Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное

образовательное учреждение высшего образования

«Сыктывкарский государственный университет имени Питирима Сорокина»

(ФГБОУ ВО «СГУ им. Питирима Сорокина»)

Институт точных наук и информационных технологий

Кафедра информационных систем

Курсовая работа по дисциплине «Информационные системы»

**Использование нейронных сетей для классификации изображений**

Направление подготовки

09.03.03 Прикладная информатика

Направленность (профиль) программы

Прикладная информатика в экономике

Исполнитель:

Гончаров Игорь Валерьевич \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Личная подпись

Научный руководитель:

Канд. педагогических наук, доцент

Бабикова Надежда Николаевна \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Личная подпись

Сыктывкар

2022

**Оглавление**

[Аннотация 3](#_Toc73879148)

[Введение 4](#_Toc73879149)

[Анализ предметной области 5](#_Toc73879150)

[1.1 Понятие нейронной сети 5](#_Toc73879151)

[Искусственные нейронные сети и их составляющие 5](#_Toc73879151)

[Перцептрон 10](#_Toc73879155)

[Сигмоидальный нейрон 10](#_Toc73879155)

[Активационная функция 10](#_Toc73879155)

[Архитектура нейронных сетей 6](#_Toc73879152)

[Глубокие нейронные сети 10](#_Toc73879155)

[1.2 Обучение нейронных сетей и связанные с этим проблемы 5](#_Toc73879151)

[1.3 Применение сверточных нейронных сетей для распознавания изображений 5](#_Toc73879151)

[Рекурсивная и рекуррентная нейронные сети 10](#_Toc73879155)

[1.4 Методы классификации 5](#_Toc73879151)

[1.5 Обзор аналогов 5](#_Toc73879151)

[Проектирование системы 11](#_Toc73879156)

[2.1 Постановка задачи 11](#_Toc73879157)

[Функциональные требования к системе 10](#_Toc73879155)

[Алгоритм обучения нейронной сети 10](#_Toc73879155)

2.2 Показатели качества нейронной сети 14

[2.3 Обзор средств разработки 14](#_Toc73879158)

[Средства разработки 10](#_Toc73879155)

[Реализация системы 24](#_Toc73879159)

[3.1 Средства реализации 24](#_Toc73879160)

[3.2 Обучающая выборка 30](#_Toc73879161)

[3.3 Тестирование системы 43](#_Toc73879162)

[Заключение 46](#_Toc73879163)

[Список источников 47](#_Toc73879164)

# **Аннотация**

Целью данной курсовой работы является создание нейронной сети, решающей задачу классификации изображений определенного типа.

В курсовой работе были изложены основные проблемы, связанные с

данной темой и рассмотрены методы решения этих проблем. Были изучены основные принципы работы нейронных сетей.

Для реализации системы было решено использовать IDE PyCharm, язык Python, фреймворк TensorFlow.

Ключевые слова:

# **Введение**

Объем информации в интернете с каждым годом увеличивается практически экспоненциально. Возникает необходимость обработки большого объема данных. На текущее время самый оптимальный способ – создание нейронной сети. Нейросетями обрабатывается любая информация, от графической до огромных массивов данных. В представленной работе рассматривается применение нейронной сети для выявления пневмонии по рентгенографии грудной клетки.

Объектом исследования являются методы на основе нейронных сетей

для классификации изображений. Предмет исследования – изучение методов создания нейронных сетей.

Целью работы является создание нейронной сети и ее обучение на подготовленном наборе типовых изображений, для дальнейшего ее использования в качестве инструмента классификации изображений.

**Задачи исследования:**

* Анализ предметной области.
* Проектирование нейронной сети.
* Создание нейронной сети на Python при помощи TensorFlow.
* Обучение сети и оценка результатов.

# **Анализ предметной области**

## **1.1 Понятие нейронной сети**

### **1.1.1 Искусственные нейронные сети и их составляющие**

Структура нейронной сети пришла в программирование из биологии. Нейронная сеть — это последовательность нейронов, соединенных между собой синапсами.

Нейронная сеть (также искусственная нейронная сеть, ИНС) представляет собой систему соединённых и взаимодействующих между собой простых процессоров (искусственных нейронов). Нейрон — это вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше. Они делятся на три основных типа: входной (синий), скрытый (красный) и выходной (зеленый):



Рисунок 1 – схема простой нейронной сети

В том случае, когда нейросеть состоит из большого количества нейронов, вводят термин слоя. Входной слой получает информацию, n скрытых слоев, которые ее обрабатывают и выходной слой, который выводит результат. У каждого из нейронов есть два основных параметра: входные данные и выходные данные.

