**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ**

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ**

**Кафедра математической кибернетики**

Апочкина Елена Федоровна

**МЕТОДЫ СПЕКТРАЛЬНОГО АНАЛИЗА ГОЛОСОВЫХ СИГНАЛОВ ДЛЯ СИСТЕМ МЕДИЦИНСКОЙ ДИАГНОСТИКИ**

Дипломная работа

Научный руководитель:

доктор технических наук,

доцент М.И. Вашкевич.

Допущена к защите «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_ 2023 г.

Зав. кафедрой математической кибернетики доктор физико-математических наук, профессор А.Л. Гладков

Минск, 2023

**Оглавление**

[**ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ, СИМВОЛОВ И ТЕРМИНОВ** 3](#_Toc137066249)

[**РЕФЕРАТ ДИПЛОМНОЙ РАБОТЫ** 4](#_Toc137066250)

[**ВВЕДЕНИЕ** 5](#_Toc137066251)

[**ГЛАВА 1 ОБЩИЕ СВЕДЕНИЯ О СПЕКТРАЛЬНОМ АНАЛИЗЕ РЕЧЕВЫХ СИГНАЛОВ ДЛЯ МЕДИЦИНСКИХ ДИАГНОСТИЧЕСКИХ СИСТЕМ** 7](#_Toc137066252)

[**1.1** **Определение спектрального анализа** 7](#_Toc137066253)

[**1.2** **Преимущества и ограничения использования голосовых сигналов для медицинской диагностики** 8](#_Toc137066254)

[**1.3** **Факторы, которые необходимо учитывать при анализе голосовых сигналов** 9](#_Toc137066255)

[**1.4** **Преобразование Фурье** 10](#_Toc137066256)

[**1.5** **Спектрограмма** 12](#_Toc137066257)

[**1.6** **Мел-спектрограмма** 13](#_Toc137066258)

[**ГЛАВА 2 НЕЙРОННАЯ СЕТЬ** 15](#_Toc137066259)

[**2.1. Искусственные нейронные сети** 15](#_Toc137066262)

[**2.1.1. Модель нейрона** 16](#_Toc137066263)

[**2.1.2. Функция активации** 17](#_Toc137066264)

[**2.1.3. Функция потерь** 19](#_Toc137066265)

[**2.2. Сверточные нейронные сети** 21](#_Toc137066266)

[**2.2.1. Архитектура сверточной нейронной сети** 21](#_Toc137066267)

[**2.2.2. Сверточный слой** 22](#_Toc137066268)

[**2.2.3. Подвыборочный слой (субдискретизирующий слой)** 23](#_Toc137066273)

[**2.2.4. Полносвязный слой** 24](#_Toc137066275)

[**2.2.5. DropOut слой** 25](#_Toc137066276)

[**2.3. Алгоритм обучения нейронной сети** 25](#_Toc137066277)

[**ГЛАВА 3 РЕАЛИЗАЦИЯ СПЕКТРАЛЬНОГО АНАЛИЗА СИГНАЛА** 27](#_Toc137066278)

[**ВЫВОД** 35](#_Toc137066279)

[**ЛИТЕРАТУРА** 36](#_Toc137066280)

# **ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ, СИМВОЛОВ И ТЕРМИНОВ**

# **РЕФЕРАТ ДИПЛОМНОЙ РАБОТЫ**

Объем дипломной работы составляет \_\_ страница. Дипломная работа содержит \_\_ рисунок, \_ таблицы, \_ приложения и \_ использованных источников.

Ключевые слова: АНАЛИЗ РЕЧЕВОГО СИГНАЛА, КРИТИЧЕСКИЕ ПОЛОСЫ, МОДУЛЯЦИОННЫЙ СПЕКТР, МОДУЛЯЦИОННЫЕ ПРИЗНАКИ, ИЗВЛЕЧЕНИЕ ПРИЗНАКОВ, НЕРОННАЯ СЕТЬ , СВЕРТОЧНЫЙ СЛОЙ

Предмет исследования: диагностика заболеваний с помощью модуляционного спектра

Целью дипломной работы является вычисление модуляционного спектра, исследование его и реализация анализа на практике с помощью нейронных сетей

Полученные результаты: разработана нейронная сеть для определения патологии голосового сигнала

Подтверждаю самостоятельность выполнения дипломной работы.

# **ВВЕДЕНИЕ**

Спектральный анализ голосовых сигналов стал передовым достижением в области медицинских диагностических систем, существенно повлияв на способы диагностики и лечения некоторых заболеваний. Анализируя частотное распределение и акустические паттерны голосовых сигналов, медицинские специалисты могут обнаружить ключевые биоиндикаторы, указывающие на различные заболевания и проблемы со здоровьем. Спектральный анализ голоса показал многообещающий потенциал для раннего обнаружения, дистанционного мониторинга и оценки состояний, включая неврологические расстройства, психические расстройства и респираторные заболевания. Цель данной статьи - представить обзор важности спектрального анализа голосовых сигналов в медицинских диагностических системах.

1. Раннее выявление заболеваний

Раннее выявление заболеваний имеет решающее значение для оперативного вмешательства и эффективного лечения. Спектральный анализ позволяет медицинским работникам обнаружить тонкие изменения в голосовых сигналах, которые могут свидетельствовать о наличии основного заболевания. Анализируя закономерности в распределении частот, спектральный анализ позволяет выявить отклонения, которые могут потребовать дальнейшего изучения, что может привести к быстрой диагностике и спасению жизни.

2. Неврологические расстройства

Неврологические расстройства, такие как болезнь Паркинсона, рассеянный склероз (РС), могут изменять речевые паттерны человека. Анализ голоса может помочь обнаружить эти изменения, которые часто не видны с помощью традиционных методов диагностики. Спектральный анализ голосовых сигналов позволяет оценить двигательную функцию и предложить объективные измерения, которые позволяют отследить прогрессирование этих расстройств и определить эффективность вмешательств.

3. Мониторинг психического здоровья

Проблемы психического здоровья, такие как клиническая депрессия, тревожность и биполярное расстройство, могут существенно повлиять на уникальные особенности речи человека. Спектральный анализ может быть особенно полезен для наблюдения за пациентами с течением времени, предлагая неинвазивный метод выявления ранних признаков психических расстройств. Выявление изменений в речевых паттернах может помочь врачам обеспечить своевременное и целенаправленное лечение психических расстройств, снижая риск рецидива и госпитализации.

4. Респираторные заболевания

Спектральный анализ голоса также продемонстрировал потенциал для выявления и мониторинга респираторных заболеваний, таких как хроническая обструктивная болезнь легких (ХОБЛ), астма и пневмония. Анализируя голосовые сигналы пациентов, медицинские работники могут оценить вибрацию голосовых связок, характер дыхания и воздушный поток, что дает ценную информацию о состоянии здоровья пациента и помогает своевременно принять меры.

5. Неинвазивный удаленный мониторинг

Одним из значительных преимуществ спектрального анализа голоса в медицинской диагностике является его неинвазивный характер. Традиционные диагностические инструменты часто требуют инвазивных процедур, вызывая дискомфорт и риски, связанные с процедурой. Спектральный анализ голоса, напротив, предлагает удобный и неинвазивный метод диагностики пациента, который можно проводить удаленно. Эта функция помогает в долгосрочном наблюдении за состоянием, позволяя медицинским работникам постоянно оценивать прогресс пациента, не причиняя ему дискомфорта.

6. Персонализированное лечение

Данные, полученные в результате спектрального анализа голосовых сигналов, могут быть использованы для подбора лечения и терапевтических мероприятий для отдельных пациентов. Функция мониторинга в режиме реального времени позволяет медицинским работникам оценивать эффективность лекарств и терапии, корректировать дозировки и изучать альтернативные методы лечения для оптимизации ухода за пациентами.

Нейронные сети все чаще используются в этом процессе благодаря их способности обучаться на основе данных и выявлять сложные закономерности, которые могут быть нелегко обнаружить с помощью традиционных методов.

Использование нейронных сетей в спектральном анализе голосовых сигналов потенциально может повысить точность и надежность медицинских диагнозов, особенно для состояний, которые трудно обнаружить с помощью традиционных методов диагностики. Эта технология способна произвести революцию в области медицинской диагностики и улучшить качество обслуживания пациентов.

