**UWA BOOTCAMP 2023** 

# Skin Cancer Prediction & Classification



## Why do this study?

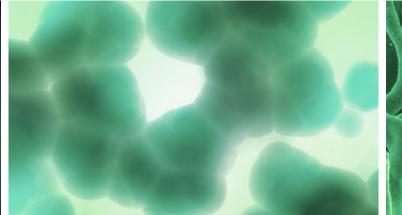
#### 1. How common is skin cancer?

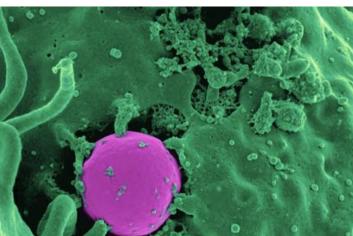
Over 95% of skin cancers are caused by exposure to UV radiation.

Most parts of Australia have high levels of UV radiation from the sun all year round.

Australia has one of the highest rates of skin cancer in the world. About 2 in 3 Australians will be diagnosed with some form of skin cancer before the age of 70.







when detected early, skin cancer is highly treatable!

## Why do this study?

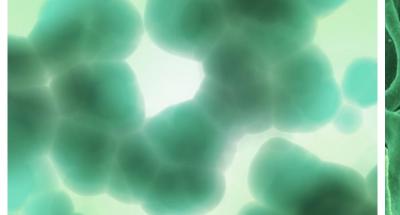
#### 1. How common is skin cancer?

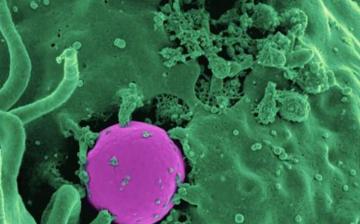
UV radiation cannot be seen or felt and it is not related to temperature. It can cause sunburn; premature skin ageing; and damage to skin cells, which can lead to skin cancer,

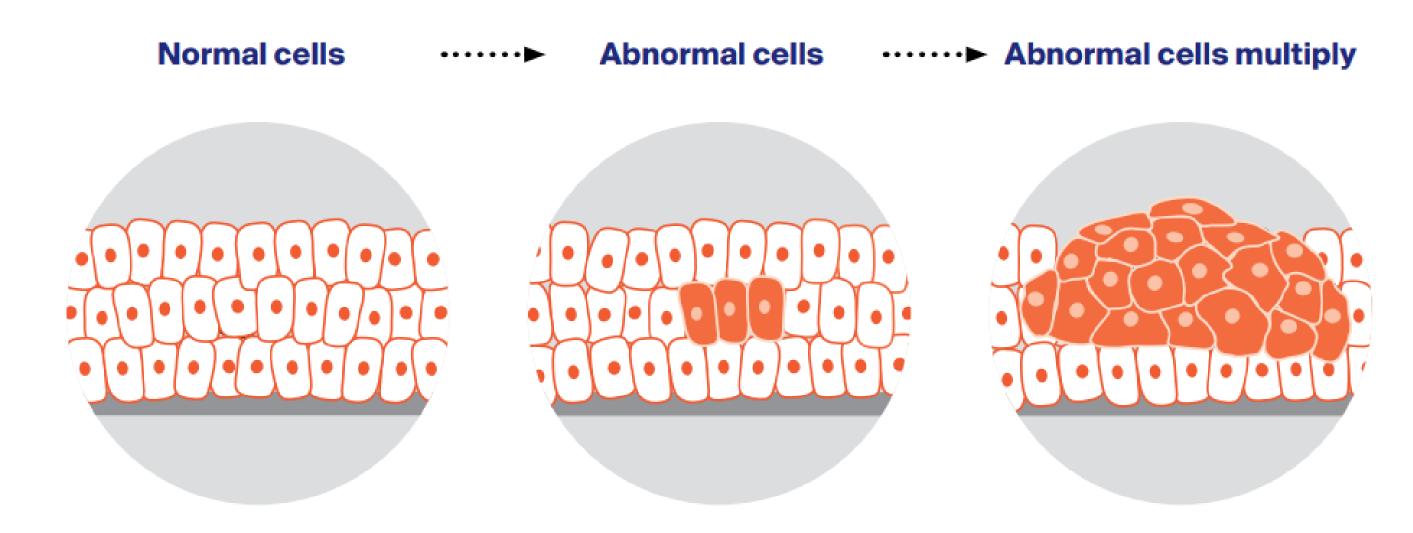
In 2012, Australia had the world's 2nd highest incidence rate of melanoma, at 35 new cases a year per 100,000 people. This was more than 11 times as high as the estimated average worldwide rate (3 per 100,000) (Globocan 2012). (AIHW & AACR 2014).

they all look the same to bare eyes, but are different! Its easy to go unnoticed!!









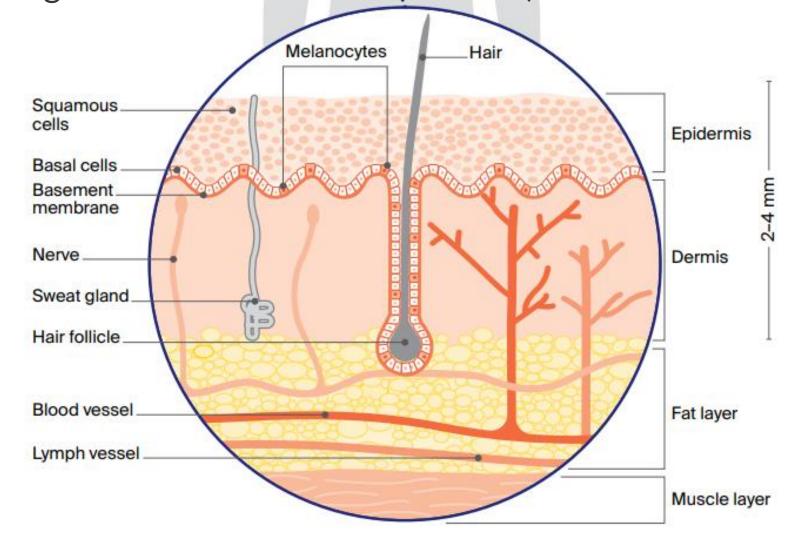
#### What is Cancer?

