DOI: 10.15514/ISPRAS-2020-32(4)-10

Диагностика гипертрофий левых отделов сердца с помощью глубокой нейронной сети

^{1,2} П.К. Андреев, ORCID: 0000-0003-4907-9808 <andreev.pk@phystech.edu>^{1,3} В.В. Ананьев, ORCID: 0000-0002-5070-8117 <surval53@gmail.com>^{1,3} В. А. Макаров, ORCID: 0000-0001-6744-4130 <vladimir.makarov@novsu.ru>^{1,2,4} Е. А. Карпулевич, ORCID: 0000-0002-6771-2163 <karpulevich@ispras.ru>^{1,5} Д. Ю. Турдаков, ORCID: 0000-0001-8745-0984 <turdakov@ispras.ru>¹ Институт системного программирования им. В.П. Иванникова РАН, 109004, Россия, г. Москва, ул. А. Солженицына, д. 25² Московский физико-технический институт, 141701, Россия, г. Долгопрудный, Институтский пер., 9³ Новгородский государственный университет имени Ярослава Мудрого, 173003, Россия, г. Великий Новгород, Большая Санкт-Петербургская ул., 41⁴ Национальный исследовательский центр "Курчатовский институт", 123182, г. Москва, пл. Академика Курчатова, 1⁵ Московский государственный университет имени М.В. Ломоносова, 119991, г. Москва, ул. Ленинские Горы, 1

Аннотация. В настоящей работе представлены результаты применения сверточной нейронной сети для диагностики гипертрофий левых отделов сердца посредством анализа электрокардиограмм (ЭКГ) в 12 стандартных отведениях. В ходе исследования был собран и обработан новый уникальный набор данных, содержащий 64 тысячи записей ЭКГ. На основе сопутствующих записям заключений были сформированы метки принадлежности к двум рассматриваемым классам: гипертрофия левого желудочка и гипертрофия левого предсердия. Набор сигналов и выделенные метки были использованы для обучения глубокой сверточной нейронной сети с остаточными блоками, получившаяся модель способна детектировать гипертрофию левого желудочка с качеством по F-мере выше 0.82 и гипертрофию левого предсердия с качеством выше 0.78. Кроме того, был осуществлен поиск оптимальной архитектуры нейросети, произведена экспериментальная оценка эффекта от включения в модель метаданных пациентов и предобработки сигнала, а также сделан сравнительный анализ трудности детектирования гипертрофий левых отделов по отношению к двум другим часто встречающимся нарушениям сердечной активности – мерцательной аритмии и блокады левой ножки пучка Гиса.

Ключевые слова: нейронные сети; ЭКГ; электрокардиография; машинное обучение; гипертрофия

Для цитирования: Андреев П.К., Ананьев В.В., Макаров В.А., Карпулевич Е.А., Турдаков Д. Ю. Диагностика гипертрофий левых отделов сердца с помощью глубокой нейронной сети. Труды ИСП РАН, том 32, вып. 4, 2020 г., стр. 141–154. DOI: 10.15514/ISPRAS-2020-32(4)-10

Благодарности: Исследование поддержано грантом РФФИ 19-57-06004 МНТИ_а.

Diagnosis of left atrial and left ventricular hypertrophies using a deep neural network

^{1,2} P.K. Andreev, ORCID: 0000-0003-4907-9808 <andreev.pk@phystech.edu>^{1,3} V.V. Ananov, ORCID: 0000-0002-5070-8117 <surval53@gmail.com>^{1,3} V. A. Makarov, ORCID: 0000-0001-6744-4130 <vladimir.makarov@novsu.ru>^{1,2,4} E. A. Karpulevich, ORCID: 0000-0002-6771-2163 <karpulevich@ispras.ru>^{1,5} D. Y. Turdakov, ORCID: 0000-0001-8745-0984 <turdakov@ispras.ru>¹ Ivannikov Institute for System Programming of the Russian Academy of Sciences, 25, Alexander Solzhenitsyn st., Moscow, 109004, Russia² Moscow Institute of Physics and Technology,⁹ Institutskiy per., Dolgoprudny, Moscow Region, 141701, Russia³ Yaroslav-the-Wise Novgorod State University,

41, Bolshaya Saint-Petersburgskaya st., Veliky Novgorod, Novgorod region, 173003, Russia

