

Дипломный проект по детекции кожных покровов

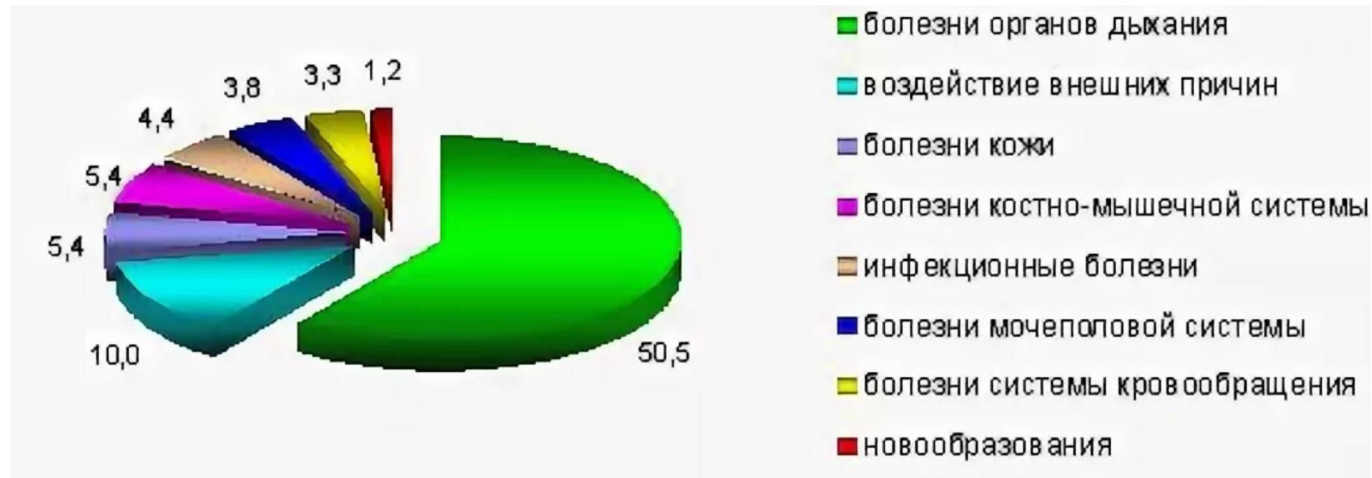


Жеманцева Татьяна

tatiana.zhemantseva@gmail.com

Почему актуальна эта тема?

Кожные заболевания могут быть различного происхождения, от дерматита до онкологических заболеваний. Среди всех заболеваний кожные занимают долю 5,4%.



Соцсети - наше все!

Интернет пестрит вопросами в форумах:

- на что это похоже?
- к какому врачу обращаться?
- чем лечиться?

“Экспертами” в большинстве случаев выступают такие же пользователи форума.

Система могла помочь бы частично ответить на эти вопросы, повысив тем самым рейтинг аккаунта сети. И при наличии дополнительного функционала связать с врачом.

```
Источник: class: Melanoma Skin Cancer Nevi and Moles, file: Melanoma Skin Cancer Nevi and Moles/atypical-nevi-24.jpg
Предсказано: класс: Melanoma Skin Cancer Nevi and Moles, с вероятностью: 0.997192
Предсказано: класс: Seborrheic Keratoses and other Benign Tumors, с вероятностью: 0.002630
Предсказано: класс: Light Diseases and Disorders of Pigmentation, с вероятностью: 0.000079
```

По ряду заболеваний разные врачи по своему трактуют его.

Например, детское акне и аллергические проявления часто путают в диагнозах.

Система могла бы стать некой первичной диагностикой, особенно в эпоху развития дистанционного оказания услуг.



Что сделано

Модель для выявления заболеваний кожи построена на нейросети Xception.

Для обучения собрано более 15тыс. изображений и поделено на 23 класса.