Cancer is a disease of the cells.

Cells are the body's basic building blocks. Normally, cells multiply and die in an orderly way. Sometimes, however, cells become abnormal and keep growing. These abnormal cells may turn into cancer.

The 3 main types of cells making up skin

- basal cells, the lower layer of the skin
- squamous cells, the top layer of skin
- melanocytes, which produce dark pigment that gives colour to the skin (melanin)



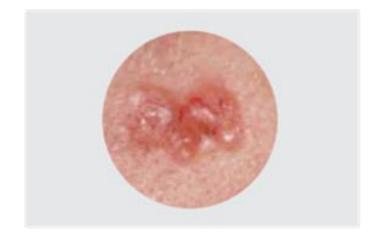
## The Skin Epidermis + Dermis

The skin is the largest organ in the human body. Its main functions include protecting the inner layers and organs from external elements, regulating body temperature and preventing dehydration (Cancer Council Victoria 2012a).

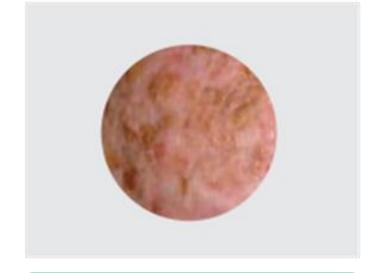
& cancer

Skin cancer is the uncontrolled growth of abnormal cells in the skin (Skin Cancer Foundation 2013).

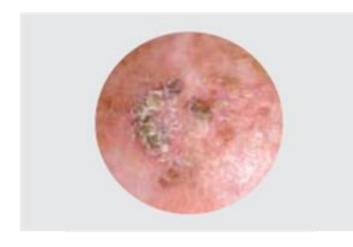
#### most common type - about 98%



Basal cell carcinoma (BCC)



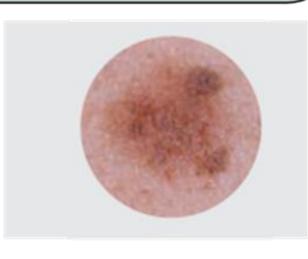
Benign keratosis-like



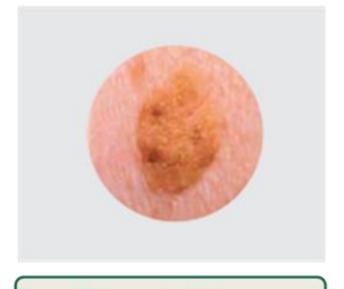
**Actinic Keratosis** 



Melanocytic Naevi

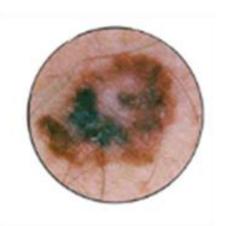


Vascular lesions



Dermatofibroma





Melanoma

#### 1. Prediction

### Lesion Classification Based on Key Parameters

To develop a predictive model capable of distinguishing between benign and cancerous skin lesions. This model will integrate four critical parameters: diagnostic type, patient age, patient sex, and lesion location. Such a model aims to assist in early-stage diagnosis, potentially improving patient outcomes.

#### 2. Classification

### RGB Dataset Exploitation for Cancer Type Detection

To harness the numerical values within the RGB dataset, representing the pixel intensity levels of skin lesion images, for the identification and prediction of seven different types of skin cancer.

#### 3. Image Classification

Image Dataset Analysis for Comprehensive Skin Cancer Prediction

This approach will involve the development and implementation of sophisticated algorithms capable of processing and interpreting complex visual data. The goal is to create a robust predictive model that enhances the accuracy and reliability of skin cancer diagnosis based on image analysis.

## Objectives

## Significance

#### Increased Accessibility and Convenience

Importance: These apps can provide valuable insights into skin health for people in remote or underserved areas, where dermatological services might not be readily available.

**Necessity:** With the proliferation of smartphones, such apps can leverage the widespread use of mobile technology to reach a broader population.

#### Early Detection and Diagnosis

**Importance:** Early detection of skin cancer, especially melanoma, is crucial for successful treatment. Apps that analyze skin lesions can alert individuals to potential problems early, leading to timely medical consultations.

**Necessity**: Considering the increasing incidence of skin cancer globally, such tools are necessary to augment traditional screening methods, especially in areas with limited access to dermatologists.

#### **Cost-Effective Screening**

**Importance**: They provide a cost-effective preliminary screening option, potentially reducing the need for in-person consultations for obviously benign lesions.

**Necessity**: Reducing unnecessary healthcare visits can save time and resources for both patients and healthcare systems.

**Dataset Used** 

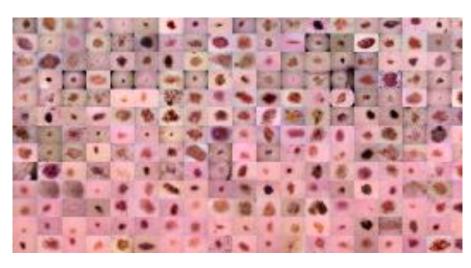
Prediction - HAM10000\_metadata.csv

Classification - hmnist\_28\_28\_RGB.csv

Image Classification - HAM10000\_images\_part\_1 & HAM10000\_images\_part\_2

lesion_id	image_id	dx	dx_type	age	sex	localization
HAM_0000118	ISIC_0027419	bkl	histo	80	male	scalp
HAM_0000118	ISIC_0025030	bkl	histo	80	male	scalp
HAM_0002730	ISIC_0026769	bkl	histo	80	male	scalp
HAM_0002730	ISIC_0025661	bkl	histo	80	male	scalp
HAM_0001466	ISIC_0031633	bkl	histo	75	male	ear
HAM_0001466	ISIC_0027850	bkl	histo	75	male	ear
HAM_0002761	ISIC_0029176	bkl	histo	60	male	face
HAM_0002761	ISIC_0029068	bkl	histo	60	male	face
HAM 0005132	ISIC 0025837	hkl	histo	70	female	back