# **ГЛАВА 1 ОБЩИЕ СВЕДЕНИЯ О СПЕКТРАЛЬНОМ АНАЛИЗЕ РЕЧЕВЫХ СИГНАЛОВ ДЛЯ МЕДИЦИНСКИХ ДИАГНОСТИЧЕСКИХ СИСТЕМ**

## **Определение спектрального анализа**

Спектральный анализ, также известный как частотный анализ или спектральная декомпозиция, - это математический метод и метод обработки сигналов, используемый для исследования и разложения сигнала или набора данных на частотные компоненты. Этот метод позволяет получить ценные сведения о важных частотах, интенсивности или временных периодах, которые характеризуют основные процессы, закономерности или структуры в данных. Он широко используется в различных областях, включая астрономию, акустику, геофизику, связь, экономику и медицину.

Спектральный анализ предполагает преобразование сигнала во временной области, который представлен как функция времени, в сигнал в частотной области, где он становится функцией частоты. Это преобразование позволяет исследователям анализировать сигнал с точки зрения его частот и амплитуд, обеспечивая более четкое понимание периодической или гармонической природы данных.

Спектральный анализ играет важную роль в целом ряде приложений, таких как:

- выявление периодичности в астрономических данных

- анализ и определение характеристик геологических и сейсмических данных

- оценка и мониторинг состояния механических систем и конструкций с помощью анализа вибрации

- диагностика медицинских состояний с помощью сигналов электроэнцефалограммы (ЭЭГ) или электрокардиограммы (ЭКГ)

- улучшение соотношения сигнал/шум в системах связи и изучение распространения волн

- понимание и прогнозирование экономических или финансовых временных рядов.

## **Преимущества и ограничения использования голосовых сигналов для медицинской диагностики**

Использование голосовых сигналов для медицинской диагностики имеет ряд преимуществ и ограничений.

Преимущества:

1. Неинвазивность: В отличие от традиционных методов диагностики, анализ голоса не требует использования игл, инструментов или физического прикосновения, что делает его неинвазивным методом диагностики.

2. Простота в использовании: Анализ голосового сигнала может быть выполнен любым человеком, имеющим смартфон, микрофон или компьютер, что делает его легкодоступным для пациентов и медицинских работников.

3. Экономическая эффективность: Анализ голосового сигнала стоит относительно недорого по сравнению с традиционными методами диагностики, что делает его экономически выгодной альтернативой для пациентов.

4. Быстрое время реакции: Технология голосовой диагностики позволяет проводить анализ в режиме реального времени, что делает его более быстрым по сравнению с традиционными методами диагностики, которые могут занимать дни или даже недели.

Ограничения:

1. Ограниченная точность: Технология анализа голоса еще недостаточно развита, чтобы обеспечить последовательные и точные результаты для широкого спектра медицинских состояний.

2. Ограниченная сфера применения: Анализ голосового сигнала в настоящее время ограничен определенными медицинскими состояниями и может не подходить для диагностики более сложных или редких заболеваний.

3. Изменчивость качества голоса: Результаты анализа голосового сигнала могут варьироваться в зависимости от таких факторов, как качество голоса, акцент и фоновый шум, что может ограничить точность результатов.

4. Проблемы конфиденциальности: Анализ голосового сигнала предполагает запись и хранение личных данных. Существуют проблемы конфиденциальности, которые необходимо учитывать и решать для обеспечения защиты частной жизни пациентов.

## **Факторы, которые необходимо учитывать при анализе голосовых сигналов**

При анализе голосовых сигналов необходимо учитывать различные факторы. Вот некоторые из этих факторов:

1.Высота тона: Высота голосового сигнала относится к частоте звуковой волны, производимой голосовыми связками. Это важный фактор, который может раскрыть информацию об эмоциональном состоянии говорящего.

2.Интенсивность: Интенсивность голосового сигнала относится к громкости или громкости звуковой волны, производимой голосовыми связками. Она может раскрыть информацию об эмоциональном состоянии говорящего, а также о его физическом состоянии.

3.Продолжительность: Длительность голосового сигнала означает продолжительность звуковой волны, производимой голосовыми связками. Она может раскрыть информацию об эмоциональном состоянии говорящего, а также о его физическом состоянии.

4.Форманты: Форманты - это резонансные частоты, которые производятся голосовым трактом. Они могут раскрыть информацию о возрасте, поле и эмоциональном состоянии говорящего.

5.Просодия: Просодия относится к ритму, мелодии и интонации голосового сигнала. Она может раскрыть информацию об эмоциональном состоянии говорящего, а также о его культурном происхождении и языковых способностях.

6.Содержание речи: Содержание речевого сигнала может раскрыть информацию о психическом состоянии говорящего, его когнитивных способностях и социальном контексте.

7.Фоновый шум: Фоновый шум может помешать анализу речевых сигналов, особенно в шумной обстановке. Важно учитывать уровень фонового шума при анализе голосовых сигналов.

## **Преобразование Фурье**

Существует несколько типов методов спектрального анализа, используемых в медицинских диагностических системах для речевых сигналов. Эти методы используются для извлечения особенностей из речевых сигналов и выявления отклонений в речи, которые могут указывать на медицинское состояние.

Преобразование Фурье - это математическая операция, которая разлагает сложный сигнал на составляющие его частоты. Оно используется в различных областях, таких как обработка сигналов, физика и инженерия.

Преобразование Фурье определяется как:

где f(t) - преобразуемая функция, а F(ω) - преобразованная функция в частотной области.

Дискретное преобразование Фурье — это одно из преобразований Фурье, применяемое в алгоритмах цифровой обработки сигналов (его модификации применяются в сжатии звука в MP3, сжатии изображений в JPEG и др.), а также в других областях, связанных с анализом частот в дискретном (к примеру, оцифрованном аналоговом) сигнале. Пример такого преобразования можно увидеть ниже (Рис.1 и Рис.2).

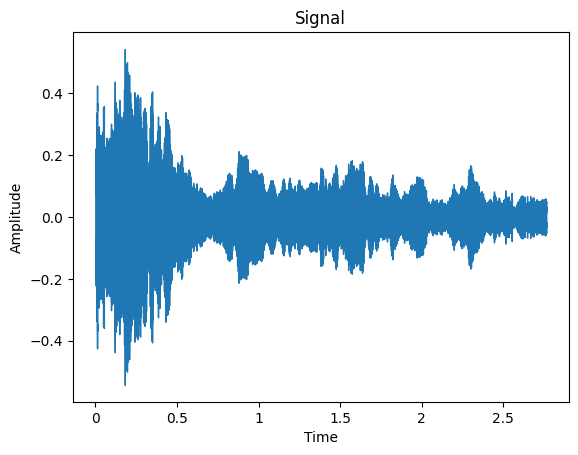


Рис.1 Преобразуемая функция

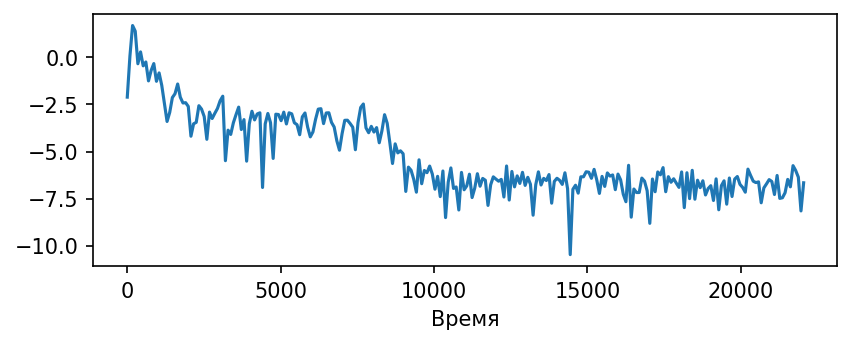


Рис.1 Дискретное преобразование Фурье

Дискретное преобразование Фурье требует в качестве входа дискретную функцию. Такие функции часто создаются путём дискретизации (выборки значений из непрерывных функций).

Прямое преобразование

Обратное преобразование

где:

* *N* – количество значений сигнала, измеренных за период, а также количество компонент разложения;
* — измеренные значения сигнала (в дискретных временных точках с номерами �=0,…,�−1) , которые являются входными данными для прямого преобразования и выходными для обратного;
* — �*N* комплексных амплитуд синусоидальных сигналов, слагающих исходный сигнал; являются выходными данными для прямого преобразования и входными для обратного; поскольку амплитуды комплексные, то по ним можно вычислить одновременно и амплитуду, и фазу;
* обычная (вещественная) амплитуда*�k*-го синусоидального сигнала;
* — фаза �*k*-го синусоидального сигнала (аргумент комплексного числа);
* *k* — индекс частоты. Частота �*k*-го сигнала равна ��*k/T*, где �*T*— период времени, в течение которого брались входные данные.