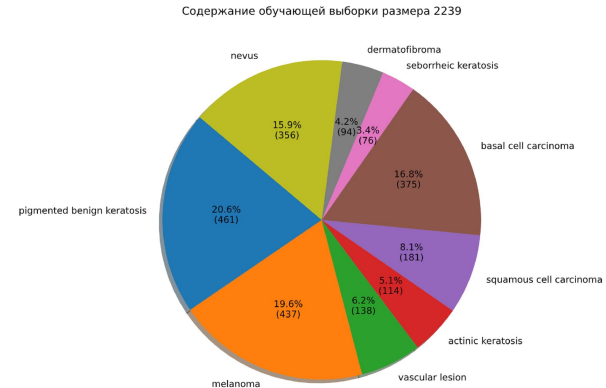
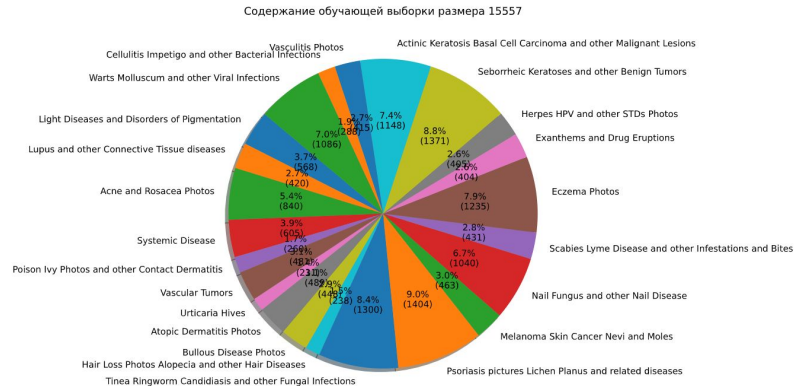
Модель способна определять кожное заболевание с определенной вероятностью соответствия. Максимальная точность модели, которую удалось достичь 60%.

С целью увеличения информирования и, таким образом, снижения ошибки системы в целом система выдает три варианта ответа.

Почему точность 60% ?

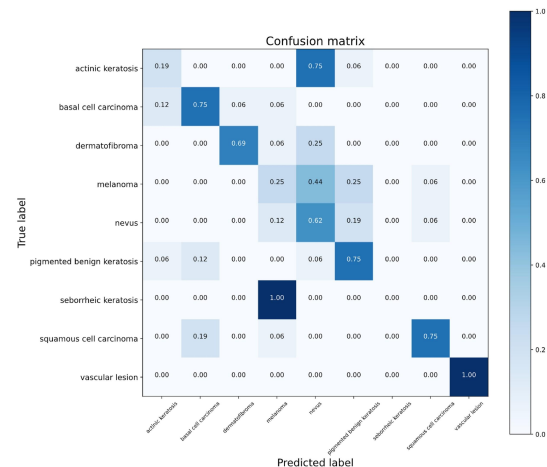
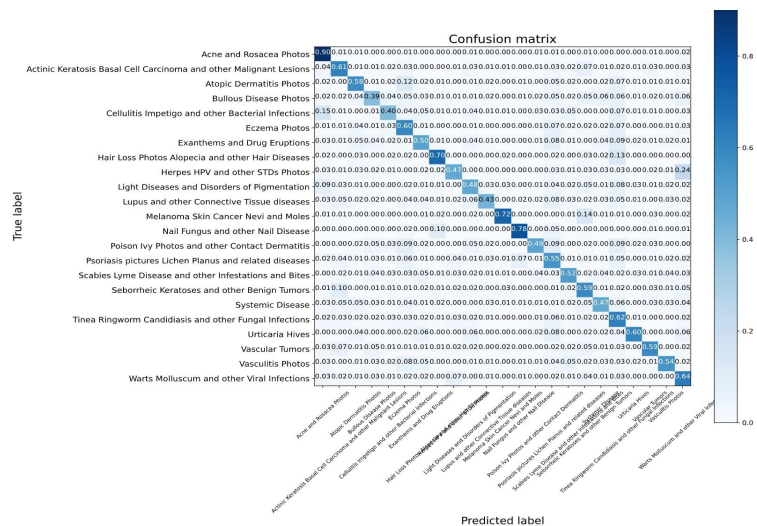
Почему же максимальная ассурансу=60% ?

