

экономической модернизации и научно-технологического обновления всех отраслей экономики. Считаем, что нашей стране требуется очень широкий комплекс различных мер, включающий в себя становление экологически-грамотного сознания, углубление принципиально новых эколого-демократических начал непосредственно в управлении, изобретение и внедрение в нашу хозяйственно-бытовую деятельность современных типов экологической как техники, так и технологии, и многое-многое другое.

Библиографический список

1. Негативное влияние транспорта на окружающую среду // Vtorothodi.ru: сайт о переработке и утилизации отходов. URL: <http://vtorothodi.ru/ecology/vliyanie-transporta-na-okruzhayushhuyu-sredu> (дата обращения: 14.10.2020).
2. Сенина Ю.П., Ветошкин А.Г. Снижение негативного влияния автотранспорта на экологическое состояние окружающей среды // Труды международного симпозиума «Надежность и качество». 2011. URL: <http://cyberleninka.ru/article/n/snizhenie-negativnogo-vliyaniya-avtotransporta-na-ekologicheskoe-sostoyanie-okruzhayushey-sredu> (дата обращения: 14.10.2020).
3. Компанеец А. Как GPS навигаторы помогают уменьшить вредные выбросы? // FacePla.net. URL: <http://www.facepla.net/index.php/content-info/182-gps-reduces-car-emission> (дата обращения: 14.11.2020).
4. Число автомобилей в России превысило 56 млн. / РИА Новости. URL: <https://ria.ru/society/20160220/1377940767.html> (дата обращения: 14.10.2020).

УДК 004.032.26

Кузнецов А.Е., Коньков В.В.

ИСПОЛЬЗОВАНИЕ АППАРАТА НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ СЕРДЕЧНО-СОСУДИСТЫХ ЗАБОЛЕВАНИЙ USING THE DEVICE OF NEURAL NETWORKS FOR PREDICTION OF CARDIOVASCULAR DISEASES

*Филиал «Протвино» государственного университета «Дубна»
Секция «Естественные и инженерные науки»*

Авторы: Кузнецов Алексей Евгеньевич, студент 4 курса направления «Информатика и вычислительная техника» филиала «Протвино» государственного университета «Дубна»; Коньков Владислав Владимирович, студент 4 курса направления «Информатика и вычислительная техника» филиала «Протвино» государственного университета «Дубна».

Authors: Kuznetsov Alexey Evgenievich 4th year students of the direction "Informatics and computer technology" of the branch "Protvino" of the State University "Dubna"; Konkov Vladislav Vladimirovich, 4th year students of the direction "Informatics and computer technology" of the branch "Protvino" of the State University "Dubna".

Научный руководитель: Соколов Анатолий Александрович, доктор физико-математических наук, профессор кафедры информационных технологий филиала «Протвино» государственного университета «Дубна».

Scientific adviser: Sokolov Anatoly Aleksandrovich, doctor of physical and mathematical sciences, professor of the department of information technologies of the branch "Protvino" state University "Dubna".

Аннотация

В данной статье рассмотрена совокупность задач по импорту данных о медицинских показателях пациентов и последующей их обработке с помощью аппарата нейронных сетей. Нейросетевой подход позволяет прогнозировать возможность сердечно-сосудистых заболеваний у пациентов.

Abstract

This article discusses a set of problems for importing data on medical indicators of patients and their subsequent processing using the apparatus of neural networks. The neural network approach makes it possible to predict the possibility of cardiovascular disease in patients.

Ключевые слова: сердечно-сосудистые заболевания, медицинская диагностика, нейронная сеть, искусственный интеллект, прогнозирование, машинное обучение, обучающая выборка, контрольная выборка.

Keywords: cardiovascular diseases, medical diagnostics, neural network, artificial intelligence, forecasting, machine learning, training sample, control sample.