## **Спектрограмма**

Спектрограмма - это визуальное представление частотного содержания сигнала во времени. Наиболее распространенным представлением спектрограммы является двумерная диаграмма ̶ это двумерный график, на котором ось x представляет время, ось y ̶ частоту, а цвет или яркость каждой точки соответствует интенсивности или мощности сигнала на данной частоте и в данное время.

Частота и амплитуда осей может быть линейными или логарифмическими, в зависимости от того, с какой целью используется график. Аудио обычно может быть представлено с логарифмической осью амплитуды (зачастую, в [децибелах](https://ru.wikipedia.org/wiki/%D0%94%D0%B5%D1%86%D0%B8%D0%B1%D0%B5%D0%BB) или дБ), и частота будет линейной, чтобы подчеркнуть гармонические отношения, или логарифмической, чтобы подчеркнуть музыкальные, тональные отношения.

Формирование спектрограммы с использованием оконного преобразования Фурье обычно осуществляется с помощью методов цифровой обработки (пример: Рис. 2). Во-первых, данные подвергаются цифровой дискретизации во временной области. Затем сигнал разбивается на фреймы, которые обычно перекрываются, и выполняется преобразование Фурье для вычисления величины частотного спектра для каждого сегмента. Каждому разделу соответствует вертикальная линия на полученном изображении, представляющая амплитудное значение как функцию частоты в каждый момент времени. Затем спектры или временные графики сопоставляются на изображении или трехмерной диаграмме.

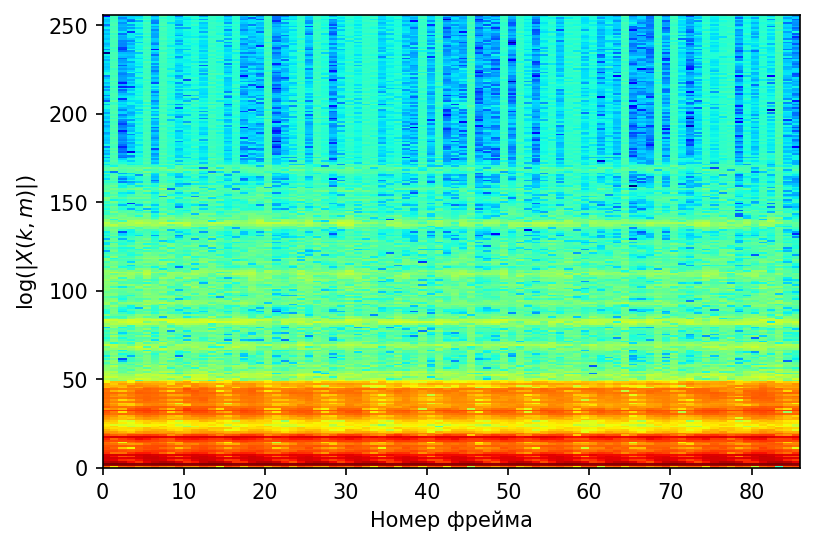


Рис. 2 Спектрограмма с использованием оконного преобразования Фурье

Спектрограммы обычно используются в обработке сигналов, акустике и музыкальном анализе для анализа частотного содержания звукового или музыкального сигнала, а также для выявления закономерностей и изменений в сигнале с течением времени.

Спектрограммы также помогают в извлечении особенностей из сигнала, например, формант при анализе речи, или для идентификации и удаления шума или помех из сигнала.

## **Мел-спектрограмма**

Мел-спектрограмма - это визуальное представление частотного содержания звукового сигнала. Она широко используется в обработке речи и музыки для анализа спектральных характеристик звуковых волн.

Шкала Мела, также известная как шкала частот Мела, является перцептивной шкалой, которая связана с тем, как люди слышат высоту тона. Мел-спектрограмма использует эту концепцию, отображая частоты из линейной шкалы (Герц) в нелинейную, основанную на восприятии шкалу (Мел). Это помогает уловить важную информацию о звуках, которая имеет значение для человеческого восприятия.

Для создания Мел-спектрограммы аудиосигнал сначала делится на небольшие кадры. Затем каждый кадр подвергается преобразованию Фурье для получения его частотного спектра. Затем эти спектры преобразуются в Мел-шкалу с помощью фильтров, разработанных по определенным правилам, основанным на психоакустических принципах. Наконец, логарифмическое масштабирование и нормализация могут быть применены перед отображением в виде изображений или графиков.

Мел-спектрограммы имеют множество применений, таких как распознавание речи, идентификация диктора, классификация музыкальных жанров и многие другие области машинного обучения и искусственного интеллекта, где анализ аудиосигналов играет важную роль.

Основное различие между Мел-спектрограммой и спектрограммой заключается в частотной шкале, используемой для представления аудиосигнала. Спектрограмма представляет частоты аудиосигнала в линейной шкале, в то время как Mel-спектрограмма представляет их в нелинейной, основанной на восприятии шкале, называемой Mel-шкалой.

С технической точки зрения, хотя оба типа спектрограмм используют преобразование Фурье для получения частотных спектров из небольших кадров аудиосигнала, в случае Mel-спектрограмм эти спектры дополнительно преобразуются с помощью фильтров, приближенных к слуховому восприятию человека. Это преобразование позволяет лучше улавливать важные акустические особенности, такие как высота тона и тембр, которые важны для обработки речи и музыки.

Другим важным отличием является то, что поскольку человеческое восприятие имеет более высокое разрешение на низких частотах, чем на высоких, большая часть информации о звуковых сигналах сосредоточена на низких частотах. При отображении этого диапазона на шкалу Мела с более тонким разрешением становится легче различать различные звуки или фонемы в речи.

В целом, несмотря на то, что традиционные спектрограммы имеют свои преимущества и сферы применения, например, визуализация гармонических структур или компонентов шума в музыкальных записях, mel-спектрограммы обеспечивают более точное представление того, как человек воспринимает звук, что делает их особенно полезными для таких задач, как автоматическое распознавание речи или идентификация диктора, где требуется тонкий анализ спектральных паттернов.

# **ГЛАВА 2 НЕЙРОННАЯ СЕТЬ**



## **Искусственные нейронные сети**

Искусственные нейронные сети – математические модели, а также их аппаратные и программные воплощения строятся по принципу организации и функционирования биологических нейронных сетей.

Нейронные сети используются для решения сложных задач, которые требуют аналитических вычислений подобных тем, что делает человеческий мозг.

Несмотря на наличие различий в архитектурах нейронных сетей, все они обладают некоторыми общими чертами. Во-первых, основу каждой нейронной сети составляют относительно простые и однотипные элементы (формальные нейроны), имитирующие работу нейронов головного мозга.

Во-вторых, общей чертой для всех нейронных сетей является возможность параллельной обработки сигналов, достигающейся за счет объединения нейронов в слои. На Рис. 3 представлена структура связей однослойной полносвязной нейронной сети, где каждый нейрон из предыдущего слоя связан с каждым нейроном из следующего слоя.

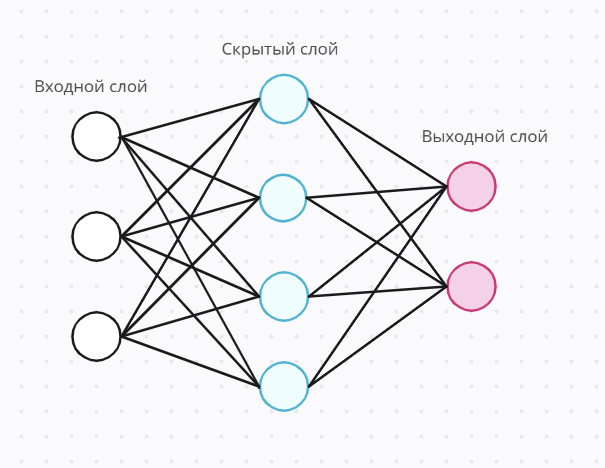


Рис. 3.Структура связей полносвязной нейронной сети

В теории, количество слоев и нейронов в слое может быть выбрано произвольным образом, но на практике оно ограничивается ресурсами компьютера, на котором будут производиться вычисления и временем, за которое сеть должна производить необходимые вычисления. Однако, чем сложнее нейронная сеть, тем более масштабные задачи она может решать.

Процесс функционирования нейронной сети зависит от величин синаптических связей (весов сети), поэтому, после определения структуры нейронной сети для решения какой-либо задачи необходимо найти оптимальные значения всех настраиваемых параметров. Этот процесс называется обучением нейронной сети. От того, насколько точно будут подобранны значения параметров (весов), зависит способность сети решать поставленные перед ней задачи.

* + 1. **Модель нейрона**

В основе любой нейронной сети лежит модель нейрона (Рис. 4), который является основной функциональной единицей нейронной сети.