pixel0000	pixel0001	pixel0002	pixel0003	pixel0004	pixel0005	pixel0006
192	153	193	195	155	192	197
25	14	30	68	48	75	123
192	138	153	200	145	163	203
38	19	30	95	59	72	143
158	113	139	194	144	174	215
8	1	3	19	5	10	26
194	147	137	197	148	139	197
161	121	105	169	128	119	172
125	84	85	165	114	118	18:
228	179	194	227	174	191	220



**Prediction** 

Classification

**Image Classification** 

## Dataset

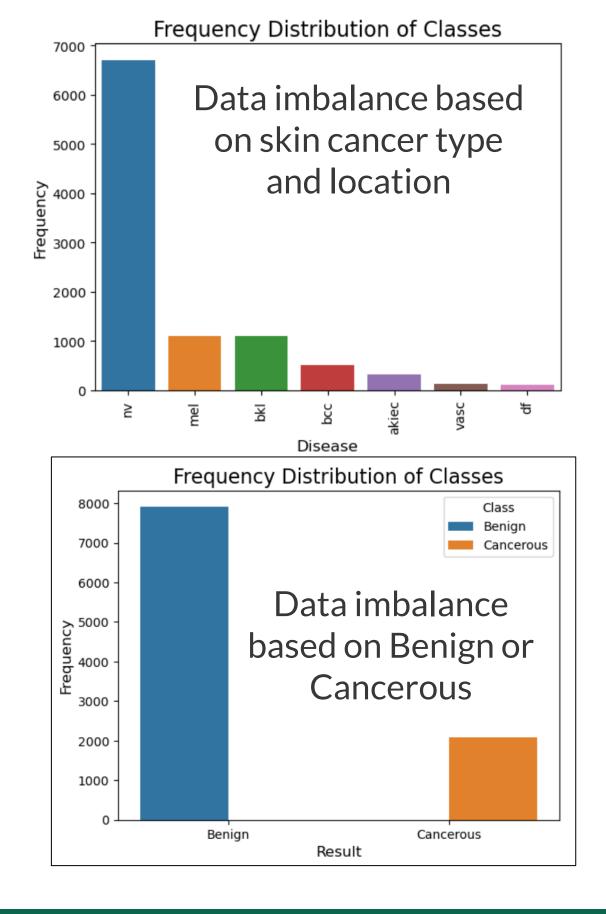
Prediction dataset- HAM10000\_metadata.csv

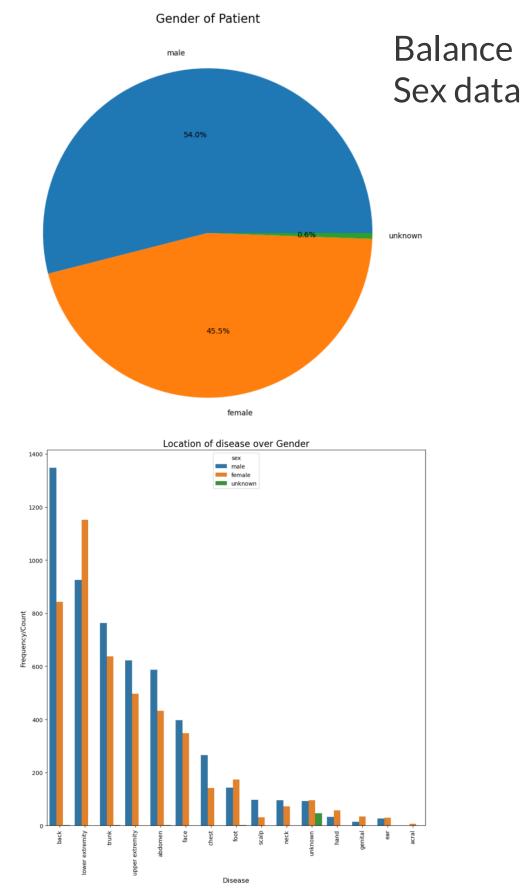
Age Columns with null values had to be changed by filling with mean age

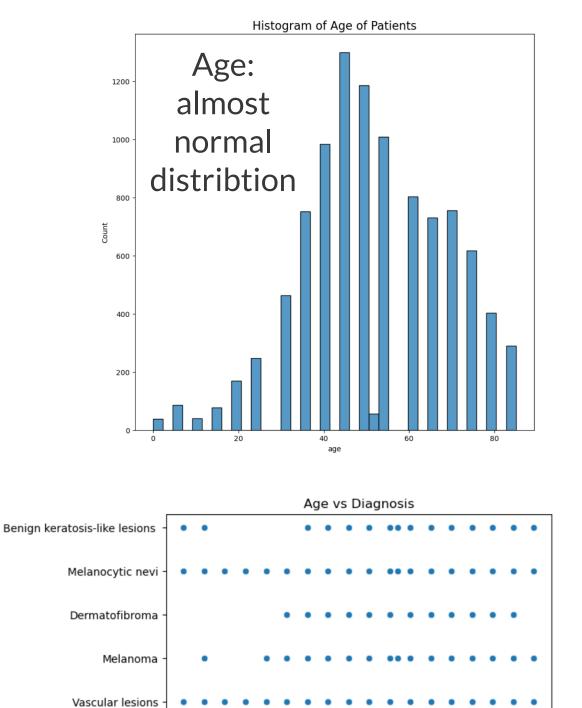
Sex Column has unknown values and treated as another class

Column names edited and loaded into cleaned\_skin\_metadata.csv









20

Basal cell carcinoma

Actinic keratoses



# Not all tumours are cancer!

Benign tumours tend to grow slowly and usually don't move into other parts of the body or turn into cancer.

