

УДК 519.688

Д. А. Кузин, Л. Г. Стаценко, П. Н. Анисимов, М. М. Смирнова  
Дальневосточный федеральный университет

## Применение методов машинного обучения для классификации акустических сигналов по спектральным характеристикам

Использование методов машинного обучения при проведении исследований и обработке любых объемов информации относится к современным тенденциям. Их применение эффективно для классификации большого количества записанных акустических сигналов. Спектральная характеристика звукового сигнала позволяет исследовать его по частотным компонентам. В качестве примера для оценки эффективности методов машинного обучения используются 3240 акустических записей работы сердца. Заболевания сердечно-сосудистой системы – самые распространенные причины смерти среди населения во всем мире. По оценкам Всемирной организации здравоохранения каждый год от заболеваний сердечно-сосудистой системы умирает около 17.9 миллиона человек, что составляет 31 % от всей смертности [1]. Существуют различные методы обнаружения патологических состояний сердца, основанные на анализе звуковых явлений, возникающих при его работе. Для классификации акустических сигналов применяются модели логистической регрессии и случайных лесов. Также для решения задачи классификации проведено обучение и тестирование нейронной сети. Эффективность работы моделей оценивается сразу по четырем метрикам, а именно accuracy, precision, recall, F1.

### Машинное обучение, классификация акустических сигналов, спектральный анализ, искусственная нейронная сеть, логистическая регрессия, метод случайных лесов

В работе используется набор акустических сигналов, состоящий из 3240 записей с различными спектральными характеристиками. Набор данных включает в себя 2575 записей здоровых сердец и 665 записей больных сердец (все файлы в формате .wav). Алгоритм, реализованный на языке программирования Python, строит спектрограмму для каждого звукового сигнала из набора. Пример спектрограммы для акустического сигнала с положительным результатом (здоровое сердце) приведен на рис. 1, а, а с отрицательным результатом (больное

сердце) – на рис. 1, б. Можно наблюдать учащение появления вертикальных полос на рис. 1, б по сравнению с рис. 1, а. Эти вертикальные полосы представляют собой биение сердца.

Затем на основании спектрограммы извлекаются признаки для построения модели машинного обучения. Извлеченные признаки записываются в табличный файл формата .csv, чтобы не затрачивать время на повторное преобразование акустического сигнала в спектр позднее. В качестве признаков акустических сигналов мы используем:

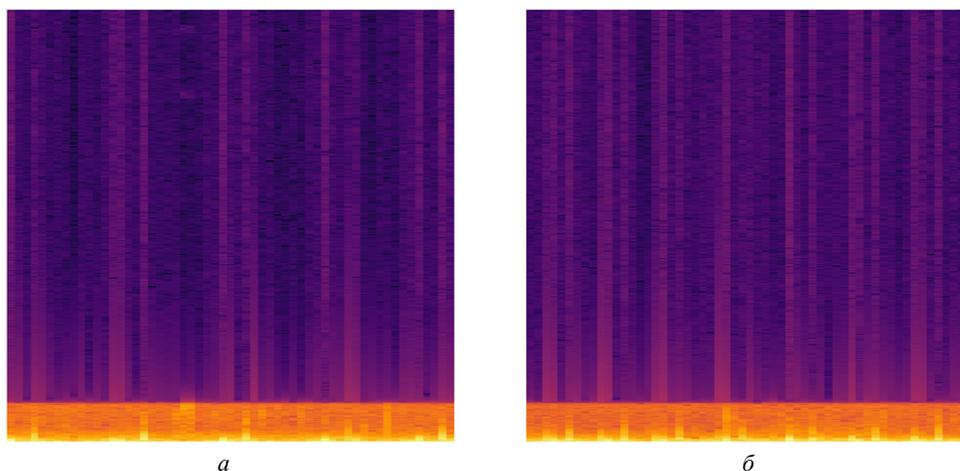


Рис. 1

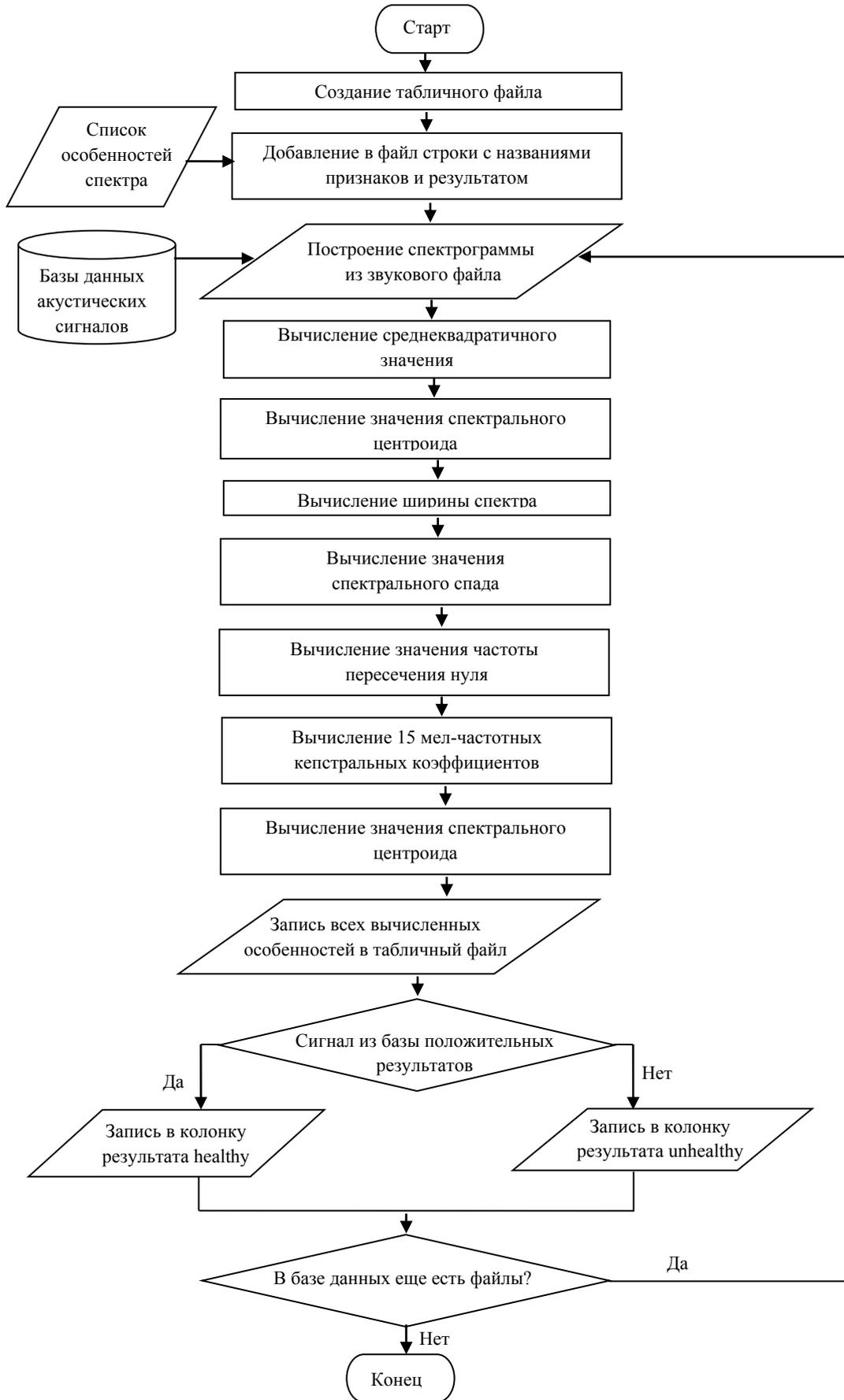


Рис. 2

среднеквадратичное значение, спектральный центр, ширину спектра, спектральный спад, частоту пересечения нуля, 15 мел-частотных кепстральных коэффициентов для описания общей формы спектральной огибающей. Мел-частотные кепстральные коэффициенты широко используются при классификации музыкальных жанров и музыкальных инструментов, которые используются в произведении [2]. Кроме того, существуют исследования, которые описывают успешное применение мел-частотных кепстральных коэффициентов и в медицине, например для выявления патологий в легких по шумовой спектральной характеристике [3]. Извлечение признаков акустических сигналов осуществляется в реализованном на языке Python алгоритме с помощью библиотеки *librosa*. Для каждого признака в таблице содержится столбец с его названием. В последний столбец записывается результат классификации, в нашем случае это бинарное значение – больное сердце или здоровое. Таким образом, строка содержит 21 признак акустического сигнала и результат их классификации. Данный файл используется для построения моделей машинного обучения и их оценки. Блок-схема алгоритма создания файла с признаками спектра звуковых сигналов представлена на рис. 2.

Выполнение преобразования всех акустических сигналов из исходного набора данных в спектр и извлечение его признаков заняло 890 с на процессоре Ryzen 5 3500U с тактовой частотой 2.1 ГГц. Данная часть алгоритма выполняется один раз и в последующем полученный результат в виде таблицы с признаками используется во всех моделях машинного обучения. Скорость обучения для моделей в дальнейшем будет приводиться без учета затрат времени на это преобразование.

В результате получен набор данных размером  $3240 \times 21$ , где 3240 – количество точек данных (звуковых сигналов), а 21 – количество признаков (спектральных характеристик для каждого звукового сигнала).

**Модель логистической регрессии.** Изначально была построена модель логистической регрессии для исходных акустических данных. Метод логистической регрессии используется некоторыми специалистами для выявления заболевания гортани по голосу [4].

Весь набор данных разделен на части: для обучения и для проверки. Реализованный алгоритм разбивает данные всегда одинаково. Этот шаг необходим для адекватного сравнения метрик – как для отдельно взятого метода, так и для сравнения методов между собой. После разделения набор данных для обучения модели имеет размер  $2592 \times 21$ , а тестовый набор данных –  $648 \times 21$ . После обучения модели логистической регрессии и ее проверки на тестовом наборе данных были получены следующие результаты метрик:

- Accuracy = 0.8070;
- Precision = 0.8355;
- Recall = 0.9462;
- F1 = 0.8874.

Для интерпретации результатов построена матрица ошибок, которая содержит:

1. Количество точек данных, для которых модель предсказала положительный результат, и этот результат на самом деле положительный (истинно положительный, ИП).
2. Количество точек данных, для которых модель предсказала положительный результат, но этот результат на самом деле отрицательный (ложноположительный, ЛП).
3. Количество точек данных, для которых модель предсказала отрицательный результат, но этот результат на самом деле положительный (ложноотрицательный, ЛО).
4. Количество точек данных, для которых модель предсказала отрицательный результат, и этот результат на самом деле отрицательный (истинно отрицательный, ИО).

Матрица ошибок для модели логистической регрессии изображена на рис. 3.

	ИП	ЛП
493	97	
28	30	
	ЛО	ИО

Рис. 3

Метрика «Ассигасу» – это точность модели, или по-другому – доля верных прогнозов в процентах. Формула для вычисления общей точности модели:

$$\text{Accuracy} = \frac{\text{ИП} + \text{ИО}}{\text{ИП} + \text{ИО} + \text{ЛП} + \text{ЛО}}$$

Метрику «Precision» можно интерпретировать как долю точек данных, которые классификатор счел положительными и которые при этом действительно положительны. Расчет «Precision» выполняется по формуле

$$\text{Precision} = \frac{\text{ИП}}{\text{ИП} + \text{ЛП}}$$

Полнота «Recall» отображает, какую долю положительных точек данных из всех объектов положительного класса нашел алгоритм. Формула полноты выглядит следующим образом:

$$\text{Recall} = \frac{\text{ИП}}{\text{ИП} + \text{ЛО}}$$

Метрика F1 сочетает в себе точность (Precision) и полноту (Recall) и вычисляется по следующей формуле:

$$F1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Необходимо отметить, что общая точность полученной модели составляет 81 %, по метрике F1 – 89 %, однако модель неверно классифицировала 94 отрицательные точки данных, а 28 положительных точек были классифицированы моделью как отрицательные. Полученная модель достаточно точно классифицирует положительные точки данных (здоровое сердце), но ошибается при классификации отрицательных (больное сердце).