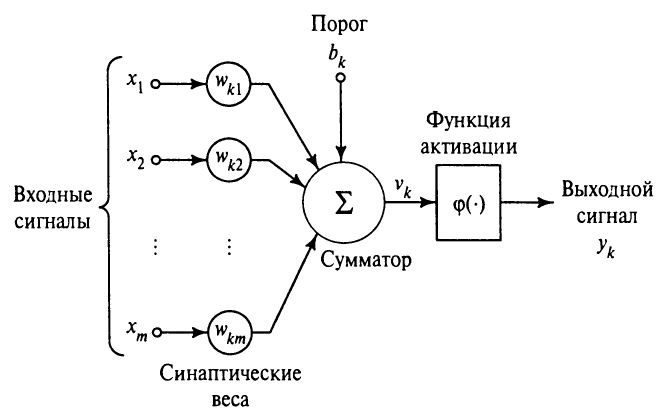


Рис. 4. Модель нейрона

В этой модели нейрон получает входные сигналы (x0...xm), которые проходят через связи-синапсы, вес которых (wk1...wkm) эмулируя различную пропускную способность синапсов в естественных нейронах. После прохождения входных сигналов по синапсам они обрабатываются сумматором, задачей которого является получение линейной комбинации всех входных сигналов, в свою очередь полученная линейная комбинация служит аргументом для функции-преобразователя (функции-активации). Именно эта функция определяет значение выходного сигнала нейрона, который затем посылается на единственный выход нейрона. Таким образом, искусственные нейроны объединяют в сети путем соединения выходов одних нейронов с входами других.

Нейрон состоит из элементов 3 типов [2]:

•**синапсов** (весов, также иногда называемых умножителями), которые характеризуют силу связи между двумя нейронами;

•**сумматор**, который выполняет сложение входных сигналов нейрона, перед этим умноженных на соответствующие веса связей;

•**преобразователь**, который реализует функцию одного аргумента. Эта функция называется функцией активации нейрона.

Исходя из полученного описания, можно представить математическую модель нейрона следующим образом:

|  |  |
| --- | --- |
|  | (2.1) |

где *x* – элементы вектора входных сигналов, *w* – значения весов связей нейрона, *b* – смещение нейрона, а *y* – выходной сигнал нейрона.

* + 1. **Функция активации**

Функция активации — это способ нормализации входных данных.

Функция активации, представленная в формуле (2.3) как f(S), определяет выходной сигнал нейрона в зависимости от взвешенной суммы входов S. Ниже приведены основные виды функций активации [1,2]:

1. **Линейная функция**. Линейная функция описывается следующим образом:

|  |  |
| --- | --- |
| f(S)= S, | (2.3) |

2. **Сигмоидальная логистическая функция**. Сигмоидальная функция является одной из самых распространенных функций, применяемых в искусственных нейронных сетях. Примером сигмоидальной функции может служить функция, описываемая следующим выражением:

|  |  |
| --- | --- |
| , | (2.4) |

где, *а* – параметр наклона, изменение которого позволит построить функцию с различной крутизной наклона (Рис. 5). Область значений сигмоидальной функции лежит в диапазоне от 0 до 1.

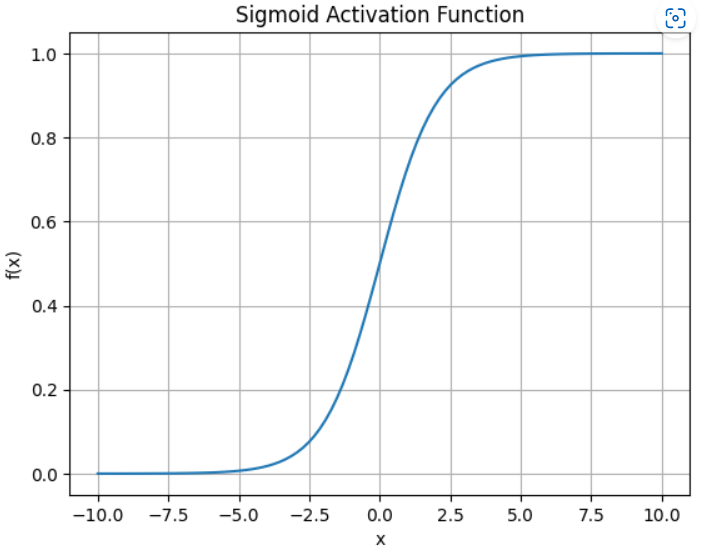


Рис. 5 График сигмоидальной логистической функции

Однако иногда требуется функция активации, имеющая область значений от-1 до +1, в этом случае функция активации должна быть симметрична относительно начала координат. Тогда пороговую функцию можно определить следующим образом:

|  |  |
| --- | --- |
|  | (2.5) |

Данная функция называется *сигнум*

*3.* **Гиперболический тангенс**. Функция используется, когда значения могут быть и отрицательными, и положительными, так как диапазон функции [-1,1]:

|  |  |
| --- | --- |
| , | (2.6) |

где *f(S)* – искомое значение элемента, *S* – взвешенная сумма входов, *a* – параметры функции. График данной функции представлен на Рис. 6.

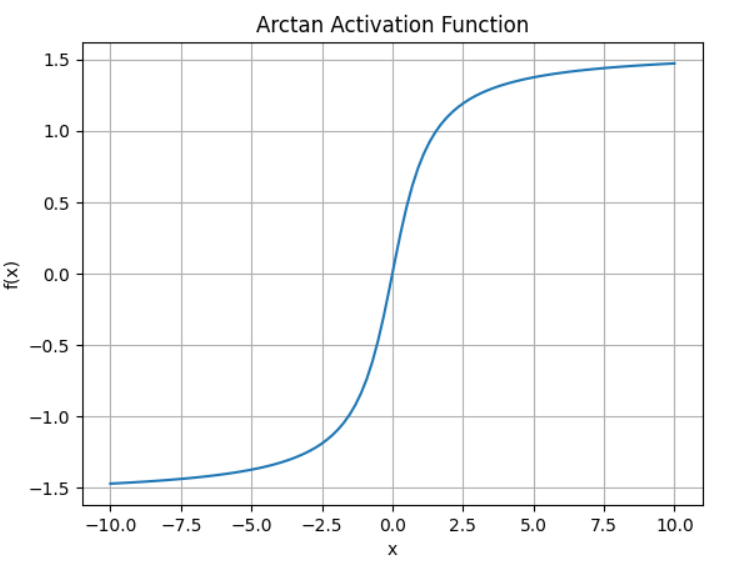


Рис.6 График гиперболического тангенса

В данной работе выбор был в пользу сигмоидальной логистической функции по следующим причинам:

* Функция очень проста в вычислении и реализации. Она требует только одного математического уравнения.
* Сигмоидальная функция дает гладкий результат. Гладкость выходного сигнала позволяет эффективно вычислять градиенты во время обратного распространения, что помогает улучшить скорость и точность обучения.
* Сигмоидальную функцию можно интерпретировать как вероятность принадлежности к классу. Это означает, что при использовании данной функции активации в задачах классификации она обеспечивает выходы как вероятности принадлежности разным классам.
  + 1. **Функция потерь**

Функция потерь характеризует величину отклонения ответа от правильного ответа – ошибку.

Ошибка — это процентная величина, отражающая расхождение между ожидаемым и полученным ответами. Для вычисления ошибки существует огромное количество способов. Для реализации программы в данном случае используется :

* Категориальная перекрестная энтропия (categorical crossentropy)
* Разреженная категориальная перекрестная энтропия   
  (sparse categorical crossentropy)

Оба эти метода используются довольно часто, поскольку они являются как достаточно точными, так и интуитивно понятными. Каждый метод считает ошибки по-своему.

Ошибки вычисляются по следующим формулам:

1. Категориальная перекрестная энтропия

|  |  |
| --- | --- |
|  | (2.7) |

где *xi* – распределение истинных ответов, а *yi* – распределение вероятностей прогнозов модели.

Хотя кроссэнтропия может использоваться для задач классификации по двум классам, не все задачи под это подходят.

Уравнение выглядит немного более сложным, и это так, но мы снова можем объяснить его чрезвычайно интуитивно.

1. Разреженная категориальная перекрестная энтропия

|  |  |
| --- | --- |
|  | (2.8) |

где (σ(yi) – softmax-функция, или номализованная экспонента)

Возможно, сначала набор данных не является категориальным и, возможно, он слишком велик для использования. В этом случае было бы довольно сложно использовать категориальную перекрестную энтропию, поскольку она зависит от категориальных данных.