⁴ National Research Center «Kurchatov Institute»,

1, Academician Kurchatov sq., Moscow, 123182, Russia

⁵ Lomonosov Moscow State University,

1, Leninskie Gory st, Moscow, 119991, Russia

Abstract. This paper presents the results of the application of a convolutional neural network to diagnose left atrial and left ventricular hypertrophies by analyzing 12-lead electrocardiograms (ECG). During the study, a new unique dataset containing 64 thousand ECG records was collected and processed. Labels for the two classes under consideration, left ventricular hypertrophy and left atrial hypertrophy, were generated from the accompanying medical reports. A set of signals and obtained labels were used to train a deep convolutional neural network with residual blocks; the resulting model is capable of detecting left ventricular hypertrophy with F-score more than 0.82 and left atrial hypertrophy with F1-score over 0.78. In addition, the search for optimal neural network architecture was carried out and the experimental evaluation of the effect of including patient metadata into the model and signal preprocessing was conducted. Besides, the paper provides a comparative analysis of the difficulty of detecting left ventricular and left atrial hypertrophies in relation to the other two frequently occurring heart activity disorders, namely atrial fibrillation and left bundle branch block.

Keywords: neural networks; ECG; electrocardiography; machine learning; hypertrophy

For citation: Andreev P.K., Ananov V.V., Makarov V.A., Karpulevich E.A., Turdakov D.Y. Diagnosis of left atrial and left ventricular hypertrophies using a deep neural network. Trudy ISP RAN/Proc. ISP RAS, vol. 32, issue 4, 2020. pp. 141–154 (in Russian). DOI: 10.15514/ISPRAS-2020-32(4)-10

Acknowledgements. The reported study was funded by RFBR according to the research project № 19-57-06004.

1. Введение

Электрокардиография как эффективный метод инструментальной диагностики в кардиологии непрерывно совершенствуется. На основе исследований создаются новые подходы к интерпретации электрокардиограмм. Для синтеза непротиворечивого диагностического заключения врачу зачастую приходится проводить многофакторный глубокий анализ с проверкой большого числа критериев, рассмотрением сочетания специфических признаков и внешних факторов. На качество диагностического заключения сильно влияет индивидуальная подготовка врача. Для существенного сокращения трудозатрат и повышения качества заключения, формулируемого врачом, постоянно совершенствуются и внедряются в практику автоматизированные инструменты.

В настоящей статье представлены результаты применения методов машинного обучения для диагностики гипертрофии левых отделов сердца. Гипертрофия желудочков и предсердий относятся к числу сложно диагностируемых заболеваний, которые часто встречаются в сочетании с другими патологиями сердца, например, блокадами. Известно не менее десятка

различных критериев диагностики гипертрофий, зависящих от конкретного набора ЭКГ-признаков и учитывающих такие факторы как пол, возраст, раса, вес пациента, площадь поверхности тела [1, 2].

Гипертрофия левого желудочка (aleft ventricular hypertrophy, LVH) – достаточно распространенное заболевание. Оно выявляется примерно у 20% обследуемых в возрасте 50 лет и этот процент растет с увеличением возраста пациентов. Поскольку существуют препараты лечения LVH, эффективность которых подтверждена клинической практикой, задача своевременного выявления этой патологии представляется весьма актуальной [1].

Как известно, патологии предсердий выявляются на основе анализа формы и длительности зубца р, для их обозначения используются термины: увеличение предсердия (atrial enlargement), патология предсердия (atrial abnormality), гипертрофия предсердия (atrial hypertrophy). В диагностических заключениях набора ЭКГ, на котором построено исследование, чаще используется термин «гипертрофия», его будем использовать в тексте статьи.

Гипертрофия левого предсердия (left atrial hypertrophy, ЛАН) является, как правило, важным первичным индикатором нарушения работы сердца. Изменение функционирования левого предсердия может стать причиной фибрилляции предсердий, сердечной недостаточности, инсульта, транзиторной ишемической атаки, острого инфаркта миокарда [3, 4]. Раннее выявление патологий левого предсердия и своевременная терапия позволяют предотвратить необратимые негативные последствия.

Электрокардиография не является наилучшим методом выявления гипертрофии. Гораздо лучшее качество диагностики дает эхокардиография. У пациентов с клинически значимой гипертрофией, которая обнаруживается по результатам эхокардиографии, ЭКГ может не содержать диагностических признаков заболевания. Однако этот факт не исключает выявления указанных патологий методом электрокардиографии. Во-первых, чувствительность и специфичность алгоритмов выявления гипертрофий на ЭКГ достаточно высоки и уступают ЭхоЭКГ на 20-35% [1]. Для людей старшей возрастной группы (> 40 лет) эти показатели находятся на вполне приемлемом уровне. Во-вторых, электрокардиография является более дешевым и наиболее часто применяемым методом скрининга, не требующим участия высококвалифицированного специалиста непосредственно для снятия ЭКГ. Использование портативных инструментов в сочетании с техниками телемедицины позволяет пациенту самостоятельно снять ЭКГ и предоставить ее врачу.