Большинство жизненно важных процессов в организме человека так или иначе связаны с работой сердечно-сосудистой системы. Из-за болезней сердца ежегодно умирает огромное количество людей. Для диагностики сердечных заболеваний необходимо использовать большое количество медицинских показателей пациента. Использование накопленных данных от огромного числа людей позволяет производить более детальный анализ протекания болезней сердца. Для проведения подобных исследований требуется найти инструмент для хранения, преобразования, обработки больших объёмов информации, поиска скрытых зависимостей и совпадений в наборах данных. Необходимым инструментом для этого является использование возможностей искусственного интеллекта – нейронных сетей глубокого обучения.

Симптомы некоторых болезней, таких как сахарный диабет, гастрит, бронхиальная астма и даже язва практически совпадают с симптомами болезней сердца – стенокардией, аритмией. Из-за этого болезни сердца иногда трудно предугадать и классифицировать. Применение аппарата нейронных сетей и машинного обучения с использованием большого объема данных позволяет более точно прогнозировать сердечно-сосудистые заболевания. Научная значимость данной работы заключается в проверке эффективности применения аппарата нейронных сетей для диагностики сердечно-сосудистых заболеваний.

Объектом данной работы являются болезни сердца, такие как ишемические болезни, заболевания кровеносных сосудов, аритмия и врожденные пороки сердца.

Предметом исследования данной работы является наиболее часто используемый для машинного обучения набор данных Cleveland Heart Disease из репозитория UCI с медицинскими показателями, в котором содержатся различные характеристики людей и информация о наличии у них заболевания.

Цель исследования – проанализировать выборку с медицинскими показателями пациентов с помощью искусственной нейронной сети, чтобы спрогнозировать у тестируемых сердечно-сосудистые заболевания.

Задачи:

- Сформировать обучающую выборку из БД Cleveland Heart Disease для обучения нейронной сети ставить в соответствие каждому набору данных правильный ответ – пациент здоров или болен.
- Осуществить нормализацию данных перед их использованием в нейронной сети.
- Выбрать архитектуру сети.
- Осуществить обучение нейронной сети с использованием обучающей выборки.
- Осуществить проверку адекватности обучения на основе тестовой выборки, которая не использовалась в обучении.
- Провести оценку качества работы сети.

Исходя из поставленных задач, наиболее подходящим языком для нейронной сети был выбран язык MATLAB для программы MathWorks MATLAB R2018b, так как в ее состав входит целый комплекс инструментов для использования сетей глубокого обучения, а также визуализации результата.

Для создания выборки был выбран набор данных Cleveland Heart Disease из свободного репозитория UC Irvine в составе которого более 600 наборов данных для машинного

обучения.[5] Он представляет собой csv-файл с численными эквивалентами представленных ниже медицинских параметров, а также информацию о наличии заболевания.

Из набора данных использовалось 13 параметров:

4. Возраст в годах;
5. Пол;
6. Тип боли в груди;
7. Артериальное давление в покое (в мм рт. ст. при поступлении);
8. Содержание холестерина;
9. Уровень сахара в крови натощак;
10. Результаты электрокардиографии в покое;
11. Максимальная частота сердечных сокращений;
12. Стенокардия, вызванная физической нагрузкой;
13. Депрессия ST, вызванная физической нагрузкой;
14. Наклон сегмента ST при пиковых нагрузках;
15. Количество крупных сосудов;
16. Талассемия;

Данные параметры позволяют спрогнозировать наличие болезни сердца у пациента. В базе данных, используемой для обучения нейронной сети, использовались данные 200 пациентов и параметр, отвечающий за состояние человека (1 – болен, 0 – не болен).[6]

Нормализация входных векторов осуществляется методом «mapminmax». [2] Метод обрабатывает матрицы, нормализуя минимальное и максимальное значения каждой строки в диапазоне [1;10]. 13 компонент x каждого входного вектора преобразуются по формуле:

$$X' = a + \frac{X - X_{min}}{X_{max} - X_{min}}(b - a). \quad (1)$$

Где:

- [$a = 1$; $b = 10$];
- X_{min} , X_{max} - минимальное и максимальное значение компонент вектора.