Однако, когда есть целочисленные цели вместо категориальных векторов в качестве целей, можно использовать разреженную категориальную перекрестную энтропию. Это целочисленная версия категориальной функции потерь, что означает, что больше не нужно конвертировать цели в категориальный формат.

* 1. **Сверточные нейронные сети**

С появлением больших объемов данных и больших вычислительных возможностей стали активно использоваться нейронные сети. Особую популярность получили сверточные нейронные сети.

Свёрточная нейронная сеть – архитектура искусственных нейронных сетей, предложенная Яном Лекуном, предназначенная для эффективного распознавания изображений.

Данная разновидность нейронных сетей использует некоторые особенности зрительной части коры головного мозга, где были открыты простые клетки, которые реагируют на прямые линии на изображении, расположенных под разными углами, и сложные клетки, реагирующих на активацию определённого набора простых клеток.

Свое название сверточные сети получили из-за присутствия операции свёртки, суть которой заключается в вычисление нового значения текущего пикселя, учитывая значения соседних пикселей. Для вычисления значения используется ядром свертки. Во время вычисления нового значения выбранного пикселя на него накладывается ядро свертки (матрица свертки), соседние пиксели так же накрываются ядром. Далее подсчитывается сумма, где слагаемыми являются произведения значений пикселей на значения ячейки ядра, накрывшей данный пиксель. Получившийся результат суммируется и записывается в аналогичную позицию выходного изображения.

В архитектуру сети заложены априорные знания из предметной области компьютерного зрения:

•пиксель изображения сильнее связан с соседним (локальная корреляция);

•объект на изображении может встретиться в любой части изображения.

Особое внимание сверточные нейронные сети получили после конкурса ImageNet в октябре 2012 года, который был посвящен распознаванию объектов на фотографиях. Победитель данного конкурса – Алекс Крижевский, используя сверточную нейронную сеть, значительно превзошёл остальных участников.

* + 1. **Архитектура сверточной нейронной сети**

Архитектура сверточной нейронной сети представляет из себя чередование сверточных слоев, субдискретизирующих слоев, подвыборочный слой и полносвязных слоев на выходе. Все перечисленные виды слоев могут располагаться в нейронной сети произвольным образом.

В свою очередь, свёрточный слой представляет собой набора плоскостей, образующийся за счет того, что изображение предыдущего слоя сканируется небольшим окном и пропускается сквозь набор весов, а результат отображается на соответствующий нейрон текущего слоя. Таким образом, получается, что каждый нейрон выполняет свертку некоторой области предыдущего слоя. Следовательно, набор плоскостей представляет собой карты особенностей входного изображения, и каждая плоскость находит соответствующие признаки в любом месте предыдущего слоя, который содержит некоторое изображение.

На Рис. 7 изображена стандартная структура сверточной нейронной сети.

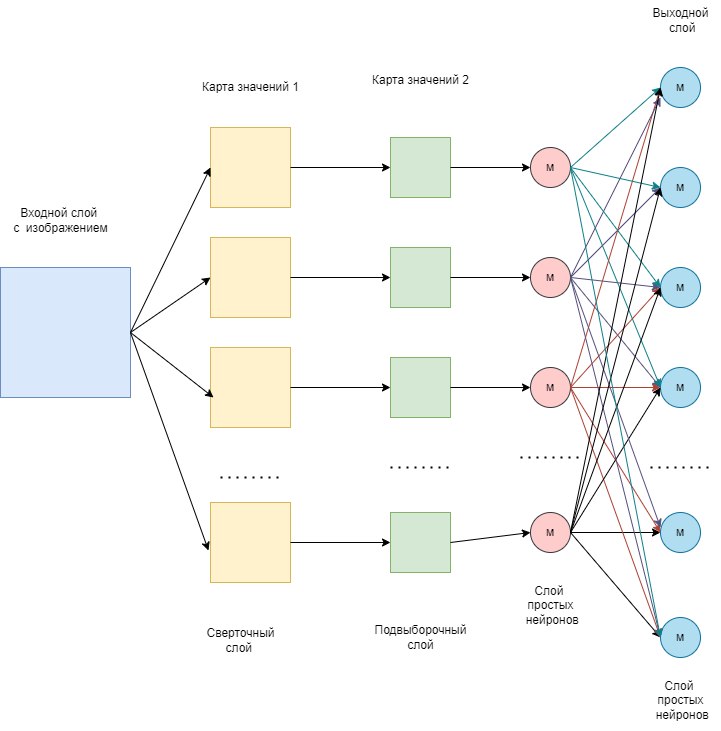


Рис. 7. Структура простой сверточной нейронной сети

Далее разберем подробнее каждый слой.

* + 1. **Сверточный слой**

Каждый нейрон находящийся в плоскости сверточного слоя получает свои входные значения от некоторой области предыдущего слоя, то есть изображение, которое подается на вход просматривается небольшим окном и пропускается сквозь набор весов, полученный результат записывается в определенный нейрон сверточного слоя.

Принцип работы сверточного слоя представлен на Рис. 8

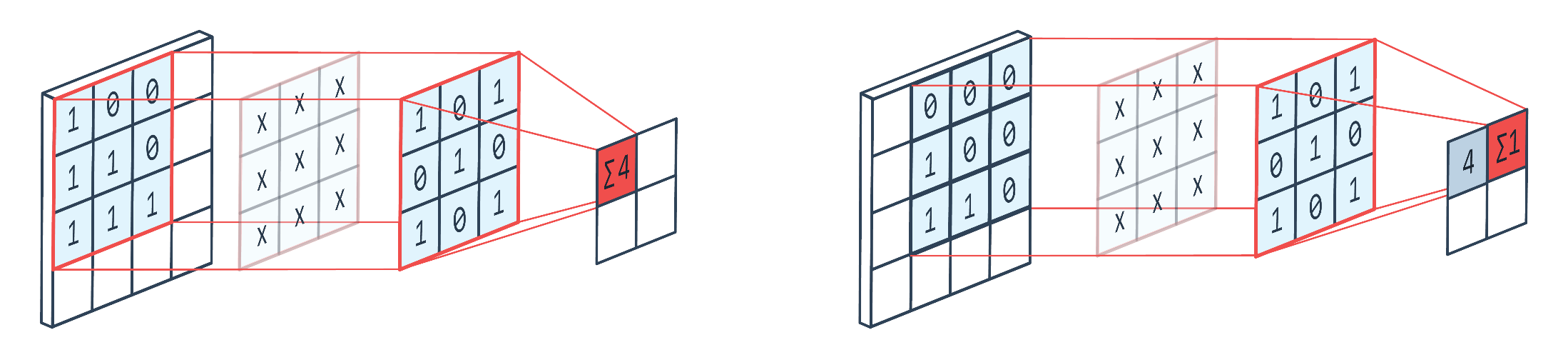


Рис. 8.Принцип работы сверточного слоя с ядром свертки 3×3

Таким образом, процесс функционирования нейрона свёрточного слоя *yk(i,j)* задается выражением [6]:

|  |  |
| --- | --- |
|  | (2.9) |

где, *yk(i,j)*– нейрон *k*-ой плоскости свёрточного слоя, *bk* – нейронной смещение *k*-ой плоскости, *К* – размер рецептивной области нейрона, *wk,s,t* –матрица синоптических коэффициентов, *х*–выходы нейронов предыдущего слоя.

Затем полученный результат свертки каждой локальной области, подается на активационную функцию:

|  |  |
| --- | --- |
| , | (2.10) |

Отличительной особенностью сверточного слоя, является то что он немного уменьшает исходное изображение за счет краевых эффектов.



### **Подвыборочный слой (субдискретизирующий слой)**

В современных ИНС используются субдискретизирующие слои , выполняющие уменьшение размерности входной карты признаков. Это можно делать разными способами, но чаще всего, для этого используется метод выбора максимального элемента в окрестности текущего элемента. После получения карты признаков, происходит разделение на ячейки, из которых выбираются максимальные по значению. Использование такого метода позволяет сделать сеть инвариантной к масштабным преобразованиям.

На Рис. 9 показан пример субдискретизирующего слоя с методом выбора максимального элемента и размером окна 2×2.

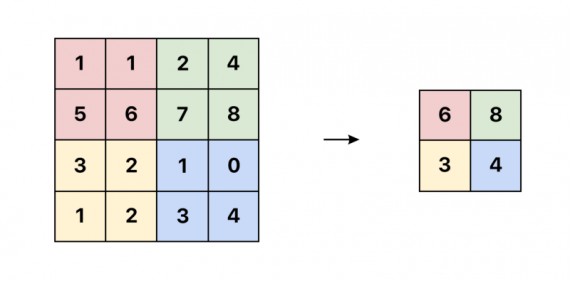


Рис. 9.Принцип работы нейронов подвыборочного слоя

Этот слой обычно идет сразу за свёрточным, он так же состоит из плоскостей, как правило, имеет такое же количество плоскостей, что и в предыдущем свёрточном слое.