За последнее десятилетие были достигнуты значительные успехи в создании алгоритмов машинного обучения, основанных на глубоких нейронных сетях. Их главной отличительной чертой является способность определять сложные закономерности в данных, так, с помощью обучения глубоких нейронных сетей были достигнуты существенные успехи в компьютерном зрении [5], распознавании речи [6], обработке естественного языка [7], а также ряде медицинских приложений [8, 9], в том числе в автоматической интерпретации ЭКГ [10, 11]. В частности, было показано, что качество диагностики некоторых состояний при использовании данных алгоритмов может превышать уровни, соответствующие врачам среднего уровня квалификации, а также классическим коммерческим программам анализа ЭКГ [12]. В силу описанных выше причин исследование применимости глубоких нейронных сетей для диагностики гипертрофий левых отделов сердца приобретает особую актуальность. Так, в рамках работы [10] для определения ритма ЭКГ была разработана сверточная нейронная сеть с остаточными блоками. Модель была обучена предсказывать тип ритма по отрезку сигнала длительностью 1.3 секунды с одного электрокардиографического отведения. Для обучения модели был собран и аннотирован новый набор данных, содержащий 91,232 тридцатисекундных ЭКГ сигналов, соответствующих 53,549 пациентов. Было продемонстрировано, что качество автоматического определения каждого из 12 рассматриваемых типов ритма (фибрилляция предсердий, атриовентрикулярная блокада I

степени, эктопический предсердный ритм, бигеминия, тригеминия, идиовентрикулярный ритм, узловой ритм, синус ритм, наджелудочковая тахикардия, желудочковая тахикардия, атриовентрикулярная блокада II степени тип 1 и зашумленный ритм) сравнимо или превосходит усредненное качество диагностики нескольких практикующих кардиологов. Несмотря на многообещающие результаты, рассмотрение в работе лишь одного из двенадцати стандартных отведений и только ритмических типов патологий ограничивает применение данного исследования в клинической практике.

В работе [11], напротив, использовался набор данных, содержащий сигналы ЭКГ в двенадцати отведениях. Набор состоит из 2,470,424 записей, соответствующих 1,676,384 пациентов. Длина каждой записи составляет от 7 до 10 секунд. Данные использовались для обучения нейронной сети подобной [10], но с большим количеством входных каналов и меньшей глубиной. В работе рассматривались следующие пересекающиеся классы отклонений сердечной активности от нормы: атриовентрикулярная блокада I степени, фибрилляция предсердий, блокада левой ножки пучка Гиса, блокада правой ножки пучка Гиса, синусовая тахикардия и синусовая брадикардия. Так же, как и в [10], полученные в данной работе результаты позволяют с оптимизмом смотреть на возможность практического применения нейронных сетей для диагностики заболеваний сердца, так как качество анализа признаков ЭКГ по F-мере в данной работе оказалось выше, чем у ряда врачей специалистов. Таким образом, несмотря на существенный прогресс в вопросе применимости нейронных сетей для интерпретации электрокардиограмм, использование данных алгоритмов для диагностики гипертрофий левых отделов сердца в настоящий момент не является полностью изученным.

2. Данные

В рамках настоящей работы в результате взаимодействия с несколькими медицинскими центрами в Великом Новгороде был получен новый, уникальный набор данных, содержащий записи ЭКГ в 12 стандартных отведениях. В наборе содержатся 64 тысячи анонимизированных записей ЭКГ, соответствующие 36 тысячам пациентов в возрасте от 13 до 95 лет (распределение пациентов по возрасту показано на Рис. 1 (а)). Длительность каждой записи составляет 4 секунды.

Каждый экземпляр ЭКГ сопровождался описанием заключения, которое было сформулировано врачом клиники. Из текста заключений по ключевым словам (табл. 1) были выделены бинарные метки принадлежности к двум рассматриваемым классам (гипертрофия левого желудочка и гипертрофия левого предсердия), при этом данные классы могут пересекаться, что и проиллюстрировано на круговой диаграмме рис. 1 (b).

