

# Методы обучения классификатора на основе звуков кашля для выявления заболевания вирусом COVID-19

А.Л. Аленин, email: freeadvice.ru@gmail.com<sup>1</sup>

В.В. Мокшин, email: vladimir.mokshin@mail.ru<sup>1</sup>

<sup>1</sup> КАЗАНСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. А.Н. ТУПОЛЕВА – КАИ

***Аннотация.** В данной статье были рассмотрены методы извлечения признаков из звука кашля, а также последующее обучение нейронных сетей для определения является ли кашель COVID положительным или отрицательным. Для обучения использовались свёрточные нейронные сети и сети долгой краткосрочной памяти. Были подобраны оптимальные гиперпараметры которые давали наименьшую ошибку распознавания при относительно коротком времени обучения. Таким образом, был подобран алгоритм обработки звуков кашля, позволяющий определять болеет ли человек вирусом COVID-19.*

***Ключевые слова:** машинное обучение, Мел-кепстральные коэффициенты, свёрточная нейронная сеть, долгая краткосрочная память.*

## Введение

Кашель - один из преобладающих симптомов COVID-19 и также симптом более 100 других заболеваний и его влияние на дыхательную систему, как известно, различается. Например, заболевания легких могут вызывать сужение или закупорку дыхательных путей, и это может повлиять на акустику кашля. Голосовые связки ведут себя по-разному при разных патологических состояниях, и это позволяет различать кашель туберкулеза, астмы, бронхита и коклюша. Простые инструменты машинного обучения, например бинарные классификаторы, умеют различать дыхательные шумы COVID-19 от здоровых сверстников. Похоже, что при кашле от COVID-19 существуют уникальные закономерности, которые позволяют предварительно обучить классификаторы для выявления кашля COVID-19 с высокой точностью. В данном случае образцы кашля собирались анонимно у 3621 человека с подтвержденным COVID-19. Помимо искажения непосредственно голоса, при заболевании COVID-19, также имеет место быть ослабление

мышц всего организма, что влияет на длительность и частоту кашля и дыхания в общем [2].

Выявление заболевания – главный элемент в системах поддержки принятия решений в медицинской сфере. Возможность определять заболевание по звуку кашля, записанному на смартфон или другую аудиоаппаратуру, даёт множество преимуществ. Такой вид досмотра бесконтактный, простой в применении и может снизить рабочую нагрузку в испытательных центрах, а также ограничить передачу вируса за счет рекомендации ранней самоизоляции тем, у кого кашель свидетельствует о наличии COVID-19.

### **1. Предварительная обработка аудиозаписей звуков кашля**

Необработанные аудиозаписи кашля из набора данных имеют частоту дискретизации ( $\mu$ ) 44,1 кГц и подвергаются некоторым простым этапам предварительной обработки. Длина временного окна ( $\lambda$ ) равна 0,05 с, а пороговое значение амплитуды ( $\phi$ ) равно 0,005, причем оба эти значения определялись вручную и интерактивно, поскольку удаление тишины во всех случаях подтверждалось визуально.

Исходный звук кашля  $c_i(t)$  нормализуется по уравнению [3]:

$$c_i(t) = 0.9 \times \frac{c_i(t)}{|\max(c_i(t))|} . \quad (1)$$

Обработанный окончательный звук кашля показан на изображении (рис. 1). Здесь  $I$  обозначает временное окно, и мы определяем:

$$C_I(t) = C_{j\mu\lambda}(t) \dots C_{(j+1)\mu\lambda}(t) . \quad (2)$$

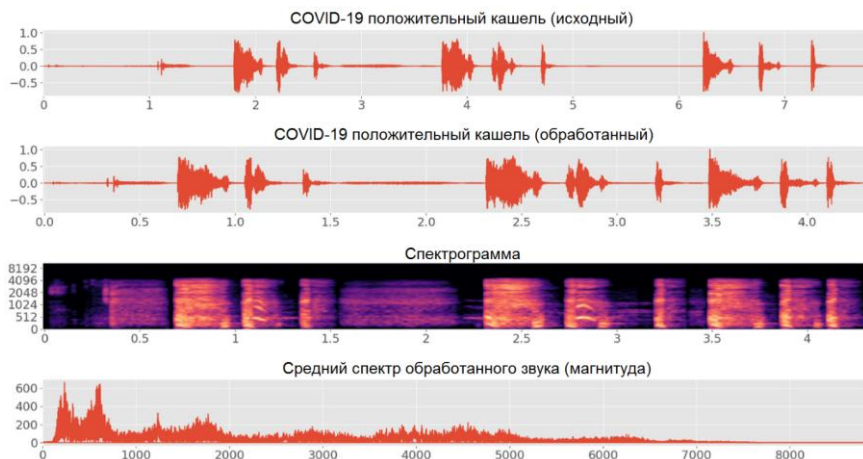


Рис. 1. Обработанный звук кашля COVID-19

Обработанный звук кашля COVID-19 короче исходного звука кашля, но сохраняет разрешение всего спектра. При предварительной обработке амплитуды нормализуются, а длительные паузы удаляются.