Поскольку проблема обучения имеет последовательный характер, то типом данной нейронной сети является глубокая сеть долгой краткосрочной памяти (LSTM), в которой LSTM-модули группируются в блоки, содержащие LSTM-модули, что характерно для глубоких многослойных нейронных сетей.[3] Это позволяет запоминать значения как на длинные, так и на короткие промежутки времени.

В итоговой конфигурации была использована сеть LSTM сеть с 2-мя полносвязными слоями скрытого блока. Первый скрытый слой содержит 14 нейронов, второй - 7 нейронов. Входной вектор имеет 13 компонентов. Выходной слой содержит 1 нейрон.

Для нейронов сети используется функция активации softmax[4], также известная как нормированный экспоненциал – обобщение мультикласса логистической сигмоидальной функции. Функция активации нейрона выходного слоя преобразует вектор x размерности k в вектор softmax той же размерности, где каждая координата полученного вектора представлена вещественным числом в интервале [0,1] и сумма координат равна единице. Функция softmax представлена формулой:

$$softmax(x)_i = \frac{e^{x_i}}{\sum_{k=1}^k e^{x_k}} \quad (2)$$

Корректировка параметров обучения осуществляется функцией «trainingOptions», которая позволяет задать длину последовательности, максимальное количество эпох и порог градиента. [1]

Точность классификации предсказаний представлена формулой:

$$acc = \frac{k}{n} \quad (3)$$

Где:

- k – количество верно угаданных пациентов, участвующих в выборке,
- n – общее количество пациентов, участвующих в выборке.

Для минимизации общей ошибки LSTM был использован стохастический градиент спуск с импульсом такой, как метод обратного распространения ошибки, развернутый во времени, что позволяет изменять значения весов пропорционально его производной в зависимости от величины ошибки.

При запуске программы веса нейронов могут быть установлены двумя способами - случайно или считыванием значения конфигурации настроенной сети. В процессе обучения веса принимают свои оптимальные значения. Данные для обучения считываются из базы данных автоматически.

Результаты нашего исследования показали, что в ходе работы с большим объемом медицинских данных с использованием нейронной сети долгой краткосрочной памяти (LSTM), цель исследования была достигнута. Обучение сети с использованием обучающей выборки продемонстрировало точность обучения 94%, что подтверждает выбор LSTM как эффективного типа нейронных сетей.

Был взят во внимание тот факт, что нейронная сеть не всегда «понимает», что от нее требуется, а ищет то, что проще всего обобщить. В связи с этим была произведена проверка адекватности обучения на основе тестовой выборки, которая не участвовала в обучении. В этой выборке содержится 100 векторов.

Для тестовой выборки были вычислены чувствительность и специфичность сети.

Чувствительность вычислялась как отношение числа правильных ответов для больных, к истинному числу больных. Она позволяет оценить вероятность того, что больной будет классифицирован как больной.

Специфичность отображает вероятность того, что здоровые будут классифицированы именно как здоровые, вычислялась как отношение числа правильных ответов здоровых, к истинному числу здоровых.

На Рис. 1 проиллюстрирован график зависимости точности прогнозирования от количества эпох обучения. Синим цветом выделена точность для обучающей выборки, красными точками – точность для тестовой выборки.