Табл. 1. Словосочетания, покрываемые регулярными выражениями, использованными для получения меток принадлежности к классам

Table 1. Phrases covered by regular expressions used to get class membership labels

Класс	Словосочетания
Гипертрофия левого желудочка	Гипертрофия левого желудочка, Гипертрофия левых отделов, Гипертрофия обоих желудочков
Гипертрофия левого предсердия	Гипертрофия левого предсердия, Гипертрофия левых отделов, Гипертрофия обоих предсердий

Рис 1. (а): Гистограмма распределения пациентов по возрасту. (б): Круговая диаграмма распределения ЭКГ записей по следующим группам: LVH – гипертрофия левого желудочка (при этом отсутствует гипертрофия левого предсердия), LVH & LAH – гипертрофия левых отделов (гипертрофия левых желудочка и предсердия одновременно), LAH – гипертрофия левого предсердия (при этом отсутствует гипертрофия левого желудочка), NH – отсутствие гипертрофий левых отделов

Fig 1. a): Distribution of patients by age. (b): Distribution of ECG records into the following groups: LVH - left ventricular hypertrophy only (left atrial hypertrophy is absent), LVH & LAH - left ventricular and atrial hypertrophy at the same time, LAH - left atrial hypertrophy only (left ventricular hypertrophy is absent), NH - left ventricular hypertrophy and left atrial hypertrophy are absent

3. Модель

В качестве алгоритма классификации электрокардиограмм была использована сверточная нейронная сеть с остаточными блоками, схожая с использованной в работе [11], архитектура сети приведена на рис. 2.

Входные данные представляют собой ЭКГ сигнал с 8 отведений (I, II, VI, V2, V3, V4, V5, V6) в виде матрицы с размерностями 768×8 , где 768 – число измерений (при частоте дискретизации 250 ГЦ соответствует длине записи в 3.072 секунды), 8 – число отведений (ввиду линейной зависимости остальных стандартных отведений III, aVR, aVF, aVL от I и II, их включение в модель не имеет смысла), а также метаданные пациента: возраст – натуральное число от 13 до 95, пол – число 0 (мужской) или 1 (женский).

ЭКГ-сигнал подается на вход нескольким (N) последовательным остаточным блокам [13], выделяющие релевантные для классификации признаки. Каждый остаточный блок представляет собой два сверточных модуля, состоящих из последовательных слоев батч нормализации [14], нелинейности (ReLU), дропаута (англ. dropout [15]) и свертки, начало и конец остаточного блока соединены остаточной связью (англ. skip connection), кроме того, для понижения пространственной размерности каждые k ($= 3$) блоков в остаточной связи применяется операция пулинга (соответственно, в одном из сверточных блоков шаг фильтра больше на единицу).

Выход сверточной части модели конкатенируется с нормализованными метаданными пациентов и подаются на вход двум последовательным полносвязным слоям, сопровождающимися сигмоидной функцией активации. В результате для каждого образца модель выдает 2 числа в промежутке от 0 до 1, которые могут интерпретироваться как вероятности принадлежности к 2 рассматриваемым классам. Таким образом, в настоящей работе использовалась одна модель для решения обеих задач бинарной классификации, альтернативный подход мог бы заключаться в использовании двух независимых моделей для каждой из задач в отдельности, однако гипертрофии левых отделов часто встречаются в

комбинации, кроме того, признаки релевантные для детекции данных состояний могут пересекаться.

Рис 2. Архитектура использованной нейросети. Conv – операция свертки, ResBlk – остаточный блок, BN – батч-нормализация, ReLU – функция нелинейности, Drop – операция дропаута, MaxPool – операция пулинга функцией максимума, Lin – полносвязный линейный слой, Concat – конкатенация векторов, Sigm – сигмоидная функция активации

Fig 2. Neural network architecture. Conv - unidimensional convolution operation, ResBlk - residual block, BN - batch normalization, ReLU - nonlinearity function, Drop - dropout operation, MaxPool - pooling by the maximum function, Lin - fully connected linear layer, Concat - vector concatenation, Sigm - element-wise sigmoid function

4. Методология и результаты экспериментов

Рассматриваемая модель обучалась на описанном в разделе 2 наборе данных с помощью алгоритма обратного распространения ошибки [16], при этом на каждой итерации обучения («эпохе») из записей ЭКГ длиной 4 секунды вырезался случайный участок длительностью 3 секунды, что позволяет значительно снизить эффект переобучения. В качестве функции потерь использовалась средняя по классам бинарная кросс-энтропия, которая оптимизировалась на обучающей выборке по методу Adam [17] с параметрами $\beta_1 = 0.9$, $\beta_2 = 0.999$, размером батча 64, начальной скоростью обучения (learning rate), равной 0.001, и её понижением с мультипликативным фактором 0.1 при отсутствии значимых улучшений на валидационной выборке в течение нескольких эпох. Во всех представленных ниже экспериментах, если не оговорено иначе, вероятность дропаута была фиксирована 0.5, начальное число каналов – 64, размер сверточных ядер – 7, число блоков – 9.