Cancerous tumours, also known as cancerous tumours, have the potential to spread.

## Supervised Models Tested

**Skin Cancer Prediction** 

This model will integrate four critical parameters: diagnostic type, patient age, patient sex, and lesion location.

#### **Decision Tree**

Confusion Matrix

	Predicted Benign	Predicted Cancer
Benign	292	121
Cancer	105	321

Accuracy Score : 0.7306317044100119

Classification Depont

Classification	n Report precision	recall	f1-score	support
0	0.74	0.71	0.72	413
1	0.73	0.75	0.74	426
accuracy			0.73	839
macro avg	0.73	0.73	0.73	839
weighted avg	0.73	0.73	0.73	839

#### Random Forest

Confusion Matrix

	Predicted Benign	Predicted Cancer
Benign	297	116
Cancer	111	315

Accuracy Score : 0.7294398092967819

Classificatio	n Report precision	recall	f1-score	support
0	0.73	0.72	0.72	413
1	0.73	0.74	0.74	426
accuracy			0.73	839
macro avg	0.73	0.73	0.73	839
weighted avg	0.73	0.73	0.73	839

#### Logistic Regression

	Predicted Benign	Predicted Cancer
Benign	374	150
Cancer	141	383

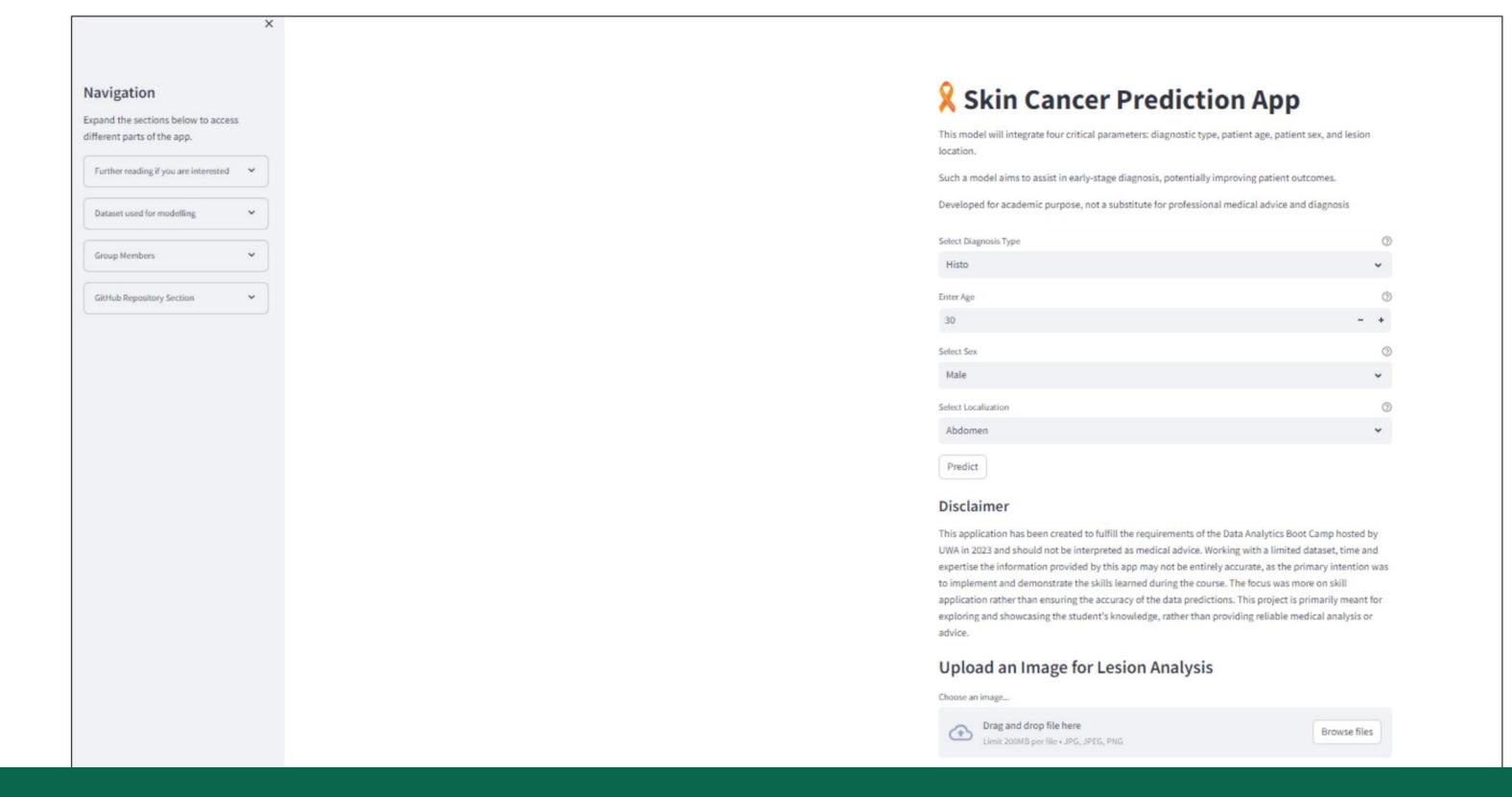
Accuracy Score : 0.7223282442748091

Classification Report

Classificacio	ii kepoi c			
	precision	recall	f1-score	support
Benign	0.73	0.71	0.72	524
Cancerous	0.72	0.73	0.72	524
accuracy			0.72	1048
macro avg	0.72	0.72	0.72	1048
weighted avg	0.72	0.72	0.72	1048