1. * 1. **Полносвязный слой**

Слой, в котором каждый нейрон соединен со всеми нейронами на предыдущем уровне, причем каждая связь имеет свой весовой коэффициент. На Рис. 10 показан пример полносвязного слоя.

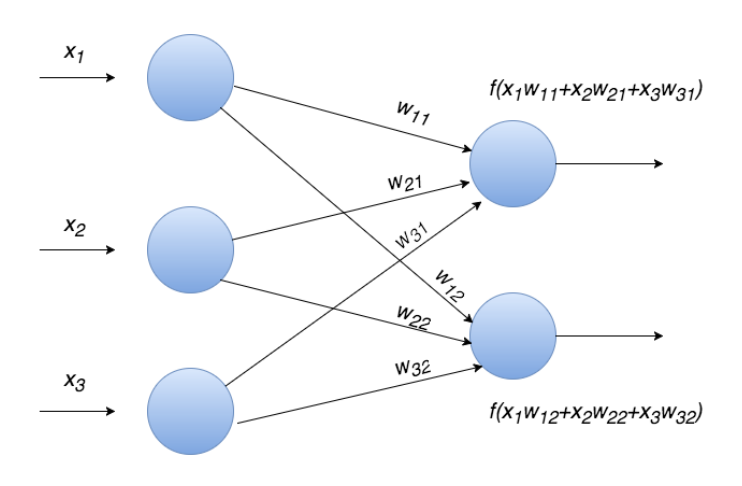


Рис. 10. Полносвязный слой

где *wij* – весовые коэффициенты связей, *xi* – входные значения сигналов, *fi* – выходные значения сети.

* + 1. **DropOut слой**

Dropout является простым и эффективным методом регуляризациии заключающийся в том, что в процессе обучения сети из её совокупной топологии многократно случайным образом выделяется подсеть, и очередное обновление весов происходит только в рамках выделенной подсети. Каждый нейрон исключается из совокупной сети с некоторой вероятностью, которая называется коэффициентом dropout.

* 1. **Алгоритм обучения нейронной сети**

Обучение нейронной сети можно производить с учителем и без учителя.

**Обучение с учителем** — этот тип присущий таким проблемам как регрессия и классификация. Нейронной сети предоставляются входные данные и желаемый результат, и нейронная сеть посмотрев на входные данные должна понять, что нужно стремиться к тому результату который ей предоставили.

**Обучение без учителя** — подобное обучение встречается не так часто. Здесь нет учителя, поэтому сеть не получает желаемый результат или же их количество очень мало.

Наиболее распространенным алгоритмом обучения нейронных сетей является метод **градиентного спуска** или как его еще называют **метод обратного распространения ошибки**, и его модификации. Данный алгоритм относится к методам обучения с учителем [2, 3, 4].

**Градиентный спуск** - это способ нахождения локального минимума или максимума функции с помощью движения вдоль градиента.

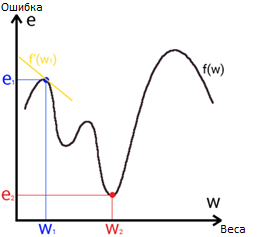


Рис. 11. График градиентного спуска

Градиент — это вектор, который определяет крутизну склона и указывает его направление относительно какой либо из точек на поверхности или графике. На Рис. 11 видно, что график функция *f(w)* является зависимостью ошибки от выбранного веса. На этом графике глобальный минимум — точка *(w2,e2)*. Эта точка будет означать, что выбрав вес *w2* мы получим самую маленькую ошибку — *e2* и как следствие, самый лучший результат из всех возможных.

В процессе обучения нейронной сети с применением алгоритма **обратного распространения ошибки**, на входы сети подаются множества примеров из обучающей выборки. Один цикл предъявления всего набора примеров из обучающей выборки называют эпохой. Процесс обучения происходит от эпохи к эпохе, пока веса и уровни порога (смешения) сети не стабилизируются, а ошибка сети на всем обучающем множестве не сойдется к некоторому заданному минимальному значению.

Алгоритм обратного распространения ошибки может быть реализован тремя способами:

**Последовательный режим**. Последовательный режим обучения так же иногда называют стохастическим градиентным спуском. В этом режиме изменение весов связей происходит после подачи каждого примера из обучающей выборки.

**Пакетный режим**. В пакетном режиме обучения корректировка весов связей происходит после подачи на вход сети всех обучающих примеров одной эпохи обучения.

**Мини-пакетный метод (**Mini-batch**)**. Между этими двумя видами метода обратного распространения ошибки (градиентного спуска) существует компромисс, называемый иногда «mini-batch». В этом случае корректировка синоптических весов сети происходит после небольшого количества обучающих образцов.

С точки зрения производительности, последовательный режим обучения является более предпочтительным, чем пакетный, так как для хранения каждой синоптической связи требуется меньший объем внутренней памяти. Помимо этого, предъявление обучающих примеров в случайном порядке в процессе обучения для последовательного режима, делает поиск в пространстве весов стохастическим. Таким образом уменьшая возможность остановки алгоритма в точке какого-либо локального минимума.

# **ГЛАВА 3 РЕАЛИЗАЦИЯ СПЕКТРАЛЬНОГО АНАЛИЗА СИГНАЛА**

В этой главе представлен практический аспект данной работы. Вначале рассматривается набор данных, используемый для оценки эффективности алгоритма, а затем описывается сам процесс обучения. Также освещаются проведенные эксперименты и вопросы, с которыми пришлось столкнуться на этом этапе. Наконец, приводится подробный отчет о полученных результатах и потенциальных улучшениях для сегментации изображений.

* 1. **Данные**

Голосовая база данных была собрана в Республиканском научно-клиническом центре неврологии и нейрохирургии (Минск, Беларусь). Она состоит из 128 устойчивых гласных фонаций (64 гласных /a/ и 64 гласных /i/) от 64 дикторов, 31 из которых был поставлен диагноз ALS(заболевание двигательных нейронов, а также описывается как "болезнь двигательных нейронов", которое представляет собой группу неврологических нарушений, влияющих на нейроны, клетки, контролирующие произвольные мышцы тела). (Пример базы на Рис.12) Каждого диктора просили произносить устойчивые фонограммы гласных /a/ и /i/ на комфортной высоте тона и громкости как можно дольше. Видно, что база данных голосов практически сбалансирована и содержит 48% патологических голосов и 52% здоровых голосов.

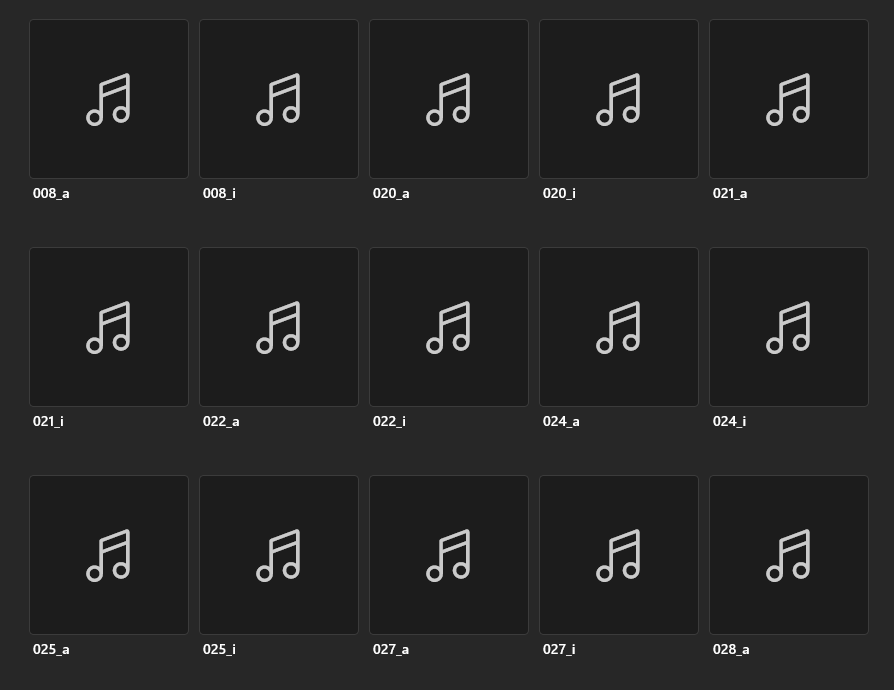


Рис. 12 голосовая база данных

Возраст 17 пациентов мужского пола варьируется от 40 до 69 лет (средний 61,1 ± 7,7), а 14 пациентов женского пола - от 39 до 70 лет (средний 57,3 ± 7,8). В случае здорового контроля (ЗК) возраст 13 мужчин варьировался от 34 до 80 лет (в среднем 50,2 ± 13,8), а возраст 20 женщин - от 37 до 68 лет (в среднем 56,1 ± 9,7). Образцы были записаны на частоте 44,1 кГц с помощью различных смартфонов с обычной гарнитурой и сохранены в виде 16-битных несжатых файлов PCM. Средняя продолжительность записей в группе HC составила 3,7 ± 1,5 с, а в группе ALS - 4,1 ± 2,0 с. Подробная информация о пациентах ALS представлена в статье[20].

