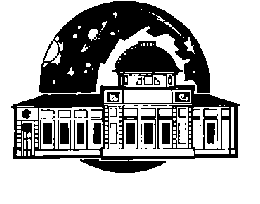
**Муниципальное общеобразовательное учреждение «Средняя школа с углубленным изучением отдельных предметов «Провинциальный колледж»**

***Исследовательское направление – информатика***

**Определение характера человека по почерку**

**с помощью нейронных сетей**

***Исследовательская работа***

**­**

Выполнена ученицей 11 информационно-технологического класса Средней школы «Провинциальный колледж» **Чвановой Валерией Олеговной**

Научный руководитель – учитель математики и информатики Средней школы «Провинциальный колледж» **Легкова Мария Николаевна**

Оглавление

[Введение 3](#_Toc24987330)

[1. Нейронные сети 4](#_Toc24987331)

[1.1 Определение нейронных сетей 4](#_Toc24987332)

[1.2 Обучение с учителем 5](#_Toc24987333)

[1.3 Алгоритм работы нейронной сети 5](#_Toc24987334)

[2. Графология 7](#_Toc24987335)

[2.1 Определение графологии 7](#_Toc24987336)

[2.2 Размер букв 7](#_Toc24987337)

[2.3 Наклон букв 7](#_Toc24987338)

[2.4 Расстояние между строками 8](#_Toc24987339)

[2.5 Расстояние между словами 8](#_Toc24987340)

[2.6 Форма букв 8](#_Toc24987341)

[3. Реализация 9](#_Toc24987342)

[3.1 Среда, используемая для разработки нейронных сетей 9](#_Toc24987343)

[3.2 Описание датасета 9](#_Toc24987344)

[3.3 Описание работы нейронной сети 9](#_Toc24987345)

[3.4 Статистика определения характеристик почерка 11](#_Toc24987346)

[Заключение 12](#_Toc24987347)

[Список использованных источников и литературы: 13](#_Toc24987348)

# Введение

**Актуальность**: Установлено, что индивидуальные особенности человека проявляются в его почерке. Для того, чтобы «читать» человека, словно книгу, необходимо знать систему признаков почерка, каждый из которых соответствует определенному свойству характера. Давно замечено, что почерки так же заметно отличаются друг от друга, как и индивидуальная природа людей, как отпечатки их пальцев. Почерк может очень много рассказать о характере человека, о его наклонностях. Многие кампании, принимая на работу клиента, просят написать пару строк, чтобы оценить основные черты его характера, и сделать вывод, подходит он для данной работы, или нет. Знания о почерке человека также используются в криминалистике для установления принадлежности написанного, проводятся почерковедческие экспертизы. Но все это делается экспертом индивидуально и на такую работу уходит огромное количество времени. Из-за относительной новизны машинного обучения тема автоматизации оценки характеристик почерка человека не рассматривалась ранее.

**Цель исследования**: обучить сверточную нейронную сеть определять основные признаки почерка и на их основе делать выводы о характере человека.

**Задачи исследования**:

1. Изучить искусственные нейронные сети, проанализировать возможности их применения.
2. Исследовать структуру и принцип работы сверточной нейронной сети.
3. Изучить связь основных характеристик почерка и характера человека.
4. Собрать датасет и подготовить его для обучения нейронной сети.
5. Запрограммировать сверточную нейронную сеть и обучить ее распознавать основные характеристики почерка.
6. Написать код, позволяющий определять характер человека по имеющимся знаниям о его почерке (характеристики почерка получены в результате работы нейронной сети).

**Объект исследования**: машинное обучение

**Предмет исследования**: нейронные сети

**Методы исследования**:

1. анализ (используется для изучения литературы и источников по машинному обучению и связи почерка с характером человека),
2. сравнение (используется для сравнения существующих решений),
3. моделирование (используется для написания кода нейронной сети),
4. эксперимент (используется для подбора параметров нейронной сети),
5. обобщение (используется для подведения итога работы нейронной сети).