Confusion Matrix

	Predicted	Benign	Predicted Ma	lignant
Actual Beni	gn	367		157
Actual Maligna	ant	48 476		476
Classificatio	on Report: precision	recall	f1-score	support
Benign Malignant	0.88 0.75	0.70 0.91	0.78 0.82	524 524
accuracy macro avg	0.82	0.80	0.80 0.80	1048 1048



### Predicting Benign or Cancerous - Streamlit

## Skin Cancer Classification

Introduction

#### Cancerous

nv	Melanocyticnevi
bkl	Benignkeratosis-likelesions
df	Dermatofibroma

#### Benign

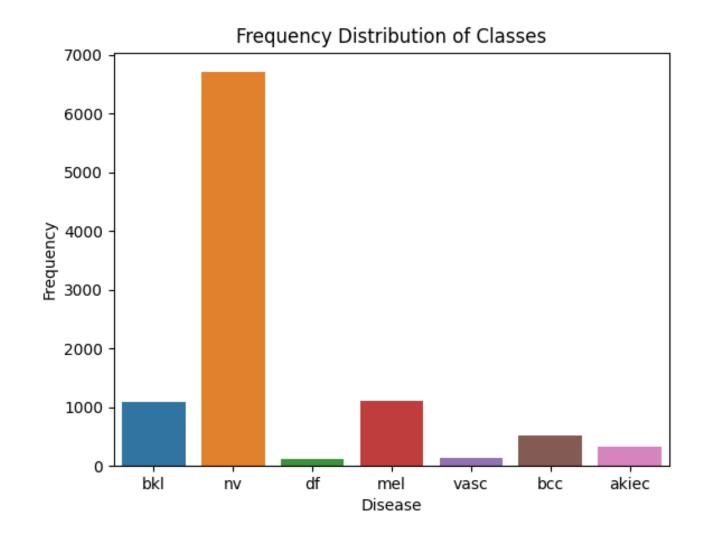
mel	Melanoma
bcc	Basalcellcarcinoma
akiec	Actinickeratoses
vasc	Vascularlesions

## Skin Cancer Classification

**Initial Analysis** 

Dataset: hmnist\_28\_28\_RGB.csv (10016 rows x 2532 cols)

1	Α	В	C	D	E	F	G	Н	1	J	K	L	M
1	pixel0000	pixel0001	pixel0002	pixel0003	pixel0004	pixel0005	pixel0006	pixel0007	pixel0008	pixel0009	pixel0010	pixel0011	pixel0012
2	192	153	193	195	155	192	197	154	185	202	162	192	208
3	25	14	30	68	48	75	123	93	126	158	128	158	172
4	192	138	153	200	145	163	201	142	160	206	149	165	207
5	38	19	30	95	59	72	143	103	119	171	125	134	177
6	158	113	139	194	144	174	215	162	191	225	179	214	232
7	8	1	3	19	5	10	26	8	13	34	13	24	100
8	194	147	137	197	148	139	197	148	132	200	154	142	202
9	161	121	105	169	128	119	172	129	116	176	134	125	181
10	125	84	85	165	114	118	181	120	125	188	133	142	189
11	228	179	194	227	174	191	226	165	182	215	157	175	206
12	23	13	16	22	12	14	29	20	24	76	63	65	122
13	159	129	127	166	136	135	170	140	142	170	137	128	173
14	37	22	22	100	74	61	145	119	111	155	128	126	159
15	155	131	142	160	135	146	165	139	157	168	142	162	172
16	193	164	192	194	162	190	195	166	193	195	168	191	197
17	197	158	179	201	160	182	201	160	182	202	163	181	203
18	171	133	142	176	141	153	180	144	153	183	140	149	193
19	212	160	173	214	157	173	219	169	185	220	168	183	220
20	166	131	138	174	139	145	185	155	163	187	154	162	199



# Analysis by Neural Networks (No Oversampler)

Skin Cancer Classification

#### Units = 64, Epoch = 20

Classificatio	on Report:					Predicted akiec	bcc	bkl	df	nv	vasc	mel
	precision	recall	f1-score	support	Actual akiec	28	7	7	0	14	1	4
akiec bcc	0.37 0.50	0.46 0.52	0.41 0.51	61 96	bcc	13	50	4	1	17	6	5
bkl	0.47	0.39	0.42	228	bkl	8	14	88	4	97	2	15
df	0.20	0.05	0.09	37	45	0			0	0	4	4
nv	0.82	0.92	0.87	1327	df	9	9	6	2	9	1	1
vasc mel	0.52 0.43	0.50 0.20	0.51 0.28	32 222	nv	12	10	46	2	1221	3	33
accuracy			0.72	2003	vasc	0	5	2	0	7	16	2
macro avg	0.47	0.43	0.44	2003	mel	6	6	34	1	128	2	45
weighted avg	0.69	0.72	0.70	2003								

#### Units = 100, Epoch = 50

Classification	on Report:					predicted akiec	DCC	DKI	αт	ΠV	vasc	mer
	precision	recall	f1-score	support	Actual akiec	31	6	4	1	17	0	2
akiec bcc		0.51 0.52	0.48 0.52	61 96	bcc	7	50	15	0	17	3	4
bkl df	0.48	0.45 0.16	0.46 0.26	228 37	bkl	8	11	102	1	76	0	30
nv vasc	0.84	0.90 0.38	0.87 0.48	1327 32	df	6	9	7	6	7	1	1
mel	0.44	0.34	0.39	222	nv	8	12	49	0	1198	2	58
accuracy			0.74	2003	vasc	0	6	3	0	10	12	1
macro avg weighted avg		0.47 0.74	0.49 0.72	2003 2003	mel	7	3	31	2	103	0	76

# Analysis by Neural Networks (No Oversampler, With Scaling)

Skin Cancer Classification

#### Units = 64, Epoch = 20

Classificatio	n Report:			
	precision	recall	f1-score	support
-1-1	0.00	0.00	0.00	
akiec	0.00	0.00	0.00	61
bcc	0.47	0.28	0.35	96
bkl	0.42	0.34	0.38	228
df	0.00	0.00	0.00	37
nv	0.74	0.97	0.84	1327
vasc	0.00	0.00	0.00	32
mel	0.59	0.10	0.17	222
accuracy			0.70	2003
macro avg	0.32	0.24	0.25	2003
weighted avg	0.63	0.70	0.64	2003