* 1. **Метрики**

Для того чтобы оценить эффективность алгоритма, мы использовали три метрики: AUC , precision (точность) и recall (полнота), .

Часто результат работы алгоритма на фиксированной тестовой выборке визуализируют с помощью ROC-кривой (ROC = receiver operating characteristic, иногда говорят «кривая ошибок»), а качество оценивают как площадь под этой кривой – AUC (AUC = area under the curve).

Сетка на рис. 13 разбила квадрат на m×n блоков. Ровно столько же пар вида (объект класса 1, объект класса 0), составленных из объектов тестовой выборки. Каждый закрашенный блок на рис. 1 соответствует паре (объект класса 1, объект класса 0), для которой наш алгоритм правильно предсказал порядок (объект класса 1 получил оценку выше, чем объект класса 0), незакрашенный блок – паре, на которой ошибся.

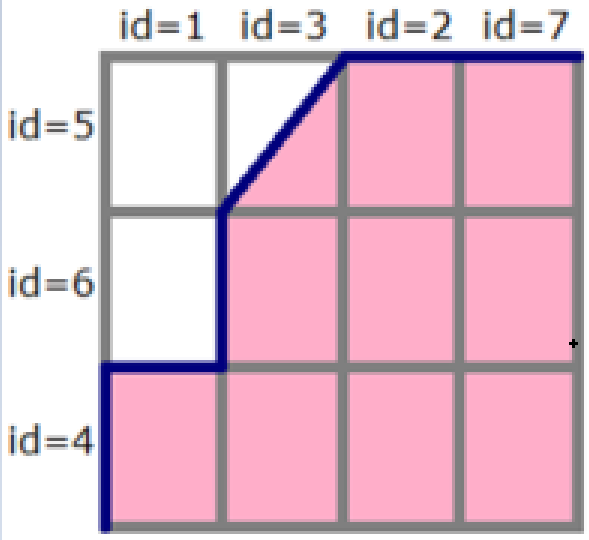


Рис. 13 Разбиение графика AUC ROC

Таким образом, AUC ROC равен доле пар объектов вида (объект класса 1, объект класса 0), которые алгоритм верно упорядочил, т.е. первый объект идёт в упорядоченном списке раньше. Численно это можно записать так:

|  |  |
| --- | --- |
| AR = | (2.11)  (2.11) |
|  |
|  | (2.12) |

где *ai* – ответ алгоритма на *i*-том объекте, *yi* – его метка (класс), *q* – число объектов в тесте.

Для оценки качества работы алгоритма на каждом из классов по отдельности введем метрики precision (точность) и recall (полнота). Precision можно интерпретировать как долю объектов, названных классификатором положительными и при этом действительно являющимися положительными, а recall показывает, какую долю объектов положительного класса из всех объектов положительного класса нашел алгоритм. Их формулы выглядят следующим образом:

|  |  |
| --- | --- |
|  | (2.13) |
|  | (2.14) |

где, *TP* — true positive: классификатор верно отнёс объект к рассматриваемому классу. *FP* — false positive: классификатор неверно отнёс объект к рассматриваемому классу. *FN* — false negative: классификатор неверно утверждает, что объект не принадлежит к рассматриваемому классу.

* 1. **Обучение** 
     1. **Средства, используемые при обучении**

Для разработки приложения сегментации зданий, представленного в данной работе, мы использовали Python версии 3.6 из-за его удобства для математических расчетов. В Python есть различные библиотеки для научных вычислений, такие как NumPy и Matlab, предназначенная для работы с N-мерными массивами и быстрыми операциями линейной алгебры. Кроме того, для python существует ряд фреймворков машинного обучения, включая Keras, который широко используется благодаря своему высокоуровневому интерфейсу и простоте использования. Он также служит оберткой для других популярных фреймворков глубокого обучения, таких как TensorFlow, Theano и CNTK. В нашем исследовательском проекте мы использовали Keras вместе с бэкендом TensorFlow.

Обычно нейронные сети требуют параллельной обработки данных во время обучения и часто обучаются на графических процессорах (GPU). Наши эксперименты проводились на видеокарте GTX 1050, оснащенной 4 ГБ памяти, что обеспечило эффективное параллельные вычисления во время обучения модели.

* + 1. **Подготовка данных**

Как уже упоминалось ранее, исходные данные состоят из аудиозаписей длительностью не более 6 секунд. Эти записи должны быть преобразованы в график сигнала с помощью встроенной функции из библиотеки librosa, которая отображает гармоническое колебание. Затем этот сигнал в большинстве случаев преобразуется в спектр с помощью ДПФ. Однако такая спектрограмма слишком громоздка, чтобы сразу использовать ее для обучения нейронной сети. Для решения этой проблемы каждая полученная спектрограмма нарезается на участки длительностью 800 мс с перекрытием в 400 мс, в результате чего получается несколько наборов "картинок" - будущих входов для классификатора - с флагами, указывающими на принадлежность к группе здоровых или больных (0 или 1). Все эти фрагменты заносятся в трехмерный массив , который потом будет подаваться на вход НС.

Пример фрагментов, подготовленных для входа в нейронную сеть :

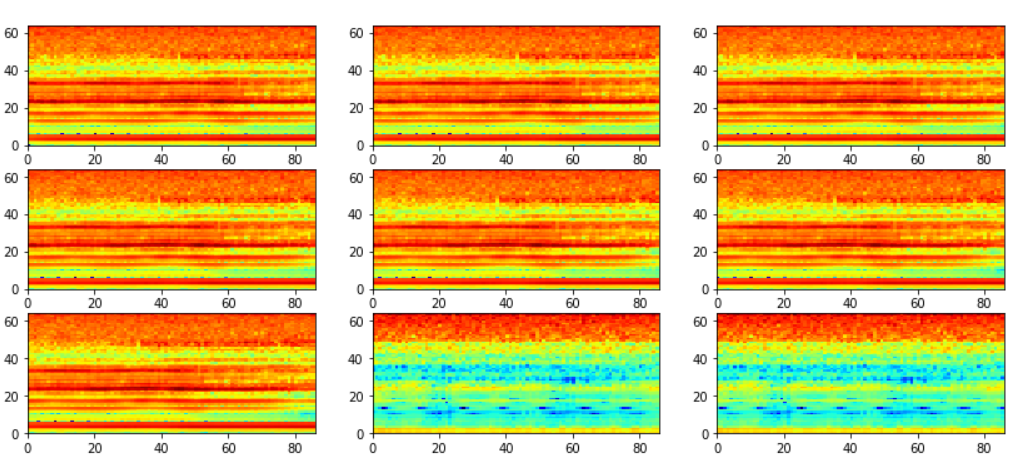


Рис. 13. Пример 9 (нормализованных) спектрограмм из обучающей выборки

Чтобы разделить набор данных на обучающие и тестовые данные, выполняются следующие действия:

Случайно перемешивается набор данных, чтобы убедиться, что данные имели произвольный порядок.

Берется определенный процент данных , который хотят использовать для обучения и тестирования (в нашем случае, 90% для обучения и 10% для тестирования).

Разделяют перемешанный набор данных на две части в соответствии с этим соотношением, одна часть будет использоваться для обучения, а другая - для тестирования.

* + 1. **Результаты экспериментов**

Первый этап

Была создана простейшая нейронная есть имеющая лишь входной слой и выходной. Она обучалась на наборе из 17575 фрагментов спектрограмм и была проверена на 1897 фрагментах.

Как видно на следующем графике(Рис.13.): с обучающим набором проблем не возникло. Однако они стали видны на тестовой выборке, поскольку с каждой эпохой количество ошибок должно падать, т.е. график – сходиться.