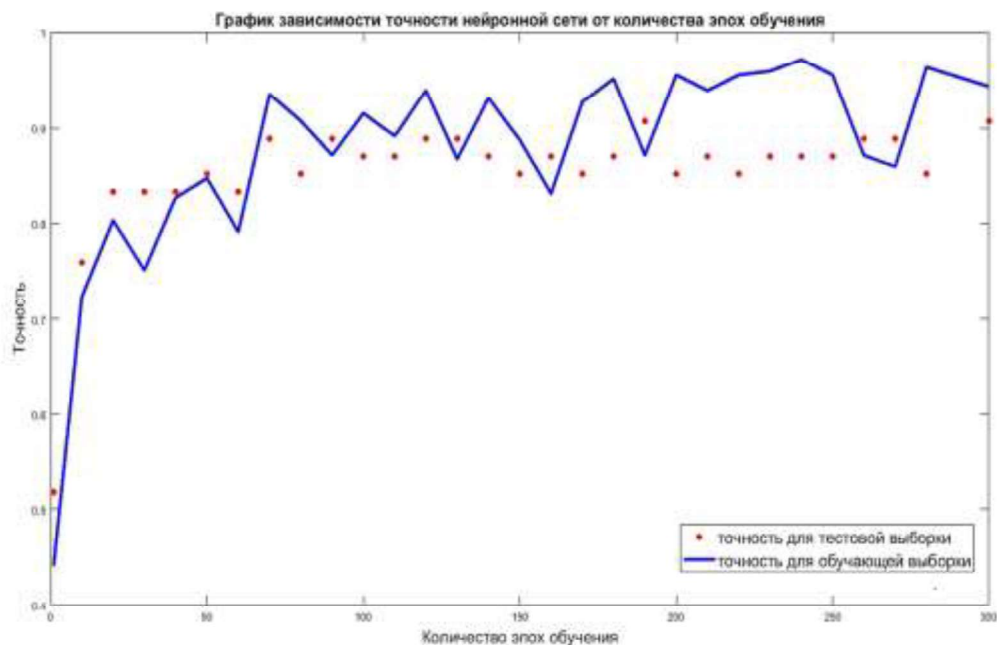


Рис. 1. График зависимости точности нейронной сети от количества эпох обучения

Таким образом, после проверки адекватности работы программы с использованием тестовой выборки, основанной на данных 100 пациентов, максимальная точность составила

91%. Это показывает, что возможно прогнозирование сердечно-сосудистых заболеваний с использованием аппарата нейронных сетей с высокой эффективностью прогнозирования.

Для оценки качества бинарной классификации используется ROC-кривая (receiver operating characteristic, рабочая характеристика приемника). Количественную интерпретацию ROC дает показатель AUC (area under ROC curve, площадь под ROC-кривой) – площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций (см. пример на Рис. 2). Значение $AUC = 0.7$ можно интерпретировать, как вероятность того, что модели удастся разделить классы с вероятностью 70%.

Если:

- $0,8 < AUC < 1,0$ – модель работает превосходно;
- $0,6 < AUC < 0,8$ – модель работает хорошо;
- $0,5 < AUC < 0,6$ – модель работает удовлетворительно;
- $AUC < 0,5$ – модель не работает.

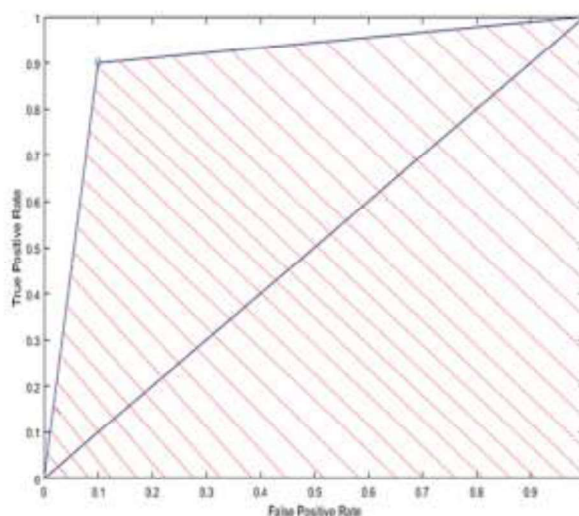


Рис. 2. Пример ROC-кривой.

На рис. 3 представлены характеристики одно(двух)параметрической диагностики ишемической болезни сердца (ИБС) с использованием различных показателей из набора данных Cleveland Heart Disease (см. выше). Здесь координаты точек X, Y имеют значения (1–чувствительность диагноза), (специфичность диагноза), соответственно. Также, здесь представлены соответствующие значения AUC. Эти характеристики даны также при диагностике ИБС с помощью нейронной сети (NN).

