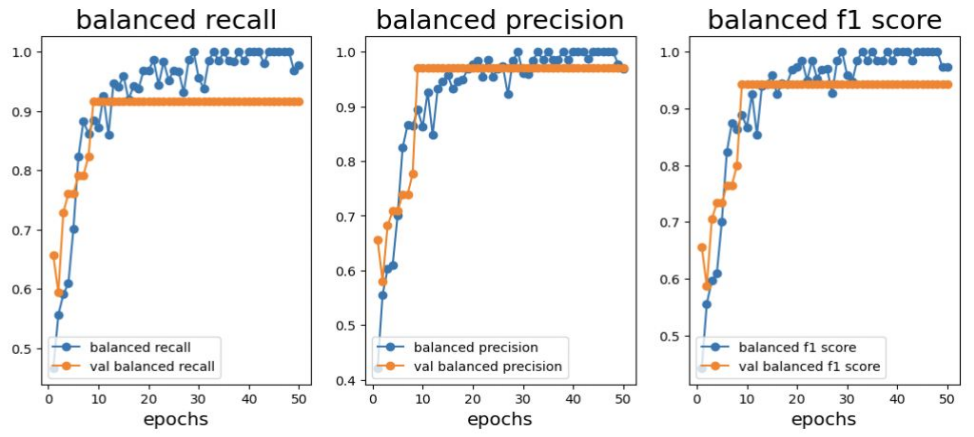
- Our custom dataset includes scraped content from the header, footer, and body of 72 accredited and non-accredited pharmacies.
- We also engineered verified address and verified phone number columns using the USPS and Verifone APIs.

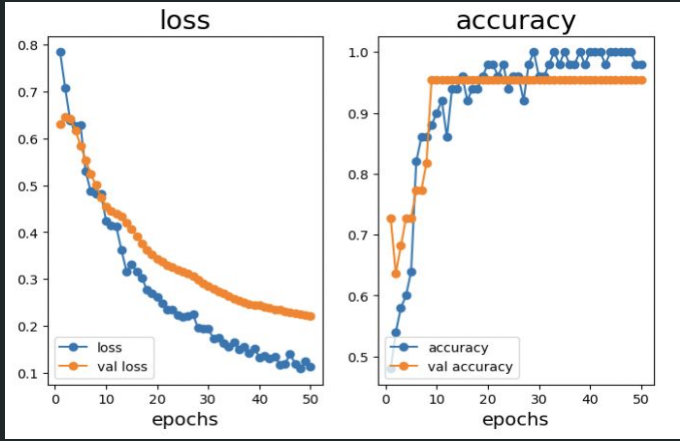
	Website	Body	Header	Footer	Image Urls	Accredited
1	https://acariahealth.e...	Skip to Main Content ...	Home Contact Insight...	Locations Referral Fo...	['https://acariahealth.e...	0
2	https://alto.com	Shop Alto Essentials t...	Shop Alto Essentials t...	Our Story Careers Dri...	['https://images.prismi...	0
3	https://www.amberph...	Loading... Skip to nav...		About Careers 340B ...	['https://www.amberph...	0
4	https://www.aoncolog...	For Patients Careers ...	For Patients Careers ...	© 2020 American On...	['https://dc.ads.linkedi...	0
5	https://www.birdirx.com	Loading... × Close ale...		Support Hours of Ope...	["https://www.birdirx.c...	0
6	https://www.cfspharm...	JavaScript seems to ...	Skip to Content Home...	Newsletter Submit yo...	['https://www.faceboo...	0
7	https://www.caremark...	Skip to main content ...	Stay healthier this flu ...	This page contains tr...	['https://www.caremar...	0
8	https://www.centerwel...	Skip to main content ...	Skip to main content ...	Back to top About I w...	['https://www.centerw...	0

BERT Model



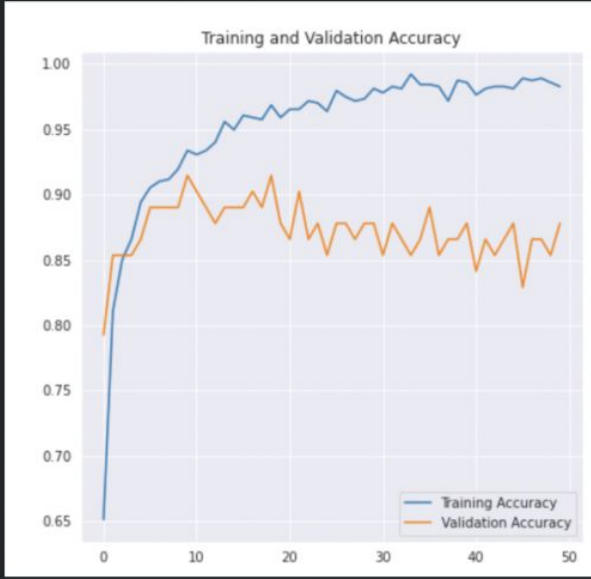
- Text Preprocessing
- 90% accuracy, 89% precision, 89% recall, 90% f1-score
- ROC&AUC - distinguishing labels
- Dropout & dense layers, softmax activation, 50 epochs



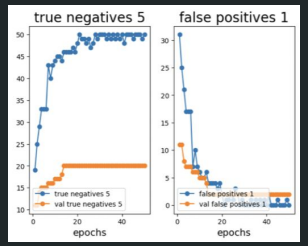


The MobileNetV2 Model was optimized further by using Adam.

	Optimizer	Accuracy
0	Adam	0.890244
1	AdaGrad	0.878049
3	NAdam	0.878049
4	RMSProp	0.878049
2	Adamax	0.865854
6	Nestrov	0.853659
5	Momentum	0.841463



	precision	recall	f1-score	support
0	0.97	0.83	0.90	47
1	0.81	0.97	0.88	35
accuracy			0.89	82
macro avg	0.89	0.90	0.89	82
weighted avg	0.90	0.89	0.89	82

Model	BERT	MobileNetV2
Accuracy (%)	95	89
Precision (%)	97	90
Recall (%)	91	89
F1 (%)	94	89
True Negatives and False Positives vs. Decision Matrix		39 | 8 1 | 34

Summary

- BERT performed really well with 95% accuracy
- MobileNetV2 was the best performing image model
- Between BERT and MobileNetV2, BERT was better able to classify accredited and non-accredited pharmacy websites

Future Work:

- Increase our dataset to include a greater number of websites - bigger training set
- Adapt the BERT model to include additional features such as zip code and phone number
- Combine the BERT and MobileNetV2 models into a multi-modal model

We hope our efforts will help future researchers identify and remove non-accredited pharmacies.