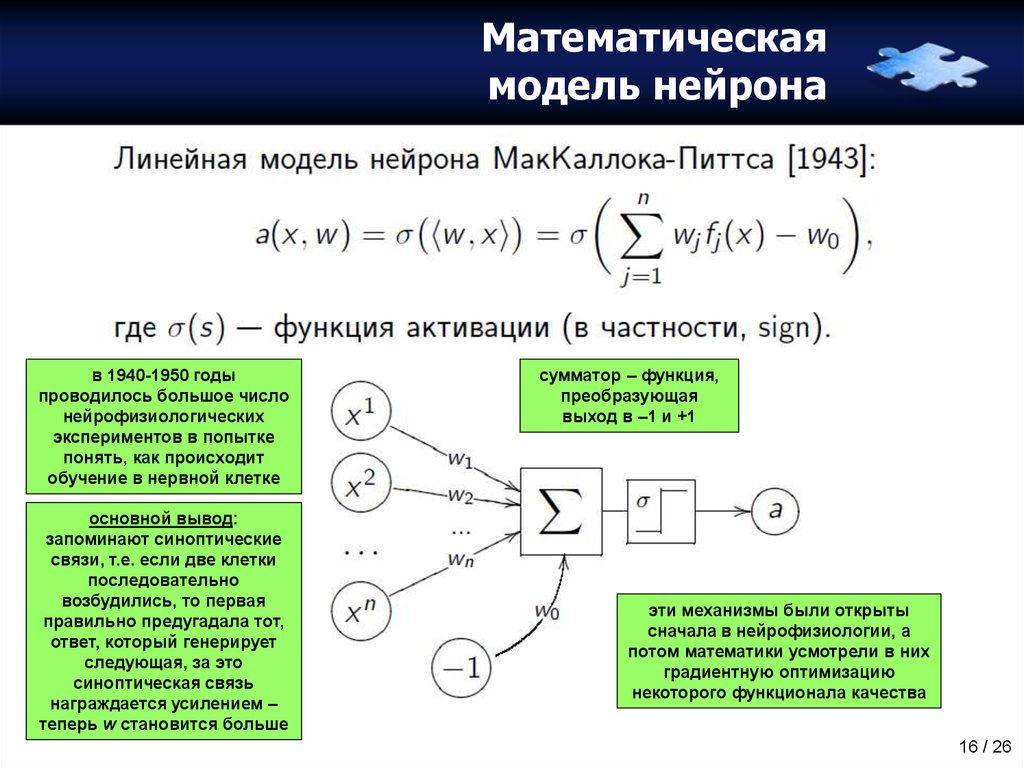
# 1. Нейронные сети

## 1.1 Определение нейронных сетей

**Нейронная** **сеть** — это математическая модель, а также ее программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма.

Мозг, как животного, так и человека, является биологической нейронной сетью, которая состоит из нейронов, что в свою очередь состоят из дендритов, аксонов и прочих отростков. Задача этих элементов – принимать сигналы извне и отправлять обработанную информацию соседним клеткам. Первая попытка смоделировать процессы, протекающие в мозге, была осуществлена У. Маккалоком и У. Питтсом, которые в 1943г. предложили систему обработки информации в виде сети, состоящей из простых вычислителей, созданных по принципу биологического нейрона.

На рис.1 представлена линейная модель нейрона МакКаллока – Питтса (1943 г.):



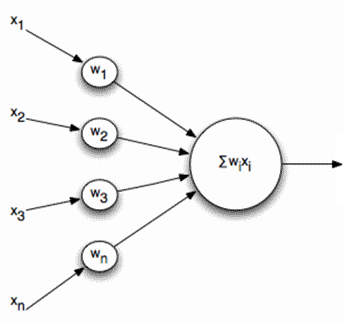


Рис.1. Линейная модель нейрона МакКаллока – Питтса

Если соединить множество нейронов, получится многослойная линейная нейросеть —вычислительная система, имеющая способность к самообучению, постепенному повышению своей производительности.