	Predicted	akiec	bcc	bk1	df	nv	vasc	mel
Actual akiec		0	7	23	0	30	0	1
bcc		1	27	14	0	53	0	1
bkl		0	7	77	0	138	0	6
df		0	4	14	0	19	0	0
nv		0	7	31	0	1282	0	7
vasc		0	4	2	0	26	0	0
mel		0	2	21	0	175	2	22

#### Units = 100, Epoch = 100

n Report:			
precision	recall	f1-score	support
0.29	0.38	0.33	61
0.50	0.58	0.54	96
0.56	0.33	0.42	228
0.33	0.03	0.05	37
0.83	0.88	0.86	1327
0.42	0.50	0.46	32
0.38	0.39	0.38	222
		0.71	2003
0.47	0.44	0.43	2003
0.70	0.71	0.70	2003
	0.29 0.50 0.56 0.33 0.83 0.42 0.38	precision recall  0.29 0.38  0.50 0.58  0.56 0.33  0.33 0.03  0.83 0.88  0.42 0.50  0.38 0.39	precision recall f1-score  0.29 0.38 0.33 0.50 0.58 0.54 0.56 0.33 0.42 0.33 0.03 0.05 0.83 0.88 0.86 0.42 0.50 0.46 0.38 0.39 0.38  0.71 0.47 0.44 0.43

	Predicted akiec	bcc	bkl	df	nv	vasc	mel
Actual akiec	23	10	5	2	11	0	10
bcc	7	56	4	0	15	5	9
bkl	18	7	75	0	88	3	37
df	7	13	3	1	7	1	5
nv	14	18	32	0	1173	10	80
vasc	3	2	1	0	7	16	3
mel	8	6	13	0	105	3	87

## Analysis by Neural Networks (With Oversampler)

**Skin Cancer Classification** 

#### Without Scaling: Units = 64, Epoch = 20

Classificatio	n Report:			
	precision	recall	f1-score	support
akiec	0.99	1.00	0.99	1359
bcc	0.98	1.00	0.99	1318
bkl	0.95	0.90	0.93	1262
df	0.99	1.00	1.00	1351
nv	0.89	0.80	0.84	1374
vasc	1.00	0.99	0.99	1358
mel	0.87	0.97	0.92	1365
accuracy			0.95	9387
macro avg	0.95	0.95	0.95	9387
weighted avg	0.95	0.95	0.95	9387

	Predicted akiec	bcc	bkl	df	nv	vasc	mel
Actual akiec	1359	0	0	0	0	0	0
bcc	2	1312	0	0	4	0	0
bkl	0	4	1139	1	88	1	29
df	0	0	0	1351	0	0	0
nv	15	23	52	5	1104	4	171
vasc	0	0	0	0	11	1347	0
mel	3	0	3	4	34	0	1321

#### With Scaling: Units = 100, Epoch = 100

Classification	Report:			
	precision	recall	f1-score	support
akiec	0.66	0.74	0.70	1359
bcc	0.58	0.73	0.65	1318
bkl	0.66	0.36	0.47	1262
df	0.76	0.92	0.83	1351
nv	0.69	0.61	0.65	1374
vasc	0.89	1.00	0.94	1358
mel	0.64	0.54	0.59	1365
accuracy			0.70	9387
macro avg	0.70	0.70	0.69	9387
weighted avg	0.70	0.70	0.69	9387

	Predicted akiec	bcc	bkl	df	nv	vasc	mel	
Actual akiec	1008	225	0	57	28	29	12	
bcc	129	964	22	101	29	55	18	
bkl	180	187	454	97	160	16	168	
df	16	58	0	1243	34	0	0	
nv	53	108	68	69	838	27	211	
vasc	0	0	0	0	0	1358	0	
mel	138	111	142	65	129	44	736	















































































# Analysis by Neural Networks (With Oversampler, With Scaling)

Skin Cancer Classification

#### Units = 100, Epoch = 50, 3 neural layers

Classificatio	n Report: precision	recall	f1-score	support		Predicted akiec	bcc	bk1	df	nv	vasc	mel
akiec	0.83	0.91	0.87	1359	Actual akiec	1243	26	49	19	0	0	22
bcc	0.87	0.78	0.82	1318	bcc	137	1024	78	21	16	17	25
bkl df	0.59 0.93	0.68 1.00	0.63 0.96	1262 1351	bkl	55	54	859	20	73	6	195
nv vasc	0.86 0.97	0.56 1.00	0.68 0.98	1374 1358	df	0	0	0	1351	0	0	0
mel	0.66	0.72	0.69	1365	nv	42	44	200	31	773	15	269
accuracy			0.81	9387	vasc	0	0	0	0	0	1358	0
macro avg weighted avg	0.81 0.82	0.81 0.81	0.80 0.81	9387 9387	mel	28	30	269	16	36	6	980

#### Units = 100, Epoch = 100, 4 neural layers

Classification	Report:			
F	recision	recall	f1-score	support
akiec	0.96	1.00	0.98	1359
bcc	0.95	0.96	0.95	1318
bkl	0.83	0.84	0.83	1262
df	0.99	1.00	0.99	1351
nv	0.83	0.67	0.74	1374
vasc	1.00	1.00	1.00	1358
mel	0.80	0.89	0.84	1365
accuracy			0.91	9387
macro avg	0.91	0.91	0.91	9387
weighted avg	0.91	0.91	0.91	9387