4.1 Поиск оптимальной архитектуры

Для поиска оптимальной архитектуры модели был произведен перебор ряда гиперпараметров: числа остаточных блоков, начального числа каналов в нейросети (её «толщины») и размера сверточного ядра. Перебор осуществлялся с помощью кросс-валидации на 5 частей, при этом при подборе каждого из рассматриваемых гиперпараметров остальные оставались фиксированными заданным значениям: число блоков – 9, начальное число каналов – 64, размер сверточных ядер – 7. Кросс-валидация проводилась по следующей

методологии: исходный набор данных был разделен на 5 приблизительно равных частей, каждая из которых содержит по 12 тысяч ЭКГ, шаг кросс-валидации заключался в последовательном выборе каждого из пяти наборов, разделения его на две равные части (тестовую и валидационную), объединения остальных четырех частей в тренировочную выборку, запуске процесса обучения на тренировочной выборке с ранним остановом по значению метрики на валидационной выборке и оценке качества модели на тестовой части набора. Результаты экспериментов приведены на рис. 3.

Рис 3. Зависимость площади под ROC-кривой от гиперпараметров (для каждой точки проведена кросс-валидация на 5 частей, доверительные интервалы оценены с помощью бутстрапа)

Fig 3. Dependence of the area under the ROC curve on hyperparameters (for each point, 5-fold cross-validation was carried out, the confidence intervals were estimated using bootstrap)

Как видно по рис. 3 (a), оптимальные результаты применения модели получены при использовании примерно 9 остаточных блоков и дальнейшее увеличение глубины приводит к ухудшению качества (на исследованном участке от 3 до 12 блоков), что находится в согласии с результатами работы [11] (с точностью до нескольких блоков, что связано с неполной идентичностью архитектур). Кроме того, по рис 3 (b) можно видеть, что качество классификации незначительно зависит от количества входных каналов на исследованном диапазоне, таким образом, «глубина» сети играет более значительную роль чем её «толщина». Зависимость качества модели от размера сверточного ядра приведена на рис 3 (c); видно что оптимальное значение достигается при размере ядра, равном 7, при этом зависимость довольно слабая.

4.2 Зависимость качества классификации от размера обучающей выборки

Оценка характера зависимости качества классификации от размера обучающей выборки проводилась по методологии схожей с использованной в пункте 4.1. На каждом шаге кросс-валидации алгоритм обучался (начиная со случайной инициализации) на разном количестве образцов в обучающей выборке, при этом размеры валидационной и тестовых выборок оставались фиксированными. Результаты приведены на рис. 4, на котором также для сравнения представлены результаты аналогичного эксперимента для двух часто рассматриваемых (например, в [11]) классов нарушений – фибрилляция предсердий и блокада левой ножки пучка Гиса.

Из сравнения рис. 4(a) и рис. 4(b) можно сделать вывод о том, что диагностика гипертрофий на основе ЭКГ является более сложной задачей, чем выявление фибрилляции предсердий и блокады левой ножки пучка Гиса. Действительно, качество диагностики данных состояний гораздо раньше выходит на высокое значение (аналогичные соотношения наблюдаются и для других метрик, например, средней точности и F-меры). Также из графиков можно видеть, что качество диагностики рассматриваемых состояний монотонно увеличивается при увеличении тренировочной выборки, таким образом, можно ожидать значительного прироста качества при дальнейшем пополнении обучающего набора.

Рис 4. Зависимость площади под ROC-кривой от размера использованной обучающей выборки выраженного в долях от 48 тысяч образцов (максимально доступный размер обучающей выборки на кросс-валидации) (a): для классификации гипертрофий левых отделов (ГЛЖ и ГЛП), (b): для фибрилляции предсердий (англ. Atrial Fibrillation, AF) и блокады левой ножки пучка Гиса (англ. Left Bundle Branch Block, LBBB)

Fig 4. Dependence of the area under the ROC curve on the size of the training sample expressed in fractions of 48 thousand samples (the maximum available size of the training sample for cross-validation) (a): for left ventricular hypertrophy (LVH) and left atrial hypertrophy (LAH), (b): for atrial fibrillation (AF) and left bundle branch block (LBBB)

4.3 Влияние предобработки сигнала и включения в модель метаданных

Для повышения качества предсказаний модели в настоящей работе были использованы техники предобработки сигнала: коррекция изолинии с помощью локального взвешенного сглаживания (LOWESS [18]) и удаление высокочастотного шума с помощью дискретного вейвлет-преобразования [19] (с мягким порогом VisuSrink [20], материнская вейвлет Symlet – 8, уровень декомпозиции – 4). Кроме того, исследовалась зависимость качества предсказаний от факта включения в модель метаданных пациентов, пола и возраста. Влияние данных факторов проиллюстрировано на рис. 5 в виде диаграмм размаха.