Построение модели логистической регрессии для набора данных и вычисление метрик заняло 0.22 с.

### Модель случайного леса (Random Forest).

В модели логистической регрессии используется параметрический алгоритм машинного обучения, задача которого состоит в том, чтобы математически подобрать оптимальные коэффициенты, которые будут определять линию, разделяющую положительные и отрицательные прогнозы. Модель случайного леса использует непараметрический алгоритм, большое число (ансамбль) решающих деревьев и является статистическим методом, предназначенным для решения задач классификации и регрессии. Каждое дерево решений в модели случайного леса строится на основе выборки, которая получается из исходной обучающей выборки с помощью выборки с возвращением (бутстреп). Каждый лист дерева содержит наблюдения только одного класса. Классификация в данном методе осуществляется с помощью голосования классификаторов, определяемых отдельными деревьями [5]. Метод случайных лесов получил широкое признание среди исследователей, применяющих методы распознавания образов, благодаря своей высокой точности и защите от переобучения. Случайный лес не заикливается на деталях обучающего набора данных, позволяя увидеть картину в целом.

Для определения оптимального количества деревьев в модели случайного леса был применен метод локтя.

На графике зависимости точности от числа деревьев (рис. 4) можно наблюдать незначительный рост точности модели при значительном росте количества деревьев в диапазоне более 12. Таким образом, число 12 – оптимальное.

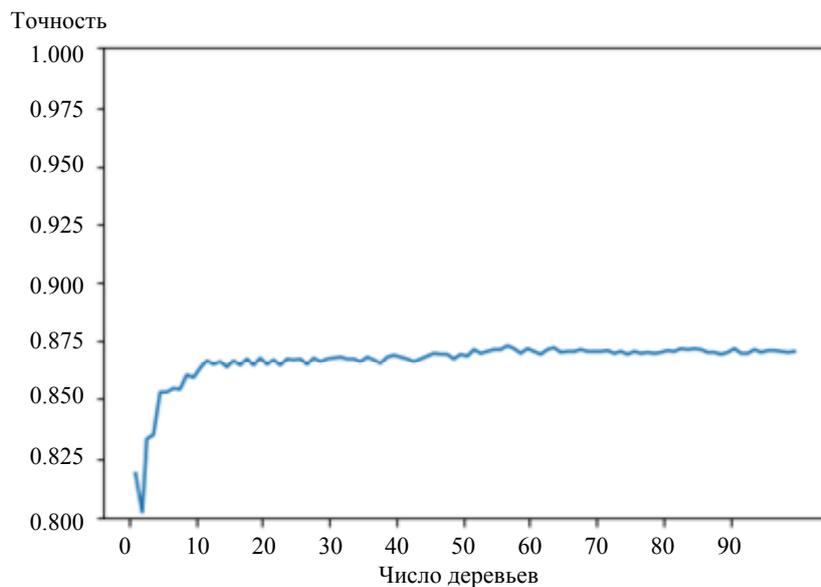


Рис. 4

Результаты метрик, рассчитанные с помощью перекрестной проверки для модели случайных лесов:

- Accuracy: 0.8735;
- Precision: 0.9111;
- Recall: 0.9317;
- F1 score: 0.9212.

Матрица ошибок для модели случайного леса представлена на рис. 5.

По сравнению с моделью логистической регрессии наблюдается значительный прирост в качестве прогнозирования: метрика F1 выросла с 0.8874 (для логистической регрессии) до 0.9212 (для случайного леса), а число неверно классифицированных отрицательных точек (большое сердце) снизилось с 97 до 39. Тем не менее, доля невыявленных сердечных больных по их спектральным характеристикам составляет 31 %.

	ИП	ЛП
ЛО	480	39
ИО	41	88

Рис. 5

Построение модели случайных лесов для набора данных и вычисление метрик заняло 0.49 с – больше, чем затрачено на построение модели логистической регрессии, но качество предсказания определенно стоит этой разницы во времени.