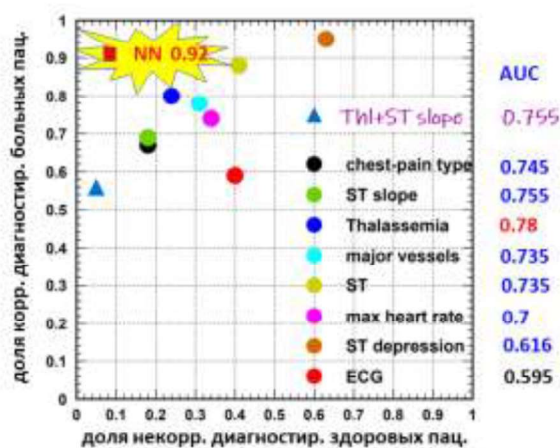


Рис. 3. Оценка качества медицинских показателей

Из Рис. 3 видно, что использование совокупности медицинских показателей, как это делается при применении нейронной сети, позволяет улучшить качество диагноза.

Действительно, использование аппарата нейронных сетей позволяет учитывать всю совокупность медицинских показателей пациента, их корреляции, производить поиск скрытых зависимостей в наборах показателей пациентов. Как результат, качество «NN-диагноза» ($AUC = 0,92$) значительно превышает качество «однопараметрического диагноза» ($AUC < 0,78$).

Данное исследование показывает, что подобный подход к прогнозированию заболеваний может оказаться востребованным, поскольку обученная нейронная сеть может помочь медицинским работникам более точно диагностировать сердечно-сосудистые заболевания.

Библиографический список

1. Documentation of Matlab – Deep Learning Training Options – URL: <https://docs.exponenta.ru/deeplearning/ref/trainingoptions.html> (дата обращения: 15.03.2021).
2. Documentation of Matlab – Matrix Normalization with MapMinMax – URL: <https://www.mathworks.com/help/deeplearning/ref/mapminmax.html.html> (дата обращения: 18.03.2021).
3. Documentation of Matlab – Sequence Classification Using Deep Learning – URL: <https://docs.exponenta.ru/deeplearning/ug/classify-sequence-data-using-lstm-networks.html> (дата обращения: 14.03.2021).
4. Documentation of Matlab – Softmax layer – URL: <https://docs.exponenta.ru/deeplearning/ref/nnet.cnn.layer.softmaxlayer.html> (дата обращения: 20.03.2021).
5. Heart Disease Prediction. Cleveland Heart Disease(UCI Repository) dataset — classification with various models. — URL: <https://towardsdatascience.com/heart-disease-prediction-73468d630cfc> (дата обращения: 13.03.2021).
6. Heart Disease UCI Dataset Download – URL: <https://www.kaggle.com/ronitf/heart-disease-uci> (дата обращения: 20.01.2021).

УДК 81-139

Курбатов С.Д.

ВЛИЯНИЕ СОВРЕМЕННЫХ РЕАЛИЙ НА ЛЕКСИЧЕСКИЙ СОСТАВ АНГЛИЙСКОГО ЯЗЫКА

THE INFLUENCE OF MODERN REALITIES ON THE LEXICAL COMPOSITION OF THE ENGLISH LANGUAGE

*Филиал «Протвино» государственного университета «Дубна»
Секция «Социальные и гуманитарные науки»*

Автор: Курбатов Сергей Дмитриевич, студент 2 курса направления «Автоматизация технологических процессов и производств» филиала «Протвино» государственного университета «Дубна».

Научный руководитель: Сотникова Инесса Юрьевна, старший преподаватель кафедры общеобразовательных дисциплин филиала «Протвино» государственного университета «Дубна».

Author: 2nd year student of the direction «Automation of technological processes and production» of the branch «Protvino» state University «Dubna».

Scientific adviser: Sotnikova Inessa Yuryevna, senior tutor of general educational subjects department of the branch «Protvino» state University «Dubna».