Рис 5. Средняя по классам площадь под ROC-кривой в зависимости от методологии (кросс-валидация на 5 частей). baseline – обучение на непереработанных сигналах и без включения в модель информации о метаданных пациентов, PREP – обучение на предобработанных сигналах, MET – обучение с включением в модель информации о метаданных пациентов

Fig 5. Class-average area under the ROC curve, depending on the methodology (5-fold cross-validation). baseline - training on unprocessed signals and without including patient metadata information in the model, PREP - training on preprocessed signals, MET - training with patient metadata information included in the model

Предобработка сигнала и включение в модель метаданных пациентов статистически значимо увеличивают качество детекции рассматриваемых состояний как в совокупности, так и по отдельности (критерий Уилкоксона, $p < 0.05$). В целом, средняя площадь под ROC-кривой увеличилась почти на 2%, что согласно рис. 4(а) соответствует аналогичному приросту в качестве при увеличении тренировочной выборки в 3 раза. Также было выяснено, что основной вклад в прирост качества, при удалении шума из сигнала, привносит коррекция изолинии, в то время как удаление высокочастотного шума не влияет на результаты значимо. Аналогично при добавлении в модель метаданных прирост качества главным образом обусловлен информацией о возрасте пациентов.

4.4 Выводы

Итоговые метрики для модели с оптимальным (из рассмотренных) набором гиперпараметров (количество блоков, начальное число каналов и размер сверточного ядра), при использовании предобработки сигнала и включении в модель метаданных приведены в табл. 2. Оценка данных значений также проводилась на кросс-валидации на 5 частей, при этом для подсчета метрик, требующих предсказания модели в виде индикаторных меток (F-мера, Точность, Чувствительность, Специфичность) была произведена бинаризация предсказаний модели с порогом, подобранным на валидационной выборке по максимуму F-меры.

Табл. 2. Доверительные интервалы (по t-распределению Стьюдента) для метрик качества обнаружения рассматриваемых состояний с помощью разработанной нейронной сети (кросс-валидация на 5 частей)

Table 2. Confidence intervals (by Student's t-distribution) for metrics of detection of the abnormalities under consideration using the developed neural network (5-fold cross-validation).

Класс	ROC-AUC	Чувствительность Sensitivity	Специфичность Specificity	Точность Precision	F-мера F1-score
ГЛЖ	0.840±0.006	0.927±0.016	0.559±0.04	0.738±0.017	0.821±0.006
ГЛП	0.889±0.007	0.864±0.02	0.766±0.012	0.717±0.013	0.783±0.011

Таким образом, в процессе проведения настоящего исследования были решены следующие задачи:

1. Собран и обработан уникальный набор данных, содержащий записи ЭКГ в двенадцати стандартных отведениях.
2. Разработана сверточная нейронная сеть, способная детектировать гипертрофию левого желудочка с качеством по F-мере свыше 0.82 и гипертрофию левого предсердия с качеством свыше 0.78.
3. Осуществлен поиск оптимальной архитектуры и произведена экспериментальная оценка эффекта от включения в модель метаданных пациентов и предобработки сигнала.
4. Проведен сравнительный анализ трудности детектирования гипертрофий левых отделов по отношению к двум другим часто встречающимся нарушениям сердечной активности – мерцательной аритмии и блокады левой ножки пучка Гиса.

5. Заключение

Электрокардиографические критерии, используемые на практике для выявления гипертрофий левых отделов сердца, как правило, имеют высокую специфичность (около 90%) и умеренную чувствительность (30-60%), например, критерий напряжения Корнелла для определения гипертрофии левого желудочка имеет специфичность равную 85-95% при чувствительности 30-50% [1]. Как видно из рис. 6, показатели качества диагностики с помощью разработанного алгоритма находятся примерно на том же уровне (при специфичности 90%, чувствительности определения ЛАН и LVH равны 48% и 62% соответственно), при этом не требуют участия специалиста в процессе интерпретации. Более того, появляется естественная возможность для балансирования между чувствительностью и специфичностью посредством варьирования порогов на присвоенные моделью вероятности обнаружения рассматриваемых состояний (см. рис. 6).