Возможность обучения — одно из главных преимуществ нейронных сетей перед традиционными алгоритмами. Технически обучение заключается в нахождении коэффициентов связей(синапсов) между нейронами.

У синапсов есть только один параметр — вес. Умножаясь на вес, входная информация изменяется, когда передается от одного нейрона к другому. Связи с положительным весом называются возбуждающими, а с отрицательным — тормозящими.

В процессе обучения нейронная сеть способна выявлять сложные зависимости между входными данными и выходными, а также выполнять обобщение. Это значит, что в случае успешного обучения сеть сможет вернуть верный результат на основании данных, которые отсутствовали в обучающей выборке.

Нейронные сети используются для решения сложных задач, которые требуют аналитических вычислений подобных тем, что делает человеческий мозг. Самыми распространенными применениями нейронных сетей является:

* Классификация
* Предсказание
* Распознавание

Классификация – распределение данных по параметрам. В нашем случае –определить по рентгеновскому снимку здоровые легкие у человека или нет.

Предсказание – возможность предсказывать развитие событий. Например, рост или падение акций, основываясь на ситуации на фондовом рынке.

Распознавание – в настоящее время, самое широкое применение нейронных сетей. Используется для определения объектов на изображении, например, выделение людей и животных на фотографии.

### **1.1.2 Активационная функция нейрона**

В теории нейронных сетей активационной называется функция, аргументом которой является взвешенная сумма входов искусственного нейрона, а значением — выход нейрона:

(1.1)

(1.2)

где:

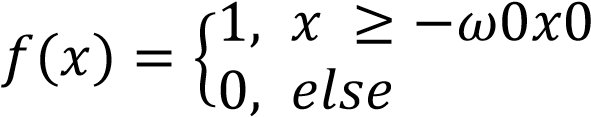
* x — взвешенная сумма входов нейрона;
* N — число входов нейрона;
* wi — вес i-го входа нейрона;
* si — значение, поступающее по i-му входу;
* f(x) — активационная функция;
* y — выходное значение нейрона (и, соответственно, активационной функции).

От вида и формы используемой активационный функции зависит выбор алгоритма обучения сети, а также качество ее обучения на конкретном обучающем множестве. Параметры активационной функции подбираются экспериментально в процессе обучения.

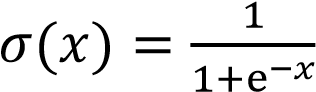
Обычно функция является монотонно возрастающей и находится в области значений [-1,1] (гиперболический тангенс) и [0,1] (сигмоида).

Основными активационными функциями являются:

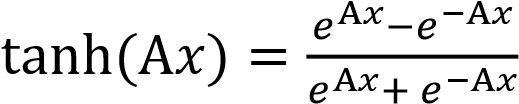
* Пороговая активационная функция (функция Хевисайда). Нельзя использовать для алгоритма обратного распространения ошибки;

 (1.3)

* Сигмоидальная активационная функция;

 (1.4)

* Гиперболический тангенс.

 (1.5)

### **1.1.3 Модели нейронов**

#### **Перцептрон**

Перцептрон — простейший вид нейронных сетей. В основе лежит математическая модель восприятия информации мозгом, состоящая из сенсоров, ассоциативных и реагирующих элементов. В самом общем своем виде он представляет систему из элементов трех разных типов: сенсоров, ассоциативных элементов и реагирующих элементов.

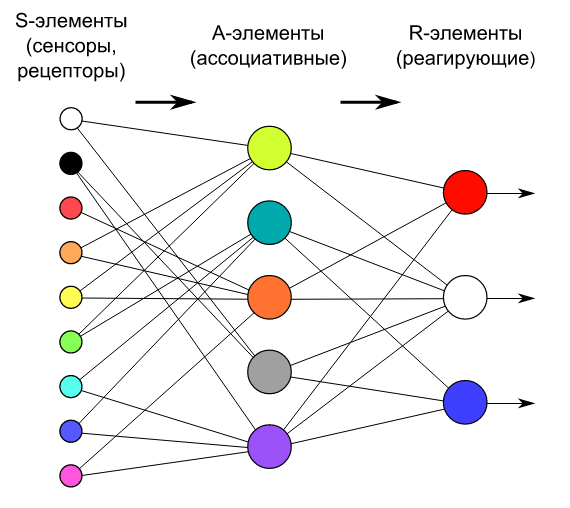


Рисунок 2 – схема перцептрона

Принцип работы перцептрона:

S-элементы находятся либо в состоянии покоя (сигнал равен 0), либо в состоянии возбуждения (сигнал равен 1). Далее сигналы от S-элементов передаются A-элементам по S-A связям. Эти связи могут иметь веса, равные только -1, 0 или 1. Затем сигналы от сенсорных элементов, прошедших по S-A связям попадают в A-элементы, которые еще называют ассоциативными элементами. Стоит заметить, что одному A-элементу может соответствовать несколько S-элементов. Если сигналы, поступившие на A-элемент, в совокупности превышают некоторый его порог, то этот A-элемент возбуждается и выдает сигнал, равный 1. В противном случае (сигнал от S-элементов не превысил порога A-элемента), генерируется нулевой сигнал.

Классификация персептронов:

*Персептрон с одним скрытым слоем.* Персептрон, у которого имеется только по одному слою S, A и R элементов.

*Однослойный персептрон.* Каждый S-элемент однозначно соответствует одному A-элементу, все S-A связи имеют вес, равный +1, порог A элементов равен 1.

*Многослойный персептрон.* Под многослойным персептроном понимают два разных вида: многослойный персептрон по Розенблатту и многослойный персептрон по Румельхарту.

* Многослойный персептрон по Розеблатту — персептрон, у которого имеется более 1 слоя А-элементов.
* Многослойный персептрон по Румельхарту — многослойный персептрон по Розенблатту, у которого обучению подлежат еще и S-A связи, а также само обучение производится по методу обратного распространения ошибки.

Даже небольшое изменение весов или смещения одного из перцептронов сети может кардинально изменить выходное значение, например, с 0 на 1. Поэтому в современных работах чаще всего используют другую модель искусственного нейрона — сигмоидальный нейрон.

#### **Сигмоидальный нейрон**

Сигмоидальные нейроны похожи на перцептроны, однако небольшие

изменения в их весах и смещениях незначительно изменяют выход нейрона. Это достигается путем использования неразрывной сигмоидальной функции активации.

Благодаря этому сеть из сигмоидальных нейронов может обучаться. На вход сигмоидального нейрона подаются любые значения между 0 и 1. На выходе также выдаётся значение между 0 и 1.

На практике используются как униполярные, так и биполярные функции активации.

Униполярная функция, как правило, представляется формулой

(1.6)

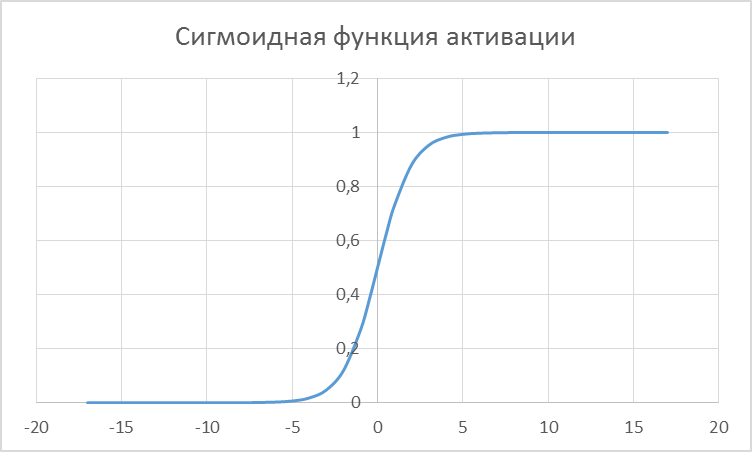


Рисунок 3 – Сигмоидальная функция активации

Чем больше β (параметр наклона сигмоидальной функции активации), тем сильнее крутизна графика. При β → ∞ сигмоидальная функция превращается в функцию ступенчатого типа, идентичную функции активации персептрона.

Важным свойством сигмоидальной функции является её дифференцируемость.

Применение непрерывной функции активации позволяет использовать при обучении градиентные методы.

### **1.1.4 Архитектура нейронных сетей**

#### **Сеть прямого распространения**

Для решения задачи классификации (обучения с учителем), нейросеть получает на вход множество тренировочных примеров X с метками Y (labels).