Многослойная нейронная сеть, имеющая в своей архитектуре кроме полносвязных слоев еще и сверточные, называется сверточной нейронной сетью. Такая нейросеть — это прототип зрительной коры головного мозга, в которой были открыты так называемые простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определённого набора простых клеток. Суть сверток в том, что каждый фрагмент изображения умножается на матрицу свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения. Благодаря своим сверточным слоям нейронная сеть не привязывается к расположению объектов на изображении.

После разработки алгоритмов обучения получаемые модели стали использовать в практических целях: в задачах прогнозирования, для распознавания образов, в задачах управления и др.

Для решения задачи определения основных характеристик почерка человека использована многослойная сверточная нейронная сеть.

## 1.2 Обучение с учителем

Задача определения основных характеристик почерка человека относится к задаче классификации.

Для распознавания образов и их классификации используют нейросети, обучающиеся с учителем, обучающиеся без учителя и со смешанным обучением. Мы использовали обучение с учителем. Рассмотрим его подробнее.

Обучение с учителем предполагает, что для каждого входного вектора существует целевой вектор, представляющий собой требуемый выход. Вместе они называются обучающей парой. Обычно сеть обучается на некотором числе таких обучающих пар. Предъявляется выходной вектор, вычисляется выход сети и сравнивается с соответствующим целевым вектором. Далее веса изменяются в соответствии с алгоритмом, стремящимся минимизировать ошибку. Векторы обучающего множества предъявляются последовательно, вычисляются ошибки и веса подстраиваются для каждого вектора до тех пор, пока ошибка по всему обучающему массиву не достигнет приемлемого уровня.

## 1.3 Алгоритм работы нейронной сети

Алгоритм работы нейросети можно представить следующим образом:

1. Инициализация весов связей между нейронами случайными значениями.

2. Вычисление ошибки для выходного слоя, т.е. общей ошибки сети, при текущих значениях весов.

3. Если ошибка = 0 или веса перестали изменяться, т.е. сеть больше обучаться не будет, то конец обучения.

4. Случайным образом выбираем батч (небольшое подмножество) из обучающей выборки.

5. Прямой проход (вычисление значений состояний нейронов всех слоев).

6. Обратный проход (вычисление значений функции потерь для скрытых слоев).

7. Обнуляем старые градиенты и вычисляем новые для каждого слоя.

8. Считаем изменения весов и изменяем веса.

9. Возврат к пункту 2.

Первый слой нейронной сети всегда фиктивный (ничего не делает, только принимает значения). Последний слой содержит n нейронов, выдающих вероятности принадлежности к каждому из классов, на которые наша сеть должна классифицировать объекты.

Функция активации нейрона – мера того, насколько положительна соответствующая взвешенная сумма. Вычислив взвешенную сумму, мы получаем любое число в широком диапазоне значений. Функция активации сжимает этот диапазон и определяет, насколько большое значение будет у нейрона.

Для минимизации функции потерь используют градиентный спуск. Он находит локальный минимум функции потерь с помощью движения вдоль градиента, причем движение должно осуществляться в сторону антиградиента.

Если изменять веса, просто сдвигая их текущие значения против значения градиента, то такой подсчет изменения весов будет застревать в локальных минимумах, кроме того, градиентный спуск будет слишком долго сходиться в минимум функции, а еще сеть может переобучиться. Чтобы избежать этих ошибок, нужно добавить к обычному градиентному спуску метод моментов (для лучшего схождения к минимуму функции), применить регуляризацию (ограничение на веса ­=> борьба с переобучением) и разбить обучающую выборку на батчи.

Описанная часть нейронной сети отвечает за предсказание нужного класса и называется classificator. До перехода к ней изображения проходят через сверточные слои нейронной сети (М блоков сверток и pooling-ов). Эта часть нейросети называется feature extractor. Она отвечает за формирование новых, более сложных признаков, поверх тех, которые подаются.