Во-первых, у нас многоклассовая классификация и к тому же налицо дисбаланс классов. Проведем эксперимент применения модели на разных наборах данных.



Confusion matrix

Для обеих моделей были заданы веса. Обучение проведено на одинаковых архитектурах. Модель с большим набором данных для обучения в confusion matrix содержит большинство предсказаний по диагонали, что является правильным. Модель с маленьким набором данных настойчиво ошибается в нескольких классах.



Отчет о результатах классификации

Точность моделей accuracy, и macro f1 не сильно отличаются. Но есть большой разброс f1-score по классам.

	precision	recall	f1-score	support
Acne and Rosacea Photos	0.75	0.90	0.82	312
Actinic Keratosis Basal Cell Carcinoma and other Malignant Lesions	0.63	0.61	0.62	288
Atopic Dermatitis Photos	0.54	0.58	0.56	123
Bullous Disease Photos	0.40	0.39	0.40	113
Cellulitis Impetigo and other Bacterial Infections	0.35	0.40	0.37	73
Eczema Photos	0.62	0.60	0.61	309
Exanthems and Drug Eruptions	0.45	0.50	0.47	101
Hair Loss Photos Alopecia and other Hair Diseases	0.44	0.70	0.54	60
Herpes HPV and other STDs Photos	0.43	0.47	0.45	102
Light Diseases and Disorders of Pigmentation	0.51	0.48	0.49	143
Lupus and other Connective Tissue diseases	0.57	0.43	0.49	105
Melanoma Skin Cancer Nevi and Moles	0.72	0.72	0.72	116
Nail Fungus and other Nail Disease	0.85	0.78	0.81	261
Poison Ivy Photos and other Contact Dermatitis	0.51	0.49	0.50	65
Psoriasis pictures Lichen Planus and related diseases	0.59	0.55	0.56	352
Scabies Lyme Disease and other Infestations and Bites	0.49	0.52	0.50	108
Seborrheic Keratoses and other Benign Tumors	0.65	0.59	0.62	343
Systemic Disease	0.55	0.47	0.51	152
Tinea Ringworm Candidiasis and other Fungal Infections	0.58	0.62	0.60	325
Urticaria Hives	0.60	0.60	0.60	53
Vascular Tumors	0.58	0.59	0.58	121
Vasculitis Photos	0.62	0.54	0.58	105
Warts Molluscum and other Viral Infections	0.60	0.64	0.62	272
accuracy			0.60	4002
macro avg	0.57	0.57	0.57	4002
weighted avg	0.60	0.60	0.60	4002

	precision	recall	f1-score	support
actinic keratosis	0.50	0.19	0.27	16
basal cell carcinoma	0.71	0.75	0.73	16
dermatofibroma	0.92	0.69	0.79	16
melanoma	0.33	0.25	0.29	16
nevus	0.29	0.62	0.40	16
pigmented benign keratosis	0.60	0.75	0.67	16
seborrheic keratosis	0.00	0.00	0.00	3
squamous cell carcinoma	0.86	0.75	0.80	16
vascular lesion	1.00	1.00	1.00	3
accuracy			0.57	118
macro avg	0.58	0.56	0.55	118
weighted avg	0.60	0.57	0.56	118

Что показал эксперимент?

1. Большое количество примеров для обучения положительно влияет на метрики предсказания каждого класса в отдельности.
2. Точность модели лучше на больших данных, но не намного.
3. Дисбаланс классов поправим, но если это не половина классов данных.
4. При этом точность модели не превзошла 60. Можно сделать вывод, что проблема в архитектуре. Подобные задачи рекомендуется проводить с сегментацией и классификацией. Но это уже более сложная работа.

Что дальше?

Self-Supervised Image Classification