Можно представить классическую нейросеть в виде вычислительного графа содержащего:

* входные вершины x;
* вершины, являющиеся нейронами со значениями их выхода a;
* вершины, отвечающие за bias b;
* ребра, умножающие значения выхода предыдущего слоя на соответствующие им коэффициенты матрицы весов w;
* гипотезу hw,b(x) — результат выхода последнего слоя.

Нейронная сеть в общем случае строится как соединение множества нейронов, объединенных в слои так, что выходы одного слоя являются входами следующего:



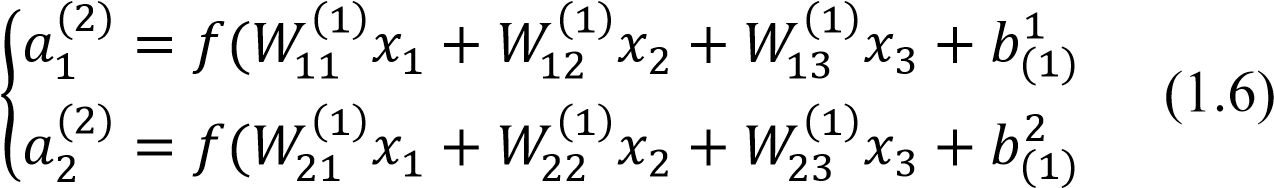
Рисунок 4 – схема сети прямого распространения

Самый левый слой сети называется входным, самый правый — выходным (на рисунке он состоит из одного нейрона), остальные слои называют скрытыми, потому что их значения отсутствуют в обучающем наборе. Таким образом, данная сеть содержит 3 входных нейрона, 3 скрытых и 1 выходной.

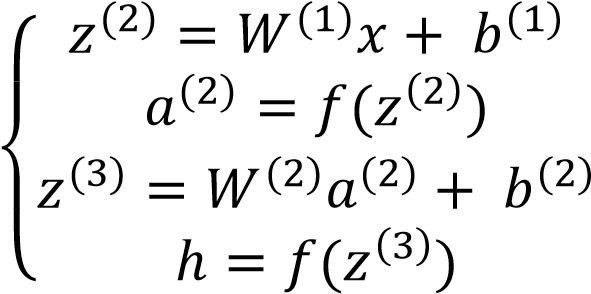
Нейрон смещения (bias нейрон) — его вход и выход всегда равняются 1 и они никогда не имеют входных синапсов. Соединения у нейронов смещения такие же, как у обычных нейронов — со всеми нейронами следующего уровня, за исключением того, что синапсов между двумя bias нейронами быть не может. Следовательно, их можно размещать на входном слое и всех скрытых слоях, но никак не на выходном слое, так как им попросту не с чем будет формировать связь. Нейрон смещения нужен для того, чтобы иметь возможность получать выходной результат, путем сдвига графика функции активации вправо или влево.

Нейросеть параметризуется значениями (*W,b*) = (*W*(1)*,b*(1)*,W*(2)*,b*(2)*,...*), где под понимается параметр, или вес, который отвечает соединению между *j*−м нейроном в слое *l* и *i*−м нейроном в слое *l* + 1. За  обозначается смещение или вес, который связан с константными единичными входами на каждом слое сети. Таким образом на рисунке выше *W*(1) ∈ R3×3 и *W*(2) ∈ R1×3.

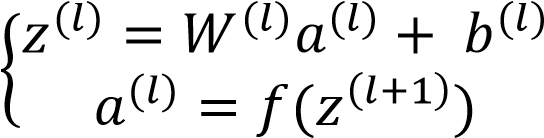
Результат применения функции активации обозначается ai для i-ого элемента. Получаем такую систему:

 (1.7)

Обозначив функцию суммирования через z, получим:

 (1.8)

Общая формула будет выглядеть таким образом:

 (1.9)

Эти преобразования называются прямым проходом или прямым распространением (forward propagation).

Сетью прямого распространения называются нейронные сети, которые используют выход одного слоя в качестве входных данных для следующего слоя.

У сети прямого распространения есть существенный недостаток – слишком много параметров. Например, нейросеть из 3 скрытых слоев, которой нужно обрабатывать картинки 100\*100 пикселей на входе будет иметь 10 000 пикселей, и они заводятся на 3 слоя. Иными словами, каждый нейрон такой сети получает на вход все пиксели изображения. В конечном итоге такая нейросеть будет иметь порядка миллиона параметров, т.е. классифицировать примеры по миллиону признаков.

Данный недостаток сетей прямого распространения исправлен в сверточных нейросетях.

