**Министерство образования и науки Российской Федерации**

**(МИНОБРНАУКИ РОССИИ)**

**ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (ТГУ)**

**Автономные магистерские программы**

**Интеллектуальный анализ больших данных**

**ОТЧЕТ**

**о выполнении технического задания**

**Выполнил**

**студент группы №148БИ**

**\_\_\_\_\_\_\_\_\_\_\_\_\_О.А. Майер**

**Проверил д-р. техн. наук**

**доцент кафедры**

**теоретических основ**

**информатики**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_А.В.Замятин**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Отчет защищен**

**с оценкой\_\_\_\_\_\_\_\_\_\_\_**

**\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Томск 2018**

Оглавление

[Введение 3](#_Toc531111934)

[Предметная область 3](#_Toc531111935)

[Описание исходных данных 4](#_Toc531111936)

[Предобработка исходных изображений 5](#_Toc531111937)

[Построение архитектуры модели 6](#_Toc531111938)

[Интерпретация полученных результатов 7](#_Toc531111939)

[Выводы. 8](#_Toc531111940)

# Введение

Распознавание типов заболеваний человеческого тела по накопленным медицинским данным, в эпоху развития систем машинного обучения, является одной из важнейших задача, для улучшения качества жизни каждого человека. Различные алгоритмы классификации уже показывают весьма высокие показатели при решении данного типа задач. Как пример, Исследователи Стэнфордского университета [1], или исследователи s, Университета Брандейса(Расписать чем они там занимались более подробно)

Главными целями данной статьи будут являться:

* Предварительная обработка данных – получение новых признаков, смена формата исходных изображений для оптимального построения свёрточной нейронной сети.
* Построение модели нейронной сети для классификации изображений пигментных поражений.
* Интеграция обученной модели в прототип прикладной системы поддержки принятия решений при постановке анамнеза.

В процессе работы над тремя основными задачами, нам придётся столкнуться со следующим спектром подзадач, таких как. Предобработка изображений, повышение яркости изображения, адаптивные пороговые преобразования, искусственное увеличение размерности выборки изображений, нормализация исходных данных, преобразование по принципу one-hot-encoding, построение топологии Свёрточной нейронной сети, интерпретация и анализ полученных результатов работы свёрточной нейронной сети, визуализация промежуточных изображений в слоях свёрточной нейронной сети, разработка прототипа системы классификации изображений используя построенную модель нейронной сети.

Решение всего вышеуказанного спектра задач приведет нас к решению основной цели данного отчёта(статьи), а именно построению системы поддержки принятия решений при постановке диагноза кожного поражения.

## Предметная область

Виды поражений кожи, для определения моделью

1) Actinic keratoses and intraepithelial carcinoma / Bowen's disease - Актинические кератозы, интраэпителиальная карцинома и болезнь Боуэна

2) basal cell carcinoma – Базальноклеточная карцинома: постоянное обновление клеточного состава за счет непрерывного размножения клеток базального (самого глубокого) слоя.

3) benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses)

4) dermatofibroma – Фиброма кожи: Фиброма кожи — это доброкачественная опухоль, которая состоит из соединительной ткани. Она располагается под верхним слоем кожи. Фиброма практически не представляет опасности в плане перерождения в злокачественную опухоль. Но она является косметическим дефектом и доставляет человеку дискомфорт.

5) melanoma - злокачественная опухоль, развивающаяся из меланоцитов — пигментных клеток, продуцирующих меланины. Наряду с плоскоклеточным и базальноклеточным раком кожи относится к злокачественным опухолям кожи. Преимущественно локализуется в коже, реже — сетчатке глаза, слизистых оболочках (полость рта, влагалище, прямая кишка). Одна из наиболее опасных злокачественных опухолей человека, часто рецидивирующая и метастазирующая лимфогенным и гематогенным путём почти во все органы.

6) melanocytic nevi - Меланоцитарные невусы (или родинки) представлены многими разновидностями и есть почти у всех людей – они выявляются у ¾ части населения земного шара. Такие образования на коже являются доброкачественными опухолями

7) vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage) - К сосудистым поражениям относятся приобретенные (например, пиогенная гранулема) и врожденные или проявляющиеся вскоре после рождения (сосудистые родимые пятна). К сосудистым родимым пятнам относятся сосудистые опухоли (например, врожденная гемангиома) и сосудистые мальформации.

Ещё немного информации про поражения кожи, вводной и общей.

# Описание исходных данных

Дополнительные данные и csv

Условное обозначение целевого признака

Размер выборки: 10000 цветных изображений пигментных поражений.

Репрезентативность: Выборка репрезентативна

Дополнительная информация: Каждое изображение имеет дополнительные параметры

Примеры изображений

Размерность

Предварительная обработка данных (Распределение по группам, в зависимости от входных параметров. Построение модели нейронной сети для классификации изображений пигментных поражений

Интеграция обученной модели в веб сервис.