Таким образом, амплитуды необработанных аудиоданных в наборах данных были нормализованы, после чего периоды тишины были удалены из сигнала с запасом в 50мс с помощью простого детектора энергии [4].

После предварительной обработки набор данных содержит 92 положительных на COVID-19 и 1079 здоровых субъектов. При этом положительный на COVID-19 кашель на 15–20% короче, чем кашель, не связанный с COVID.

## 2. Извлечение признаков из звуков

Процесс выделения признаков следующий. Извлекаются такие признаки, как кепстральные коэффициенты мел-частоты (MFCC), энергия логарифмического кадра, скорость пересечения нуля (ZCR). MFCC часто используются в качестве функций в аудиоанализе и особенно в автоматическом распознавании речи. Они полезны для дифференциации сухого кашля от влажного кашля и классификации туберкулезного кашля. В данном случае был использован традиционный метод извлечения MFCC, рассматривая MFCC с более высоким разрешением, а также скорость (разность первого порядка) и ускорение (разность второго порядка), поскольку их добавление показало улучшение классификатора [5]. ZCR — это количество раз, когда сигнал

меняет знак в пределах кадра, указывая на изменчивость, присутствующую в сигнале. Для образцов звукового сигнала это указывает на преобладание более высоких амплитуд. Эти функции были извлечены с использованием гиперпараметров для всех записей кашля (табл. 1).

Таблица 1

*Гиперпараметры извлечения признаков, оптимизированные с использованием перекрестной проверки.*

Гиперпараметры	Описание	Диапазон
MFCC	Количество MFCC более низкого порядка, которое необходимо сохранить	$13 \times k_1$ , где $k_1 = 1, 2, 3, 4, 5$
Frame	Размер кадра, в котором аудио сегментируется	$2k_2$ , где $k_2 = 8, \dots, 12$
Seg	Количество кадров, извлеченных из аудио	$10 \times k_3$ , где $k_3 = 5, 7, 10, 12, 15$

Признаки извлекаются таким образом, чтобы сохранить информацию о начале и конце эпизода кашля, чтобы можно было обнаружить закономерности во временной области в записях, сохраняя при этом фиксированную размерность входных данных. Из каждой записи мы извлекаем фиксированное количество характеристик, равномерно распределяя кадры анализа фиксированной длины по временному интервалу кашля. Матрица входных признаков для классификаторов всегда имеет размерность  $(3MFCC + 3Seg)$  для количества MFCC вместе с числом скорости и числом ускорения [6]. Если  $\Lambda$  — количество выборок в звука кашля, мы можем рассчитать количество выборок между последовательными кадрами  $\delta$ , используя уравнение:

$$\delta = \left\lceil \frac{\Lambda}{\zeta} \right\rceil. \quad (3)$$

Так, например, звук кашля продолжительностью 2,2с содержит 97020 семплов, поскольку частота дискретизации составляет 44,1 кГц. Если длина кадра составляет 1024 выборки, а количество сегментов равно 100, то пропуск кадра ( $\delta$ ) составляет 971 образец.

В отличие от более традиционно применяемых фиксированных частот кадров, этот способ извлечения признаков гарантирует, что вся запись захватывается в пределах фиксированного количества кадров, что особенно позволяет классификаторам нейронных сетей обнаруживать более полезные временные шаблоны и обеспечивать лучшую производительность классификации.

### 3. Архитектура классификаторов

Свёрточная нейронная сеть (CNN) — популярная архитектура глубокой нейронной сети, в основном используемая для классификации изображений. Например, за последние два десятилетия CNN успешно применялись для решения сложных задач, таких как распознавание лиц. Она также хорошо зарекомендовала себя при классификации речи. Архитектура CNN вместе с оптимизированными гиперпараметрами (рис. 2). Модель долгой краткосрочной памяти (LSTM) представляет собой тип рекуррентной нейронной сети, архитектура которой позволяет запоминать ранее просмотренные входные данные при принятии решения о классификации. Она успешно используется для автоматического обнаружения кашля, а также для других типов обнаружения акустических событий. Гиперпараметры, оптимизированные для классификатора LSTM (рис. 3).

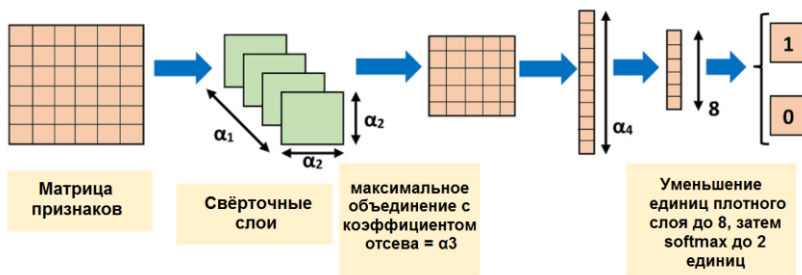


Рис. 2. Классификатор CNN

Классификатор CNN использует  $\alpha_1$  двумерных сверточных слоев с размером ядра  $\alpha_2$ , выпрямленными линейными единицами в качестве функций активации и коэффициентом отсева  $\alpha_3$ . После максимального объединения следуют два плотных слоя с  $\alpha_4$  и 8 единицами соответственно и выпрямленными линейными функциями активации. Сеть завершается двумерным softmax, где один выход (1) представляет

положительный класс COVID-19, а другой (0) — здоровый или отрицательный класс COVID-19. Во время обучения функции представляются нейронной сети партиями размером  $\beta_3$  для эпох  $\beta_4$ .