	Predicted akiec	bcc	bkl	df	nv	vasc	mel
Actual akiec	1353	6	0	0	0	0	0
bcc	25	1265	11	4	10	0	3
bkl	17	20	1061	1	92	0	71
df	0	0	0	1351	0	0	0
nv	17	37	153	10	927	5	225
vasc	0	0	0	0	0	1358	0
mel	2	6	56	0	89	0	1212











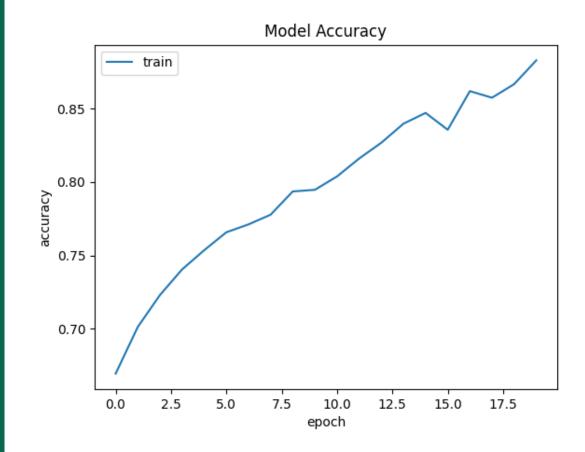


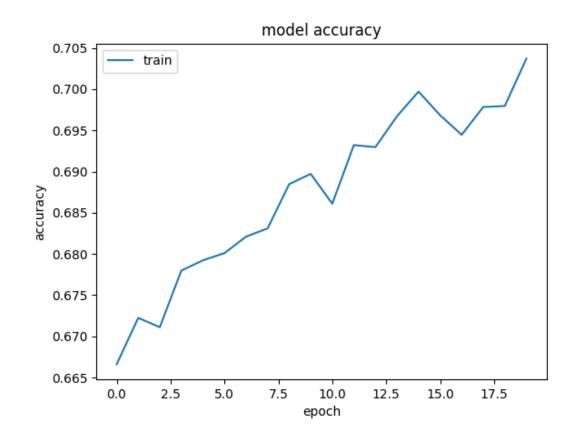


# Analysis by Neural Networks (Some Conclusions)

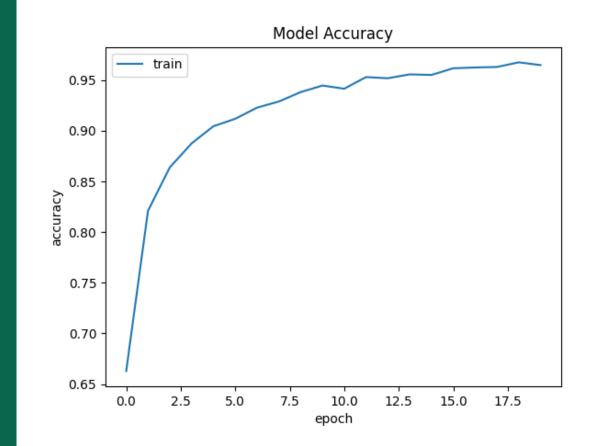
Skin Cancer Classification

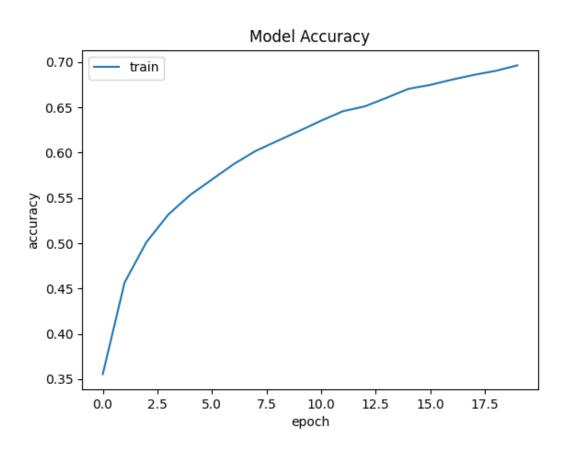
#### No OverSampler





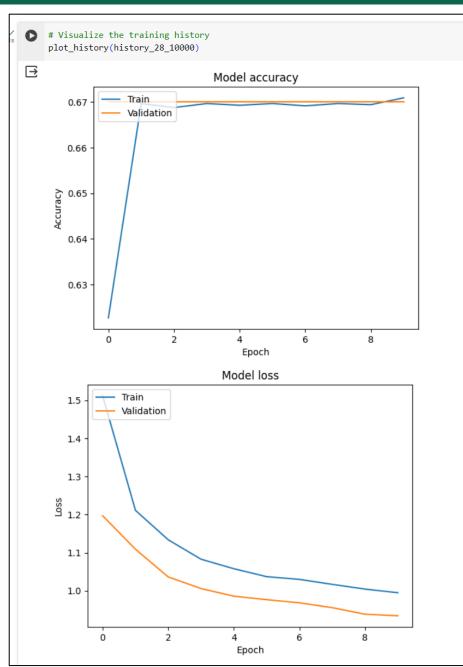
#### With OverSampler





# Skin Cancer Classification using Images

```
↑ ↓ ⊕ ■
# Assuming you have the necessary parameters and data generators ready
 # Define img_height, img_width, num_classes, train_generator, validation generator, epochs
 img_height, img_width = 28, 28
 num_classes = 7
 epochs = 10
 batch_size = 128
 subset size = 10000 # Adjust this value as needed
 json filename = '/content/Model3 28 10000 augmentation params.json
 metadata['diagnosis_no'] = metadata['diagnosis_no'].astype(str) # diagnosis_no has to be string type due to class_mode=categorical
 # Run the function to create generators
 train_gen, validation_gen = create_data_generators(subset_size=10000, json_filename=json_filename)
 # Create and train the model
 model 28 10000, history 28 10000 = create model(img height, img width, train gen, validation gen, epochs=epochs, num classes=num classes)
 Found 8000 validated image filenames belonging to 7 classes.
 Found 2000 validated image filenames belonging to 7 classes.
 Epoch 4/10
 Epoch 6/10
 Epoch 7/10
```



- Poor result.
- Need more time to improve on model architecture.
- Model take a long time to run even on Google Colab (GPU).