На рис.2 изображена общая схема работы нейронной сети:

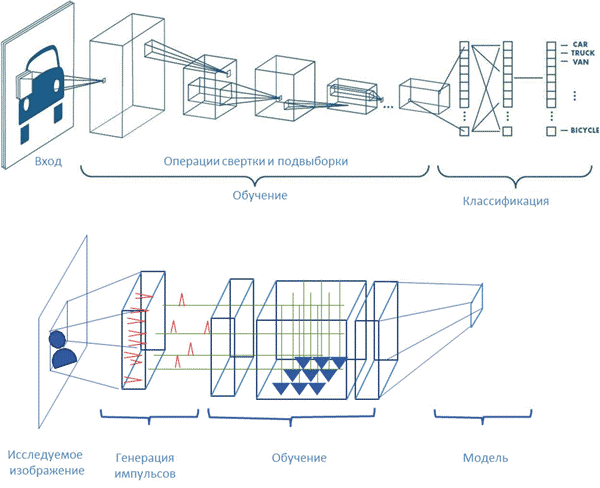


Рис.2. Общая схема работы нейронной сети

# 2. Графология

## 2.1 Определение графологии

Один из самых точных способов определить характер человека — анализ его почерка. Эффективность этого метода многократно подтверждена практикой (в частности, используется в криминалистике при определении или подтверждении авторства документа).

Графология - наука о почерке человека, изучение проявлений в почерке индивидуальных особенностей человеческой личности. Можно выделить основные характеристики почерка: размер букв, поля, наклон букв, форма букв(округлость/угловатость), расстояние между словами, расстояние между строками.

## 2.2 Размер букв

Размер букв – это первый важный момент в изучении почерка. Самый популярный размер при написании - средний, поэтому он не несёт какой-то специфической информации. ​Буквы, написанные большим шрифтом без эмоциональной и физической напряжённости, говорят о лидерских качествах человека, его эмоциональной устойчивости, открытости и лёгкости налаживать взаимоотношения с людьми. Маленький шрифт свидетельствует о том, что человек способен долго просидеть на одном месте при выполнении работы. У таких людей заметна целеустремлённость и сдержанность, но они не приживаются в обществе и являются замкнутыми личностями.

## 2.3 Наклон букв

Почерк, наклоненный вправо, говорит об эмоциональности, импульсивности, чувствительности к эмоциональным состояниям других людей. Такие люди открытые, общительные, целеустремленные и настойчивые. Человек с таким почерком более восприимчив к новым идеям и начинаниям и отличается большой жизненной энергией.

Левый наклон свидетельствует о том, что человек повернут скорее внутрь себя, ему бывает трудно адаптироваться к меняющейся внешней среде, поэтому он предпочитает жить в своем внутреннем мире. При внешней холодности и отрешенности эти люди способны к очень сильным внутренним переживаниям. Наклон влево чаще всего встречается у индивидуалистов: собственные интересы они всегда ставят выше групповых. Такие люди отличаются критическим складом ума.

Вертикальный почерк присутствует у человека, обладающего высокой степенью самоконтроля. Он независим, и, скорее всего, не любит работать в команде. Это сильный человек, полагающийся только на себя. Также это говорит о внутренней гармонии, балансе рациональности и эмоциональности одновременно. Человек пишущий так, очень хорошо может продумывать свои действия, оценивать ситуацию и действовать наверняка и эффективно.

## 2.4 Расстояние между строками

Достаточно большой промежуток между строками говорит о внимательности и сознательности человека. Средний интервал — свидетельство организованности и ясности ума, а путающиеся строки — признак неорганизованного мышления и эмоционального или умственного замешательства.

## 2.5 Расстояние между словами

Ширина интервала между словами — один из показателей отношения человека к окружающим. Узкий пробел между словами указывает на то, что человеку не требуется много пространства, он может легко ладить с окружающими и чувствовать себя комфортно даже при тесном общении. Человек, оставляющий широкий пробел между словами, предпочитает значительную часть времени проводить в одиночестве и нуждается в обширном личностном пространстве. А умеренно-ровные пробелы — это признак социальной зрелости и трезвого ума.

## 2.6 Форма букв

Чаще встречающиеся округлые буквы в почерке человека говорят о готовности личности к сотрудничеству. Люди, использующие круглый почерк - творческие, склонны смягчать противоречия, они часто идут на компромиссы, редко резко меняют свое мнение или взгляды. Их почерк свидетельствует о природной мягкости и миролюбии. Они редко проявляют инициативу, их больше устраивает активность коллег по работе или друзей.