**Искусственная нейронная сеть.** Нейронные сети в настоящее время являются мощным и популярным инструментом для работы с любыми данными [6]–[8]. Они позволяют работать с данными, содержащими большое число признаков, при этом не требуя от пользователя глубокого погружения в предметную область.

Была построена полносвязная нейронная сеть (многослойный перцептрон), состоящая из четырех слоев: слой 1 – 256 элементов, функция активации ReLU [9]; слой 2 – 128 элементов, функция активации ReLU; слой 3 – 64 элемента, функция активации ReLU; слой 4 – 10 элементов, функция активации «softmax». Для адекватности результатов при ее тестировании и получении метрик использовались данные, которые не участвовали в обучении.

В итоге получены следующие показатели общих метрик для тестового набора данных из 648 звуковых файлов:

- Accuracy: 0.91;
- Precision: 0.9529;
- Recall: 0.9328;
- F1 score: 0.9427.

Матрица ошибок для искусственной нейронной сети представлена на рис. 6.

	ИП	ЛП
ЛО	486	24
ИО	35	103

Рис. 6

Нейронная сеть неверно классифицировала 24 отрицательных результата, но по сравнению с двумя другими методами машинного обучения продемонстрировала лучший результат. Итоговое время, затраченное на обучение искусственной нейронной сети и расчет метрик, составило 7 с при 100 полных проходах набора данных по нейронной сети в прямом и обратном направлении.

В статье проанализированы три метода машинного обучения для классификации акустических сигналов на примере шумов сердца. Классификация осуществлялась по особенностям спектра звукового сигнала с помощью следующих методов машинного обучения: метод логистической регрессии (ЛР), метод случайных лесов (СЛ), искусственная нейронная сеть (ИНС). Эти методы использовали одинаковый набор признаков спектра звуковых сигналов и разбиение набора данных для обучения и тестирования. Были посчитаны метрики точности, полноты, F1, матрица ошибок, а также время, затраченное на обучение и тестирование модели. Общие результаты метрик приведены в таблице. Результаты работы моделей машинного обучения в виде гистограммы представлены на рис. 7.

Метод	Accuracy	Precision	Recall	F1
ЛР	0.8074	0.8355	0.9462	0.8866
СЛ	0.8735	0.9111	0.9317	0.9212
ИНС	0.91	0.9529	0.9328	0.9427

Можно наблюдать превосходство нейронной сети над методами случайного леса и логистической регрессии при классификации акустических сигналов. Результаты метрик хочется дополнить таким показателем, как доля неверно классифи-

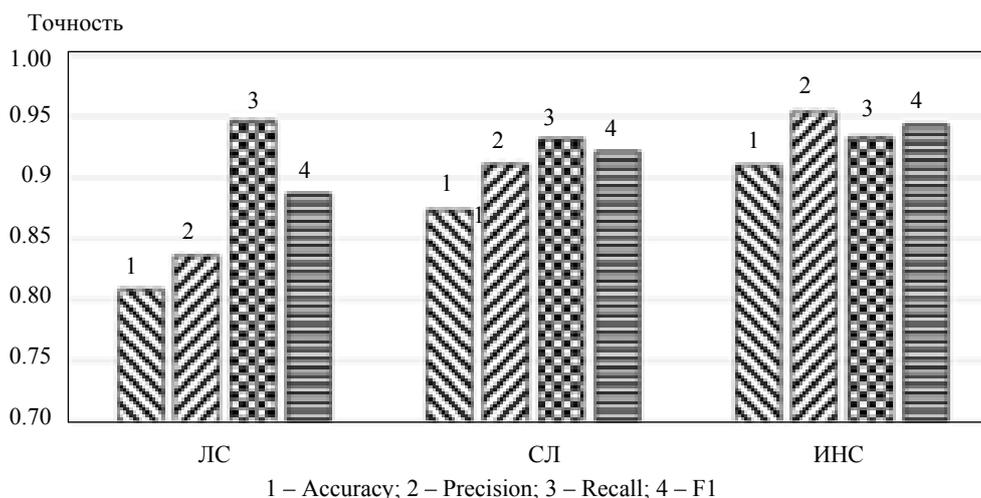


Рис. 7

цированных отрицательных точек данных (неверно диагностированных больных сердец) в процентах. Если алгоритм отнесет здоровое сердце к больным – это не так критично, как пропустить больное сердце и отнести его к здоровым. Доля неверно классифицированных больных сердец от общего количества больных сердец в тестовой выборке составила:

- метод логистической регрессии – 76 %;
- метод случайного леса – 31 %;
- искусственная нейронная сеть – 19 %.