Рис 6. Кривые чувствительность-специфичность и их характерные точки: красная – точка, соответствующая максимуму F-меры на валидационной выборке (см. Табл. 2), синяя – точка равенства чувствительности и специфичности, зеленая – точка кривой со специфичностью равной 90%. (а): кривая для диагностики гипертрофии левого желудочка, (б): кривая для диагностики гипертрофии левого предсердия

Fig 5. Sensitivity-specificity curves and their characteristic points: red - the point corresponding to the maximum of F1-score on the validation sample (see Table 2), blue - the point of equality of sensitivity and specificity, green - the point of the curve with specificity equal to 90%. (a): curve for the diagnosis of left ventricular hypertrophy, (b): curve for the diagnosis of left atrial hypertrophy

Существенным ограничением настоящего исследования является способ получения разметки использованного набора данных. Полученная разметка по нашим оценкам содержит значительное количество ошибок как ввиду способа выделения меток принадлежности к классам (по ключевым словам), так и ввиду ошибок в самих заключениях. Количество

ошибок, возникающих из-за несовершенства метода анализа заключений, может быть снижено посредством использования более продвинутых алгоритмов выделения меток, так, в работе [11] была разработана методология выделения меток на основе ленивой ассоциативной классификации (lazy associative classifier) [21] по выделенным из заключений n-граммам. Ошибки же в самих заключениях не могут быть достоверно исправлены без привлечения нескольких экспертов для перерасметки всего набора данных, однако возможно снизить влияние неправильно размеченных образцов на процесс обучения [22, 23], что является одним из перспективных направлений дальнейших исследований.

Вероятно, наиболее значимым направлением дальнейшего изучения применения нейронных сетей в контексте интерпретации ЭКГ является обобщение полученных результатов на более широкий спектр аномалий сердечной активности, в особенности на состояния, угрожающие жизни пациента. Многие из таких состояний крайне редко встречаются в клинической практике, и в частности в наборе данных, использованном в настоящей работе, что делает затруднительным применение к их диагностике рассмотренных методик. В этой связи особенно актуальным может быть дальнейшее развитие техник аугментаций [24] и самообучения (self-supervised learning) [25] в ЭКГ домене.

Таким образом, в настоящей работе продемонстрирована эффективность применения глубоких нейронных сетей для диагностики гипертрофий левых отделов сердца. Данное исследование, совместно с рядом других работ [10, 11, 12], показывает преимущества парадигмы построения систем поддержки принятия решений в клинической практике на основе алгоритмов глубокого обучения. Дальнейшее развитие данной парадигмы потенциально может привести к значительному снижению затрат человеческих ресурсов на интерпретацию ЭКГ, и вместе с тем снизить количество врачебных ошибок.

Список литературы / References

- [1] Gertsch Marc. The ECG: a two-step approach to diagnosis. Springer-Verlag Berlin Heidelberg, 2004, 615 p.
- [2] Kozłowski D. Method in the Chaos – a step-by-step approach to ECG interpretation. *European Journal of Translational and Clinical Medicine*, vol. 1, no. 1, pp. 74-90.
- [3] Donal Erwan et al. EACVI/EHRA Expert Consensus Document on the role of multi-modality imaging for the evaluation of patients with atrial fibrillation. *European Heart Journal - Cardiovascular Imaging*, vol. 17, no. 4, 2016, pp. 355–383.
- [4] Tsang T.S., Abhayaratna W.P., Barnes M.E. et al. Prediction of cardiovascular outcomes with left atrial size: is volume superior to area or diameter? *Journal of the American College of Cardiology*, vol. 47, no. 5, 2006, pp. 1018-1023.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1026-1034.
- [6] Amodei Dario et al. Deep speech 2: End-to-end speech recognition in English and Mandarin. In *Proc. of the International conference on machine learning*, 2016, pp. 173–182.
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention Is All You Need. In *Proc. of the 31st Conference on Neural Information Processing Systems*, 2017, pp. 5998-6008.
- [8] Bejnordi Babak Ehteshami et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Journal of the American Medical Association*, vol. 318, no. 22, 2017, pp. 2199-2210.
- [9] De Fauw Jeffrey, Ledsam Joseph R., Romera-Paredes Bernardino et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature medicine*, vol. 24, no. 9, 2018, pp. 1342–1350.
- [10] Hannun Awni Y., Rajpurkar Pranav, Haghighpanahi Masoumeh et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, vol. 25, no. 1, 2019, pp. 65-69.
- [11] Ribeiro Antônio H., Ribeiro Manoel Horta, Paixão Gabriela M.M. et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature communications*, vol. 11, no. 1, 2020, pp. 1–9