Угловатый почерк находит на мысль, что человеку свойственно упрямство, резкость, требовательность. Такие люди отличаются целеустремленностью, энергичностью, умением напряженно трудиться; в них чувствуется постоянная готовность к сопротивлению, что придает их манере общения резкость. Угловатый почерк очень часто присутствует у людей с хорошим самообладанием. Такой почерк может также свидетельствовать о стремлении к лидерству, состязательности, умении четко планировать карьеру.

# 3. Реализация

## 3.1 Среда, используемая для разработки нейронных сетей

Для определения каждой характеристики почерка использовалась своя нейронная сеть. Нейронные сети для каждого признака почерка были написаны в Google Colaboratory – облачном сервисе, направленном на упрощение исследований в области машинного и глубокого обучения.

Colaboratory позволяет запускать Jupyter Notebook на удалённой машине. Файлы Colaboratory представляют собой обычные .ipynb «ноутбуки» и хранятся в Гугл-диске. Также в Colab есть набор функций, позволяющий загружать файлы с удаленной машины на Гугл-диск и обратно. В Colaboratory есть практически все необходимые для работы Python-библиотеки. Если какой-то пакет отсутствует, он с легкостью устанавливается через pip.

## 3.2 Описание датасета

Для обучения нейронных сетей был собран тренировочный датасет из 900 и тестовый датасет из 100 фотографий и сканов различных почерков размером 512\*256 пикселей. Разметка изображений проводилась по следующим критериям:

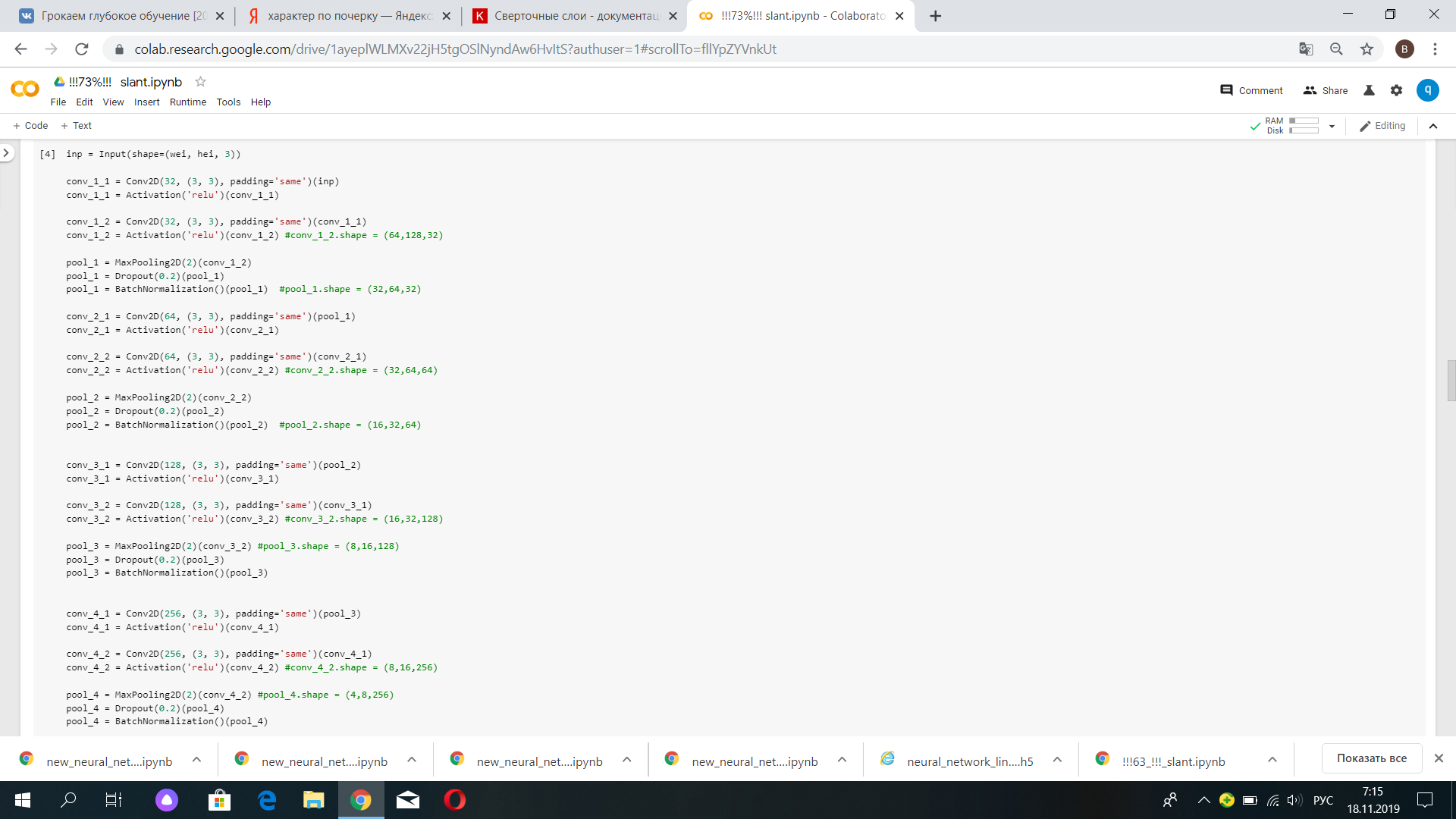
* Размер: относительный размер почерка равный 3 мм считается нормой (1), все, что больше – это крупный почерк (2), меньше – мелкий (0).
* Наклон: влево (0), вертикально и немного вправо — до 20-30 градусов (1), сильно вправо — 50-60 градусов (2).
* Расстояние между строками: малый размер интервала – пространство равно по высоте размеру одной буквы или меньше ее (0); средний размер интервала – пространство равно по высоте размеру до двух букв (1); большой размер интервала – пространство равно по высоте размеру двух и более букв (2).
* Расстояние между словами: малый размер интервала – пространство равно по ширине размеру одной буквы или меньше ее (0); средний размер интервала – пространство равно по ширине размеру двух букв (1); большой размер интервала – пространство равно по ширине размеру большему, чем две буквы (2).
* Форма букв: округлые (2), обычные (1), угловатые (0).