Улучшить качество распознавания патологий сердца по его звуковым характеристикам можно,

увеличив объем набора данных как больных, так и здоровых сердец, а также оценив важность признаков, применяемых при построении модели и используя признаки с большим качественным показателем. Кроме того, рассматривается возможность применения сверточной нейронной сети для повышения качества результатов классификации акустических сигналов.

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-37-90084\19.*

## СПИСОК ЛИТЕРАТУРЫ

1. World Health Organization, «Cardiovascular diseases (CVDs)» 17 May 2017. URL: <https://www.euro.who.int/en/health-topics/noncommunicable-diseases/cardiovascular-diseases/cardiovascular-diseases2> (дата обращения 23.12.20).

2. Станкевич Ф. В., Спицын В. Г. Нейросетевое распознавание музыкальных инструментов с использованием мел-частотных кепстральных коэффициентов // *Фундаментальные исследования*. 2014. Т. 1, № 12. С. 51–56.

3. Sengupta N., Sahidullah M., Saha G. Lung sound classification using cepstral-based statistical features // *Computers in Biology and Medicine*. 2016. Vol. 75. P. 118–129.

4. Анализ акустических параметров голоса для выявления заболеваний гортани / М. И. Вашкевич, А. А. Бурак, Н. С. Конойко, В. С. Долдова // *Информатика*. 2019. Т. 17, № 1. С. 78–86.

5. Чистяков С. П. Случайные леса: обзор // *Тр. Карельского науч. центра РАН*. 2013. № 1. С. 117–136.

6. Social network recommender system, a neural network approach / A. Rivas, P. Chamoso, A. González-Briones, J. Pavón, J. M. Corchado // *Intern. Conf. on Intelligent Data Engineering and Automated Learning*. Springer, Cham, 2020. P. 213–222.

7. Sendra-Arranz R., Gutiérrez A. A long short-term memory artificial neural network to predict daily HVAC consumption in buildings // *Energy and Buildings*. 2020. Vol. 216. P. 109952.

8. Prediction of mortality from 12-lead electrocardiogram voltage data using a deep neural network / S. Raghunath, A. Ulloa, L. Jing, D. P. van Maanen // *Nature Medicine* 2020. Vol. 26, № 6. P. 886–891.

9. Schmidt-Hieber J. Nonparametric regression using deep neural networks with ReLU activation function // *Annals of Statistics*. 2020. Vol. 48, № 4. P. 1875–1897.

D. A. Kuzin, L. G. Statsenko, P. N. Anisimov, M. M. Smirnova  
*Far Eastern Federal University*

## APPLYING MACHINE LEARNING METHODS TO ACOUSTIC SIGNAL CLASSIFICATION USING SPECTRUM CHARACTERISTICS

*Machine learning methods are modern tendencies in research and processing of any amount of information. Their use is effective for classifying a large number of recorded acoustic signals. Spectral characteristic of a sound signal allows you to carry out research on its frequency components. To assess an effectiveness of machine learning methods was used 3240 acoustic recordings of heart work as an example. Cardiovascular diseases are the most common causes of death of people all over the world. According to the estimates of the World Health Organization, about 17.9 million people die each year from diseases of the cardiovascular system, which is 31 % of all global deaths [1]. There are various methods for detecting pathological conditions of the hear. This methods based on the analysis of sound phenomena that occur during its work. Logistic regression model, a random forest model were used for classification. Also, neural network was trained and tested to solve classification problem. Performance evaluation of models will be carried out according to four metrics (accuracy, precision, recall, F1).*

**Machine learning, acoustic signal classification, spectral analysis, artificial neural network, logistic regression, random forest**

---