- [12] Smith Stephen W., Walsh Brooks, Grauer Ken et al. A deep neural network learning algorithm outperforms a conventional algorithm for emergency department electrocardiogram interpretation. *Journal of electrocardiology*, vol. 52, 2019, pp. 88-95.
- [13] He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian. Deep residual learning for image recognition. In *Proc. of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. 0
- [14] Ioffe Sergey, Szegedy Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [15] Srivastava Nitish, Hinton Geoffrey, Krizhevsky Alex et al. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, vol. 15, no. 1, 2014, pp. 1929-1958.
- [16] Rumelhart David E., Hinton Geoffrey E., and Williams Ronald J. Learning representations by back-propagating errors. *Nature*, vol. 323, 1986, pp. 533-536.
- [17] Kingma Diederik P., Ba Jimmy. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [18] Cleveland William S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, vol. 74, no. 368, 1979, pp. 829-836.
- [19] Kania Michał, Fereniec Małgorzata, and Maniewski Roman. Wavelet denoising for multi-lead high resolution ECG signals. *Measurement Science Review*, vol. 7, no. 4, 2007, pp. 30-33.
- [20] Donoho David L., and Johnstone Iain M. Threshold selection for wavelet shrinkage of noisy data. In *Proc. of the 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1994, pp. A24–A25.
- [21] Veloso Adriano, Wagner Meira, and Zaki Mohammed J. Lazy associative classification. In *Proc. of the Sixth International Conference on Data Mining (ICDM'06)*, 2006, pp. 645–654.
- [22] Nguyen Duc Tam et al. Self: Learning to filter noisy labels with self-ensembling. *arXiv preprint arXiv:1910.01842*, 2019.
- [23] Huang Lang, Zhang Chao, and Zhang Hongyang. Self-Adaptive Training: beyond Empirical Risk Minimization. *arXiv preprint arXiv:2002.10319*, 2020.
- [24] Guryanova Valeriia. Online Augmentation for Quality Improvement of Neural Networks for Classification of Single-Channel Electrocardiograms. In *Proc. of the International Conference on Analysis of Images, Social Networks and Texts*, 2019, pp. 37–49.
- [25] Kiyasseh Dani, Zhu Tingting, and Clifton David A. CLOCS: Contrastive Learning of Cardiac Signals. *arXiv preprint arXiv:2005.13249*, 2020.

Информация об авторах / Information about authors

Павел Константинович АНДРЕЕВ – выпускник Московского физико-технического института, сотрудник отдела «Информационные системы». Сфера его научных интересов включает биомедицинский анализ данных, глубокое обучение, робастные методы машинного обучения, а также обучение с частичным привлечением учителя.

Pavel Konstantinovich ANDREEV is a graduate of the Moscow Institute of Physics and Technology, an employee of the Information Systems Department. His research interests include biomedical data science, deep learning, robust machine learning, and semi-supervised learning.

Владислав Валерьевич АНАНЬЕВ является магистрантом и ассистентом кафедры информационных технологий и систем НовГУ, сотрудник ИСП РАН. Сфера научных интересов: анализ и разметка данных из различных сфер деятельности, глубокое обучение, компьютерное зрение и обработка изображений.

Vladislav Valerievich ANANAYEV is a graduate of the magistracy and assistant of the Department of Information Technologies and Systems, Novgorod State University, an employee of ISP RAS. Area of research interests: data labeling and analysis for various fields of activity, deep learning, computer vision and image processing.

Владимир Алексеевич МАКАРОВ – кандидат технических наук, старший научный сотрудник. Сфера научных интересов: анализ бинарного кода, искусственный интеллект для персонализированной медицины.

Vladimir Alexeevich MAKAROV – PhD, Senior scientist. Research interests: binary code analysis, artificial intelligence for personalized medicine.

Евгений Андреевич КАРПУЛЕВИЧ является специалистом отдела «Информационные системы». Сфера научных интересов: применение алгоритмов анализа данных к биомедицинскому домену, разработку систем распределенного хранения и анализа данных.

Evgeny Andreevich KARPULEVICH is a specialist of the Information Systems Department. Research interests: application of data analysis algorithms to the biomedical domain, development of systems for distributed data storage and analysis.

Денис Юрьевич ТУРДАКОВ – к.ф.-м.н., заведующий отделом «Информационные системы» ИСП РАН, доцент МГУ. Сфера научных интересов: машинное обучение, интеллектуальный анализ данных, извлечение информации, обработка естественного языка, сложные сети, анализ социальных сетей, большие данные.

Denis Yurievich TURDAKOV – Ph.D. head of the Information Systems Department at ISP RAS, associated professor at MSU. Research interests: machine learning, data mining, information extraction, natural language processing, complex networks, social network analysis, big data.