## Limitations & Risks

**Skin Cancer Prediction** 

#### Accuracy and Reliability:

The accuracy of these apps can vary, and there's a risk of false positives or negatives. Regular updates and improvements are necessary to ensure reliability.

## Regulatory and Ethical Concerns:

Apps handling health-related data must comply with privacy and data protection regulations. There's also a need for ethical considerations regarding how data is used and shared.

## Complementary, Not Substitute

These apps should be viewed as complementary tools and not as substitutes for professional medical advice and diagnosis. Users should be encouraged to seek professional evaluation for concerning lesions.

## User Misinterpretation and Anxiety

Users may misinterpret the results provided by the app, leading to unnecessary anxiety or, conversely, a false sense of security. This can result in delayed professional consultation for serious conditions or unwarranted visits for benign issues.

## Future Work

#### Further Work on the Classification

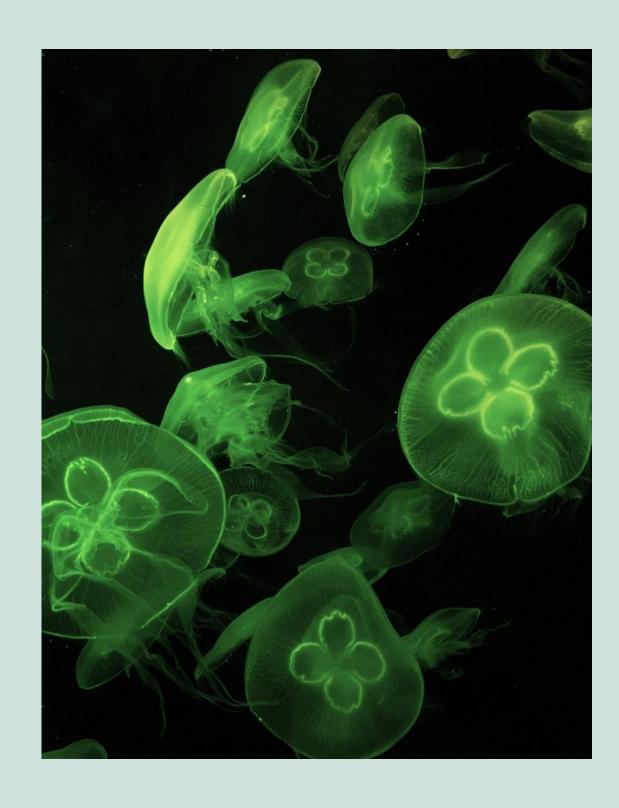
- Result seem robust
- Might consider other ML classification

#### Further Work on Cancer Prediction

- Try other methods. Gradient Boosting Machines (GBM)
- Use Neural Networks. CNNs or RNNs.
- K-Nearest Neigbors (KNN).
- Simplers Models: K-Means Clustering, L:inear Disriminant Analysis (LDA) or Principal Component Analysis (PCA).

#### Further Improvement Image Classification

- Increase Model Complexity. (Deeper CNNS or transfer learning from pre-trained models (liek VGG, ResNet or EfficientNet)
- Use higher than 28 x 28 image height and width
- Data augmentation. Already used but might need more adjustments.
- Hyperparameter Tuning. (learnign rate, batch size, optimizer)
- Class Imbalance Handling. Attempted but didn't succeed.
- Regularization Techniques. Already applied dropout but might consider L1, L2 regularization.



## Abbreviations

**ABS - Australian Bureau of Statistics** 

ACD - Australia Cancer Database

BCC - basal cell carcinoma

NMSC - non-melanoma skin cancer

SCC - squamous cell carcinoma

UV - ultraviolet

#### Resources

"Actinic Keratoses." Skin Cancer Foundation. Accessed [7<sup>th</sup> January 2024]. https://www.skincancer.org/skin-cancer-information/actinic-keratosis/.

"Basal Cell Carcinoma." Australian Cancer Research Foundation. Accessed [7<sup>th</sup> January 2024]. https://www.acrf.com.au/support-cancer-research/types-of-cancer/basal-cell-skin-cancer/?psafe\_param=1&utm\_source=google\_grant&utm\_medium=cpc&utm\_campaign={campaign}&utm\_content=154819747361&utm\_term=&gad\_source=1&gclid=EAlalQobChMI67KI1PnMgwMVVNBMAh2lfABnEAAYAiAAEglo4PD\_BwE.

"Benign Keratosis-like Lesions." Mayo Clinic. Accessed [7<sup>th</sup> January 2024]. https://www.mayoclinic.org/diseases-conditions/seborrheic-keratosis/symptoms-causes/syc-20353878.

"Dermatofibroma." The Skin Cancer Doctor. Accessed [7<sup>th</sup> January 2024]. https://www.theskincancerdoctor.com.au/education/skincancerlesions/dermatofibroma/.

"Melanocytic Nevi." Dermatology College. Accessed [7<sup>th</sup> January 2024]. https://www.dermcoll.edu.au/atoz/congenital-melanocytic-naevi/.

"Melanoma." Melanoma Institute Australia. Accessed [7<sup>th</sup> January 2024]. https://melanoma.org.au/about-melanoma/what-is-melanoma/?gclid=EAlalQobChMlvryno\_XMgwMVqswWBR1nJAC2EAAYASAAEgLiG\_D\_BwE.

"Vascular Lesions." Beauty on Rose. Accessed [7<sup>th</sup> January 2024]. <a href="https://beautyonrose.com.au/skin-condition-vascular-lesions-redness/">https://beautyonrose.com.au/skin-condition-vascular-lesions-redness/</a>.

"OesoCancerRisk+" Cancer-Risk. N.p., n.d. https://cancer-risk.streamlit.app/.