## 3.3 Описание работы нейронной сети

Написанная в ходе работы сверточная нейронная сеть содержит 4 сверточных слоя, содержащих 2 свертки и pooling, 4 скрытых слоя классификации и выходной (входной слой не учитывается). Скрытые слои содержат 128, 256, 512 и 64 нейрона соответственно, выходной – 3 (по количеству классов для каждой характеристики почерка). Используема функция активации после операций свертки и в части классификации – Relu. Функция активации для выходного слоя – Softmax. К полученной нейронной сети были добавлены Dropout и BatchNormalization. В качестве функции потерь была выбрана Кросс-энтропия, а в качестве метода оптимизации – Adam.

Каждая нейросеть обучалась на датасете из 900 изображений почерков, размером 128\*64 (проводилось сжатие). Тестировалась на 100 изображениях такого же типа. По окончании обучения точность каждой нейросети на тренировочной выборке составила от 81 до 99,5%. Точность определения размера букв составила 70% на тестовой выборке; точность определения наклона букв - 73%; точность определения интервалов между строками - 61%; точность определения интервалов между словами - 57%; точность определения формы букв - 56%.

Код архитектуры нейронной сети представлен на рис.3:



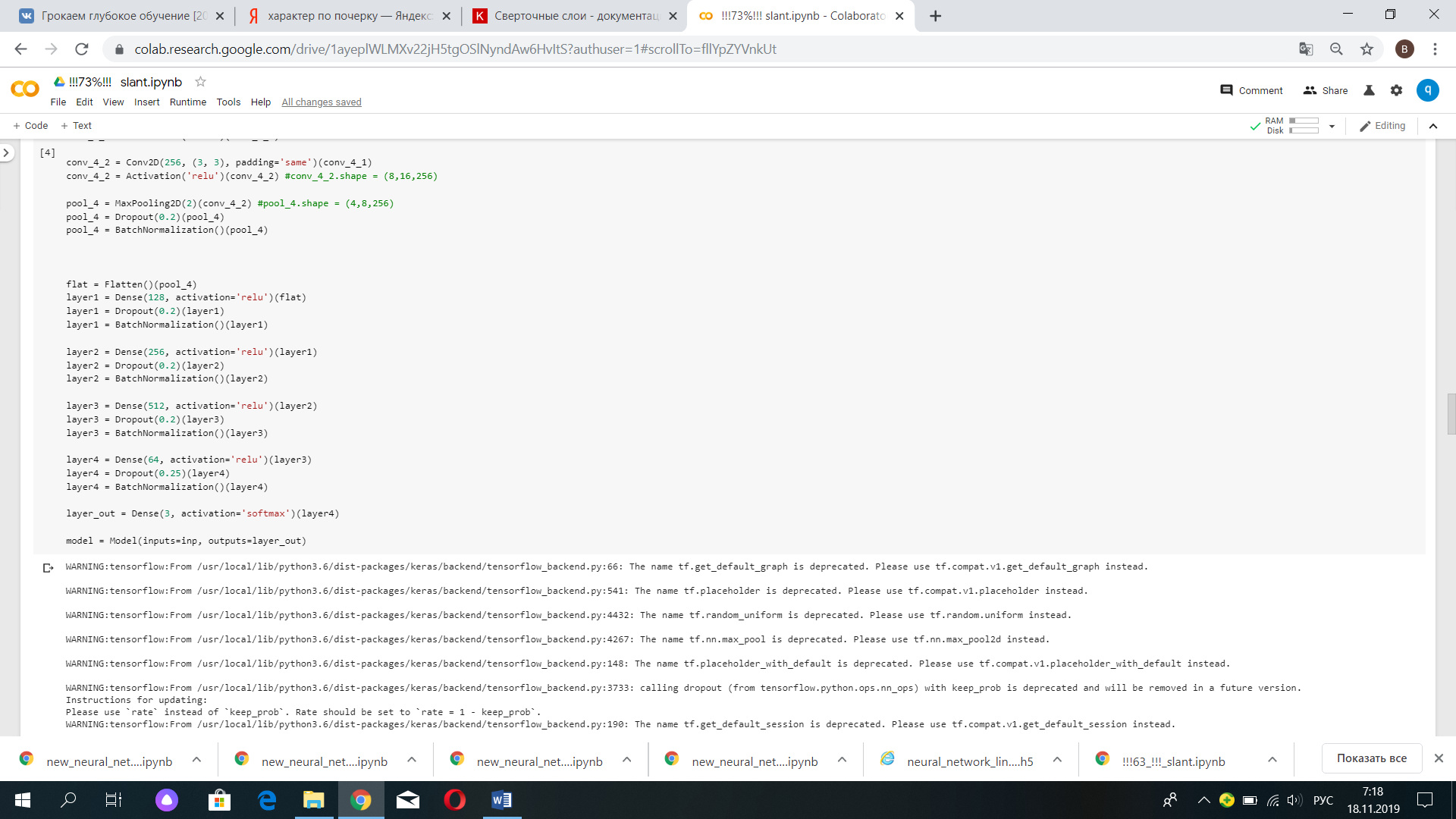


Рис.3. Код архитектуры нейронной сети

## 3.4 Статистика определения характеристик почерка

На диаграмме 1 представлен процент верного определения различных характеристик почерка нейронными сетями на тестовой выборке.

Диаграмма 1. Процент верного определения различных характеристик почерка

Для увеличения процента верного определения характеристик почерка необходимо расширить обучающую выборку.

# Заключение

В ходе работы, анализируя информацию, полученную из различных источников, были изучены искусственные нейронные сети, проанализированы возможности их применения, исследована структура и принцип работы линейных и сверточных нейронных сетей. Был сделан вывод о возможности использования сверточных нейронных сетей для анализа почерка человека.

Данная тема не рассматривалась ранее из-за относительной новизны машинного обучения, однако она достаточно актуальна, например, для криминалистических экспертиз.