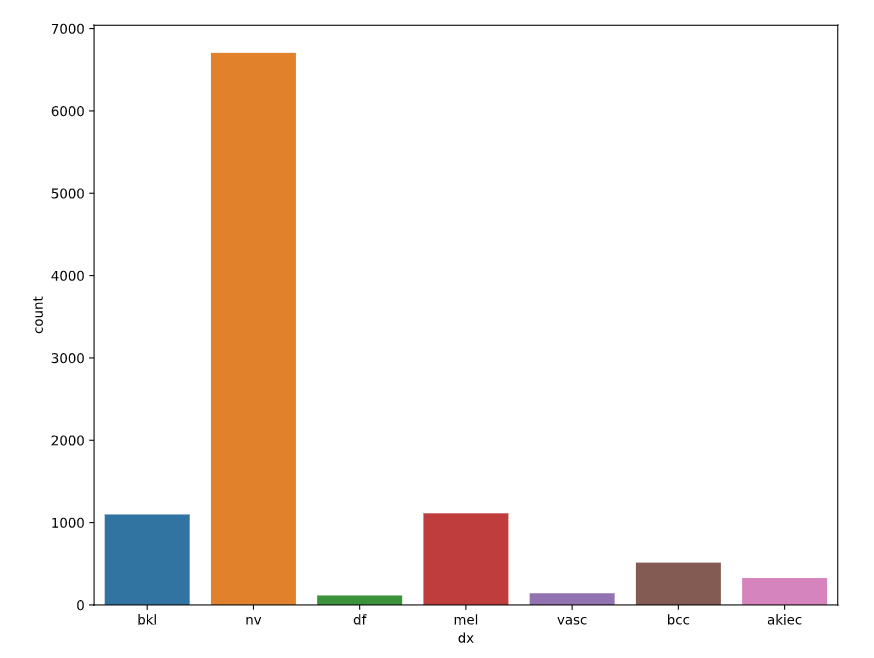


Рисунок 4. Гистограмма распределения целевого признака.

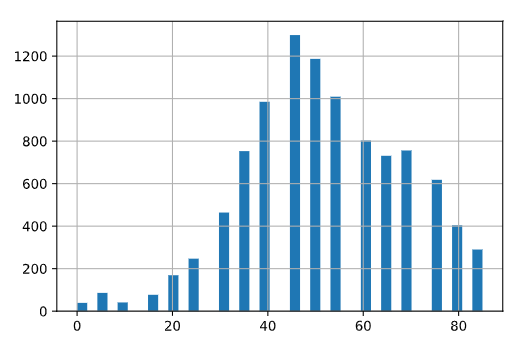


Рисунок 5. Распределение выборки по возрастам.

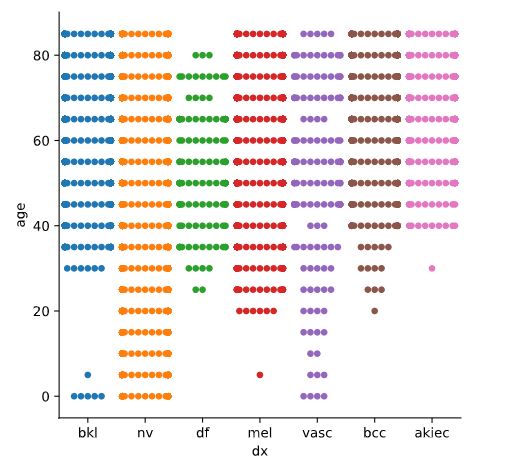


Рисунок 4. Корректно назвать .

# Предобработка исходных изображений

Необходимо описать каждый из указанных ниже процессов

Считывание данных

Уменьшение количества классов для поиска ввиду – низкого количества экземпляров на 3х классах

Image Augmentaion – увеличение выборки

Трансформация выборки

Нормализация векторов изображений

Работа с билиотекой OPENCV2 для изображений

Рзделение данных на тестовую и валидационную выборку

One hot encoding для задачи классиффикации

Аугментация далее используется так как данное действие хорошо себя показывает на всех видах модели

Выделение дополнительных каналов изображений

ADAPTIVE\_THRESH\_GAUSSIAN

Channel

ADAPTIVE\_THRESH\_MEAN

Channel

Gray channel

Общая размерность выборки

Снижение размерности изображений

# Построение архитектуры модели

Определение используемых метрик модели

Определение используемых оптимизаторов модели

Adam и RMSprop

## Построение топологии нейронной сети

Процесс перехода от одной версии к другой

Добавление

## Топология нейросетевых моделей

В процессе работы было построено и протестировано 4 основных топологии нейросети

УКАЗАТЬ ИХ

Виды топологии сети

* Топлогия A – готовая топология сети, используемая для распознавания рукописного текста
* Топология B – Улучшенная версия топологии A с 2мя слоями свертки и 2 полносвязными слоями на выходе
* Топология С – Собственная топология ОПИСАНИЕ +картинка
* Топология D – Собственная топология ОПИСАНИЕ + картинка

## Проведение экспериментов

Табличка

Про каждый эксперимент, обозначенный как v1….