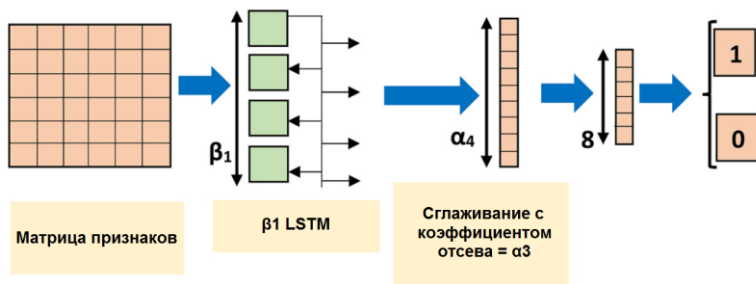


Рис. 3. Классификатор LSTM

Классификатор LSTM имеет единицы  $\beta_1$ , каждая с выпрямленными линейными функциями активации и коэффициентом отсева  $\alpha_3$ . Далее следуют два плотных слоя с  $\alpha_4$  и 8 единицами соответственно и выпрямленными линейными функциями активации. Сеть завершается двумерным softmax, где один выход (1) представляет положительный класс COVID-19, а другой (0) — здоровый или отрицательный класс COVID-19. Во время обучения функции представляются нейронной сети партиями размером  $\beta_3$  для эпох  $\beta_4$  [7].

#### 4. Результаты обучения

Результаты классификации имеют следующие результаты (табл. 2):

Таблица 2

Результаты классификации

Классификатор	Лучшие признаки гиперпараметров	Оптимальный классификатор гиперпараметров	Представление		
			Spec	Sens	ACC
CNN	26	48	99%	90%	94.57%
CNN	39	96	98%	90%	94.35%
LSTM	13	128	97%	91%	94.02%
LSTM	26	256	97%	90%	93.65%

В таблице показано, что все классификаторы могут классифицировать кашель COVID-19, но классификатор CNN демонстрирует лучшую производительность с AUC 0,953 при использовании 120-мерной матрицы признаков, состоящей из 26 MFCC с добавленными скоростью и ускорением, извлеченными из сегментов длиной 1024, а также при группировке кашля по 48 сегментов.

Из таблицы 2 также видно, что при использовании большего количества MFCC постоянно происходит повышение производительности. Поскольку спектральное разрешение используемое для вычисления 39-мерного MFCC превосходит слуховую систему человека, можно сделать вывод, что классификаторы используют информацию обычно не воспринимающуюся слушателями-людьми.

### **Заключение**

После предварительной обработки аудиозаписей кашля, было обнаружено, что кашель COVID-19 положительного субъекта на 15–20% короче, чем при кашле без COVID. Хотя описываемые классификаторы требуют более строгой проверки на более крупном наборе данных, представленные результаты очень многообещающие и указывают на то, что скрининг COVID-19 на основе автоматической классификации кашля жизнеспособен. Поскольку данные были получены на смартфонах, и поскольку классификатор в принципе может быть реализован и на таком устройстве, такая классификация кашля экономична, проста в применении и развертывании. Кроме того, его можно применять удаленно, избегая контакта с медицинским персоналом.

### **Список литературы**

1. Aydın S. Log energy entropy-based EEG classification with multilayer neural networks in seizure / S. Aydın, H.M. Saraoğlu, S. Kara // *Ann. Biomed. Eng.* – 2018. – №37 (12). – С. 2626.
2. Chung K.F. Prevalence, pathogenesis, and causes of chronic cough / K.F. Chung, I.D. Pavord // *Lancet.* – 2008. – №371 (9621). – С. 1364–1374
3. Schilt S. Wavelength modulation spectroscopy: combined frequency and intensity laser modulation / S. Schilt, L Thvenaz, P. Robert // *Appl. Opt.* – 2003. – №42. – С. 6728–6738
4. Roman R.C. Prediction of Problematic Wine Fermentations Using Artificial Neural Networks / R.C. Roman, O.G. Hernandez, U.A. Urtubia // *Bioprocess and Biosystems Engineering.* – 2011. – Vol. 34. – С. 1057–1065.
5. Efrogmson M. Multiple regression analysis / M. Efrogmson // *Mathematical methods for digital computers.* – 2019. – №1. – С.191–203.

6. Мокшин В.В., Якимов И.М. Метод формирования модели анализа сложной системы / В.В. Мокшин, И.М. Якимов // Информационные технологии. – 2011. – №5. – С. 46–51
7. Laguarda J. COVID-19 artificial intelligence diagnosis using only cough recordings / J. Laguarda, F. Hueto, B. Subirana. // IEEE Open Journal of Engineering in Medicine and Biology. – 2020. – №1. – С. 275–281.