Для решения поставленной задачи была проанализирована связь основных характеристик почерка, таких как размер и наклон букв, расстояние между строками и прочее, с характером человека. Был собран и обработан датасет, состоящий из 900 фотографий и сканов различных почерков. Используя такие методы исследования, как моделирование и эксперимент, были обучены 5 нейронных сетей (на каждую характеристику почерка) и написан код, позволяющий определять характер человека по имеющимся знаниям о его почерке.

Результатом работы является программа, состоящая из нескольких предобученных сверточных нейронных сетей. Точность определения размера букв составила 99% на тестовой выборке, 70% - на тренировочной; точность определения наклона букв - 81% на тестовой выборке, 73% - на тренировочной; точность определения интервалов между строками - 98% на тестовой выборке, 61% - на тренировочной; точность определения интервалов между словами - 99% на тестовой выборке, 57% - на тренировочной; точность определения формы букв - 98% на тестовой выборке, 56% - на тренировочной. По имеющимся данным о почерке человека составляется его личностная характеристика.

В дальнейшем планируется доработка данной программы для увеличения точности определения характеристик почерка, а именно расширение базы данных, увеличение количества определяемых параметров почерка, улучшение архитектуры нейронных сетей. На основе полученных результатов планируется разработать приложение, которое будет по написанному тексту рассказывать о характере человека.

При более детальной доработке программы ее можно будет использовать в криминалистике для упрощения проведения почерковедческих экспертиз.

# Список использованных источников и литературы:

1. Аксенов С.В., Новосельцев В.Б. Организация и использование нейронных сетей. Томск: Томский политехнический университет, 2006. 128 с.
2. Градиентный спуск. [Электронный ресурс] // Neurohive.

URL: <https://neurohive.io/ru/osnovy-data-science/gradient-descent/>. (Дата обращения: 28.03.2019)

1. И ещё раз про распознавание номеров. [Электронный ресурс] // Хабр.

URL: <https://m.habr.com/ru/company/recognitor/blog/243919/>. (Дата обращения: 24.03.2019)

1. Каллан Р. Основные концепции нейронных сетей / Пер. с англ. М.: Вильямс, 2001. 287 с.
2. Криминалистическое исследование почерка. [Электронный ресурс] // Криминалистика.

URL: https://crimlib.info/Криминалистическое\_исследование\_почерка. (Дата обращения: 12.11.2019)

1. Курс Deep Learning. [Электронный ресурс] // Deep Learning School.

URL: <https://www.dlschool.org/main-track>. (Дата обращения: 11.10.2019)

1. Нейронные сети для начинающих. Часть 1. [Электронный ресурс] // Хабр.

URL: <https://habr.com/post/312450>. (Дата обращения: 28.11.2018)

1. Нейронные сети для начинающих. Часть 2. [Электронный ресурс] // Хабр.

URL: <https://habr.com/ru/post/313216/>. (Дата обращения: 07.03.2019)

1. О методах обучения многослойных нейронных сетей прямого распространения. [Электронный ресурс] // Дом-страница Евгения Сергеевича Борисова.

URL: <http://mechanoid.kiev.ua/neural-net-backprop2.html>. (Дата обращения: 18.03.2019)

1. Рашид Т. Создаем нейронную сеть / Пер. с англ. М.: Вильямс, 2018. 272 с.
2. Самоучитель Python. [Электронный ресурс] // Python 3 для начинающих.

URL: <https://pythonworld.ru/samouchitel-python>. (Дата обращения: 21.03.2019)

1. Свёрточная нейронная сеть. [Электронный ресурс] // Википедия.

URL: <https://ru.wikipedia.org/wiki/Свёрточная_нейронная_сеть>. (Дата обращения: 07.04.2019)

1. Сверточная нейронная сеть. [Электронный ресурс] // Хабр.

URL: <https://habr.com/ru/post/348028/>. (Дата обращения: 28.03.2019)

1. Форма букв. [Электронный ресурс] // ВикиЧтение.

URL: <https://it.wikireading.ru/43480>. (Дата обращения: 10.11.2019)