Указать параметры при эксперименте и провести его заново 10 раз посчитав средним все параметры

В приложении указать табличку которая пойдёт в презентацию

Основываясь на реферате начать прогонять каждую модель по параметрам

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Характеристики  Модели | V1.0 | V1.1 | V1.2 | V1.3 | V1.4 |  |  |
| Количество классов на выходе | 7 | 7 | 3 | 4 | 4 |  |  |
| Вид топологии | A | B | C | D | D |  |  |
| Gray channel | - | - | - | - | - |  |  |
| ADAPTIVE\_THRESH\_  GAUSSIAN  Channel | - | - | - | - | - |  |  |
| ADAPTIVE\_THRESH\_  MEAN  Channel | - | - | - | - | - |  |  |
| RGB Channel | + | + | + | + | + |  |  |
| Augmentation | - | - | - | + | + |  |  |
| Нормализация данных | - | - | - | - | + |  |  |
| Вид оптимизирующей функции | Adam | Adam | Adam | Adam | Adam |  |  |
| Общая размерность выборки |  |  |  |  |  |  |  |
| Размерность изображений |  |  |  |  |  |  |  |
| Общее количество связей в сети |  |  |  |  |  |  |  |

# Интерпретация полученных результатов

# Создание СППА.

При выполнении работы были рассмотрены и применены на практике ряд наиболее используемых алгоритмов машинного обучения. Изучены их положительные и отрицательные стороны.

# Выводы.

При выполнении работы были рассмотрены и применены на практике ряд наиболее используемых алгоритмов машинного обучения. Изучены их положительные и отрицательные стороны.

References

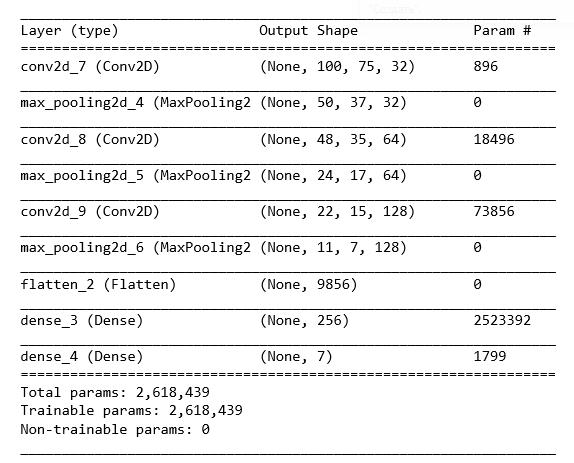
[1]: Dermatologist-level classification of skin cancer with deep neural networks

Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

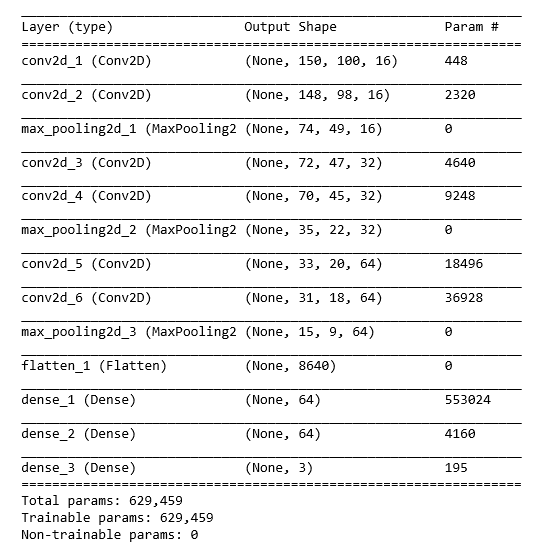
[2]: Fingerprints, Facial Recognition and Cancer Roy J Vaz\* Division of Graduate Professional Studies, Rabb School of Continuing Studies, Brandeis University, Waltham, MA 02453, USA

# Приложение А – Топологии нейронных сетей

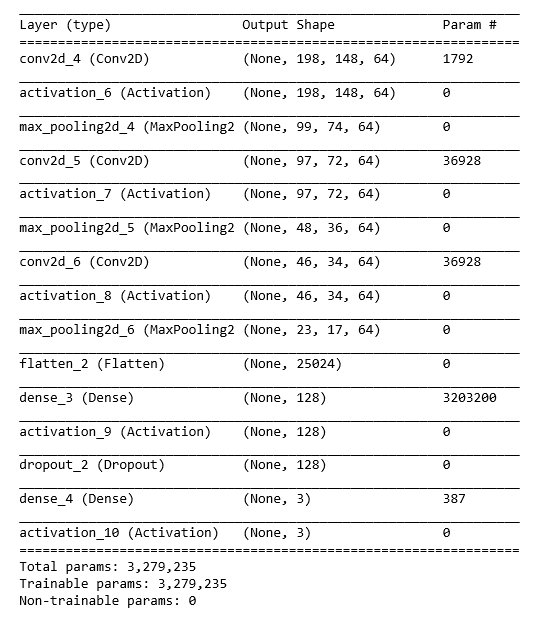
Топология A



Топология B



Топология C



Топология D