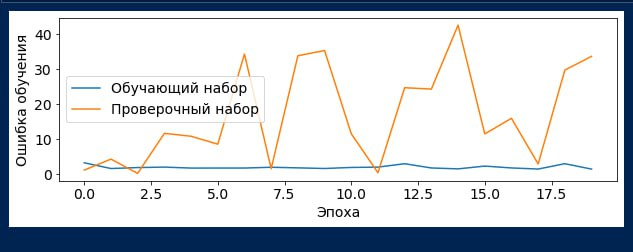


Рис. 14. Процесс обучения простейшей НС

Результат работы на итоговом тестовом наборе составил 63%. Для улучшения этого результата необходимо использовать дополнительные средства и методы.

Второй этап

Добавляется нормализация данных. Это метод, используемый при проектировании баз данных для организации данных таким образом, чтобы уменьшить избыточность и улучшить согласованность данных. Следовательно, вместо того, чтобы хранить одни и те же данные несколько раз, они организуются в таблицы, которые структурно точны и могут быть связаны между собой с помощью внешних ключей. Эта техника облегчает обслуживание и обновление систем баз данных, повышает согласованность данных, снижает риск потери или повреждения данных и повышает производительность запросов.

Регуляция весов в сети - это добавление дополнительных членов к функции потерь, которая штрафует большие значения. Это делается для того, чтобы уменьшить переобучаемость и улучшить обобщающую способность.

Так же, для более качественной оценки, из обучения были исключены случаи, когда заболевание можно определить без каких-либо технических средств. Они были перемещены в тестовую выборку.

Результат определенно улучшился, однако сходимости все ее нет (Рис.14 и Рис.15).

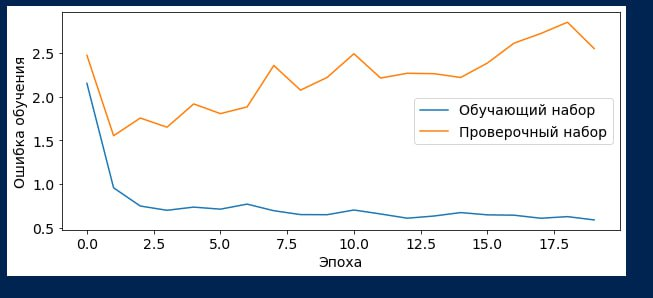


Рис.15. Процесс обучения улучшенной НС

Метрики дали такие значения: AUC = 0.818, Prec = 0.815, Recall = 0.581, при графике Рис.15.

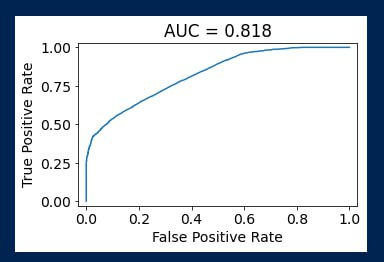


Рис.16. График кривой ROC и площадью под ней AUC для улучшенной НС

Третий этап

Скорость обучения считается одним из наиболее важных гиперпараметров для обучения моделей глубокого обучения. Изменение скорости обучения на этапах и эпохах может повысить производительность моделей глубокого обучения. Есть много способов снизить скорость обучения с течением времени. Это известно как «программирование скорости обучения».

Инициализация Ксавьера (Xavier initialization) или Инициализация Глорот (Glorot Initialization )  - это метод инициализации параметров, позволяющий избежать проблемы затухающих/взрывающихся градиентов (т.е. снижению эффективности обучения или большому накапливанию ошибок). Этот метод предполагает, что случайные реальные значения весов должны быть выбраны из нормального распределения со средним значением 0 и стандартным отклонением, равным , где *n* - количество входных и выходных соединений нейрона.

Так же для улучшения работы нейронной сети необходимо добавить DropOut слой.

В результате всего вышеописанного график обучения НС сходится, а это значит, что она работает правильно!

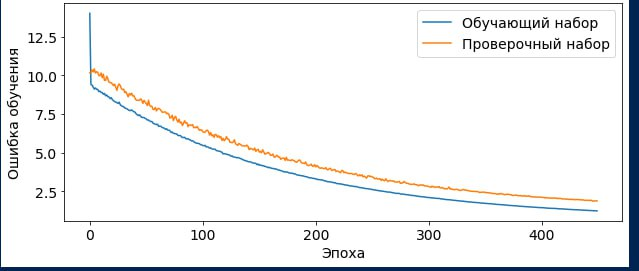


Рис.17. Процесс обучения корректно работающей НС

Значения метрик следующие: AUC = 0.994, Prec = 0.952, Recall = 0.992, где

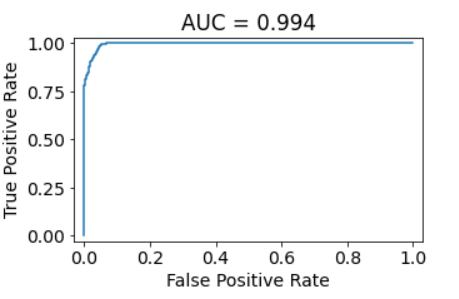


Рис.18. График кривой ROC и площадью под ней AUC для корректно работающей НС

# **ВЫВОД**

Данная работа рассматривает алгоритм, использующий нейронные сети для выявления болезни по спектрограмме. Были изучены существующие подходы и алгоритмы сегментации изображений с помощью нейронных сетей. Для решения поставленной задачи была реализована нейронная сеть. Проведена оценка эффективности.

Результатом работы является сеть способная выделять больных на ранних сроках заболевания на спектрограмме на основе обучения модели. Необходимо подготовить тренировочное множество. В работе было обращено внимание на распределение данных в обучающем множестве. Очень важно подобрать хороший набор тренировочных данных.

В заключение следует отметить, что спектральный анализ играет важнейшую роль в области медицины. Он дает ценное представление о физиологических и патологических процессах, которые не видны невооруженным глазом. Методы спектрального анализа широко используются для диагностики таких заболеваний, которые не имеют (или имеют слабовыраженные) визуальные признаки.

Использование передовых алгоритмов обработки сигналов позволяет извлечь важные особенности, которые помогают в раннем обнаружении и планировании лечения. Так же является важным инструментом для медицинских работников, поскольку он позволяет им принимать обоснованные решения по уходу за пациентами, повышая точность диагностики и снижая опасность.

# **ЛИТЕРАТУРА**

1. Смит III, Дж. О. (2010). Спектральная обработка аудиосигналов.
2. Рабинер, Л., и Жуанг, Б.-Х. (1993). Основы распознавания речи.
3. Лерч, А. (2010). Введение в анализ аудио-контента: Применение в обработке сигналов и музыкальной информатике.
4. Фант мл., Г., Лильенкрантс, Й., и Лин Квистгорд Нильсен (1985). Анализ-синтез речи на основе спектрографических методов.
5. Цанетакис, Г., и Кук, П. (2002). Автоматическая классификация музыкальных жанров аудиосигналов на международной конференции по мультимедиа и экспо (ICME).
6. Поль, К. С.(2010). Принципы цифрового аудио.
7. МакФи, Б., и др. (2011). Извлечение информации о музыке с использованием спектрально-контрастной характеристики на Международной конференции по извлечению музыкальной информации (ISMIR).
8. Ховард, Д.М., и Ангус, Дж.А.С (2006). Акустика и психоакустика (4-е издание).
9. Журнал Акустического Общества Америки.
10. Гудфеллоу, Я., Бенджио, Й., и Курвиля, А. (2016). Глубокое обучение.
11. Ли, Ф., и др. (2014). Конвюционные нейронные сети для визуального распознавания.
12. Жерон, А. (2017). Практическое машинное обучение с помощью Scikit-Learn, Keras и TensorFlow.
13. Бенджио, Й. (2009). Изучение глубоких архитектур для искусственного интеллекта.
14. Сзелиски, Р. (2010). Компьютерное зрение: Алгоритмы и приложения.
15. Бишоп, К. М. (1995). Нейронные сети для распознавания образов.
16. Бойд, С., и Ванденберге, Л. (2004). Выпуклая оптимизация.
17. Рашка, С., и Мирджалили, V. (2017). Python Machine Learning (2-е издание).
18. Гиршик, Р., и др. (2014). Богатые иерархии признаков для точного обнаружения объектов и семантической сегментации.
19. Симонян, К., и Зиссерман, А. (2015). Очень глубокие конволюционные сети для крупномасштабного распознавания изображений.
20. Maxim Vashkevich, Yulia Rushkevich (2021) Classification of ALS patients based on acoustic analysis of sustained vowel phonations -

Максим Вашкевич, Юлия Рушкевич (2021) Классификация пациентов с ALS на основе акустического анализа устойчивых гласных фонограмм

https://arxiv.org/abs/2012.07347