#### **Сверточная сеть**

Сверточные нейронные сети (СНС) – класс нейронных сетей, специализирующихся на обработке данных, предствленных в виде матрицы (например, изображения).

СНС объединяют три архитектурных идеи, для обеспечения инвариантности к изменению масштаба, повороту сдвигу и пространственным искажениям:

* локальные рецепторные поля (обеспечивают локальную двумерную связность нейронов);
* общие коэффициенты нейронов (обеспечивают детектирование некоторых черт в любом месте изображения и уменьшают общее число весовых коэффициентов);
* иерархическая организация с пространственными подвыборками.

Как было сказано ранее, в сети прямого распространения каждый нейрон связан с каждым пикселем изображения, в то время как в сверточной сети происходит разбиение – каждый нейрон связан только с частью изображения. Достигается это посредством свертки изображения:



Рисунок 5 – свертка изображения

Происходит свертка посредством пропуска изображения через ядро свертки (Рис. 5 – convolution filter). Ядро свертки – это совокупность весов данного нейрона. Применяется ядро свертки на всех пикселях изображения последовательно. Ядро представляет из себя некий фильтр, который проходится по всей области предыдущей карты и находит определенные признаки объектов. Размер ядра обычно берется в пределах от 3х3 до 7х7. Если размер ядра маленький, оно не сможет выделить какие-либо признаки, если слишком большое, увеличится количество связей между нейронами.

В зависимости от метода обработки краев исходной матрицы результат может быть меньше исходного изображения (valid), такого же размера (same) или большего размера (full).

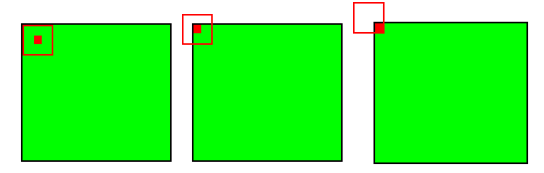


Рисунок 6 – Виды свертки



Рисунок 7 – Структура СНС

СНС состоит из нескольких видов слоев: сверточные (convolutional) слои, подвыборочные (subsampling) слои и слои нейронов (например, перцептронов).

Как видно из структуры, свертка изображения происходит несколько раз, формируя входной вектор признаков для многослойного перцептрона.

Сверточный слой представляет из себя набор карт (матриц), у каждой карты есть синаптическое ядро, о нем уже было сказано выше. Размер ядра выбирается таким, чтобы размер карт сверточного слоя был четным, это позволяет не терять информацию при уменьшении размерности в подвыборочном слое.

Размер всех карт сверточного слоя одинаков, а их количество карт определяется требованиями к задаче. Большое количество карт повышает качество распознавания, но увеличивает сложность вычислений. Часто на практике используется соотношение один к двум, то есть каждая карта предыдущего слоя связана с двумя картами сверточного слоя.

Разделяемые веса – это одна из главных особенностей сверточной нейросети. В обычной многослойной сети очень много связей между нейронами, что сильно замедляет процесс детектирования. В сверточной сети – наоборот, общие веса позволяет сократить число связей и позволить находить один и тот же признак по всей области изображения.

Вставить картинку

Изначально значения каждой карты сверточного слоя равны 0. Значения весов ядер задаются случайным образом в области от -0.5 до 0.5.

Подвыборочный слой также, как и предшествующий сверточный имеет карты, но их количество совпадает с предыдущим слоем. Цель этого слоя – уменьшение размерности карт предыдущего. Если на предыдущей операции свертки уже были выявлены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уплотняется до менее подробного. Кроме того, фильтрация уже ненужных деталей помогает не переобучаться.

Обычно, каждая карта имеет ядро размером 2x2, что позволяет уменьшить предыдущие карты сверточного слоя в 2 раза. Вся карта признаков разделяется на ячейки 2х2 элемента, из которых выбираются максимальные по значению (операция MaxPooling):



Рисунок 8 – Преобразование карты светрочного слоя ядром подвыборочного слоя

Определение топологии сети ориентируется на решаемую задачу. Можно выделить следующие действия, влияющие на выбор топологии:

* определение решаемой задачи (классификация, прогнозирование, модификация);
* определение ограничений в решаемой задаче (скорость, точность ответа);
* определение входных и выходных данных (входные: изображение, звук, размер: 100x100, 30x30, формат: RGB, в градациях серого, выходные данные – количество классов)

На данный момент сверточная нейронная сеть и ее модификации считаются лучшими по точности и скорости алгоритмами нахождения объектов на изображении.