Научная статья
УДК 004.032.26+004.056.53+004.418
ББК 32.813.5
Ч 25
DOI: 10.53598/2410-3225-2023-1-316-70-79

Нейросетевая система биометрической идентификации личности по голосу

(*Рецензирована*)

Вера Аркадьевна Частикова[1], Сергей Анатольевич Жерлицын[2], Диана Олеговна Войлова[3]

[1–3] *Кубанский государственный технологический университет, Краснодар, Россия*
[1] *chastikova_va@mail.ru*
[2] *kpytooooo@gmail.com*
[3] *diana.voylova@mail.ru*

Аннотация. *Рассматривается подход к реализации системы идентификации личности по одной из биометрических характеристик – голосу. Описываются основные факторы, приведшие к расширению области применения средств биометрической идентификации. Приводятся базовые характеристики голоса, позволяющие использовать его в качестве фактора идентификации. Раскрываются подробности предметной области проводимого исследования: перечисляются виды систем идентификации личности по голосу, основные этапы и варианты их работы. Проводится обоснование актуальности и постановка задачи: разработка и сравнение двух подходов биометрической идентификации личности по голосу на основе разных архитектур нейронных сетей – многослойного персептрона и сверточной нейронной сети (convolutional neural network, CNN). Разработанные системы включают в себя этап предварительной обработки сигнала: выделения значимых голосовых характеристик, удаление шумов, пауз, тишины, а также формирование спектрограмм в графическом виде. Приводятся алгоритмы работы модулей идентификации на основе обеих рассматриваемых архитектур нейронных сетей, описываются механизмы работы каждой из них и детальное описание конфигураций. По итогам проведенных экспериментов были сделаны выводы об эффективности применения данных архитектур: сверточная нейронная сеть показала 98,7% точность распознавания и обошла многослойный персептрон на 5,6% при обучении на той же выборке.*

Ключевые слова*: идентификация личности, биометрия, голос, нейронные сети, искусственный интеллект*

Original Research Paper

Neural network system for biometric identification of a person by voice

Vera A. Chastikova[1], Sergey A. Zherlitsyn, Diana O. Voylova[3]

[1–3] *Kuban State University of Technology, Krasnodar, Russia*
[1] *chastikova_va@mail.ru*
[2] *kpytooooo@gmail.com*
[3] *diana.voylova@mail.ru*

Abstract. *The article considers an approach to the implementation of a system of identification of a person by one of the biometric characteristics – voice. The main factors that led to the expansion of the scope of biometric identification tools are described. The basic characteristics of the voice are given, allowing it to be used as an identification factor. The details of the subject area of the research are revealed: the types of voice identification systems, the main stages and variants of their work are listed. The substantiation of the relevance and formulation of the problem is carried out: the develop-*

ment and comparison of two approaches to biometric identification of a person by voice based on different neural network architectures: a multilayer perceptron and a convolutional neural network. The developed systems include a stage of signal preprocessing: the allocation of significant voice characteristics, the removal of noise, pauses, silence, as well as the formation of spectrograms in graphical form. Algorithms for the operation of identification modules based on both considered neural network architectures are given; the mechanisms of operation of each of them and a detailed description of configurations are described. Based on the results of the experiments, conclusions were drawn about the effectiveness of using these architectures: the convolutional neural network showed 98.7% *recognition accuracy and bypassed the multilayer perceptron by* 5.6% *when training on the same sample.*

Keywords: *personal identification, biometrics, voice, neural networks, artificial intelligence*

1. Введение

Начиная с 2015 года, биометрическая идентификация охватила почти все сферы человеческой деятельности.

Первый массовый приход биометрии запустила 10 сентября 2013 года компания Apple, представив публике встроенный в iPhone 5s считыватель отпечатков пальцев – Touch ID.

Вторая веха – биометрические паспорта. В России начали выдавать паспорта нового поколения, содержащие электронный носитель информации – бесконтактный чип. Данные на чипе Российского паспорта содержат: фотографию владельца паспорта, отпечатки пальцев, информацию о дате и месте рождения владельца, дате выдачи паспорта и органе, выдавшем документ [1].

Преимущества биометрии уже привели к широкому распространению сенсоров отпечатков пальцев в мобильных устройствах, таких как смартфоны и планшеты. Но типов биометрических технологий гораздо больше, в ближайшем будущем они получат самое широкое распространение [2].

Метод идентификации по голосу человека основан на анализе уникальных характеристик речи, обусловленных анатомическими особенностями, такими как размер и форма горла и рта, строение голосовых связок, и приобретенными привычками, такими как громкость, акцент, скорость речи. Основными причинами внедрения систем, основанных на биометрической идентификации по голосу, являются практика встраивания микрофонов в компьютерные системы и периферийные устройства, а также повсеместное распространение телефонных сетей. В практической деятельности ни одна из систем идентификации, в том числе и голосовая, не гарантируют стопроцентную идентификацию личности.

2. Постановка задачи

Идентификация по голосу представляет собой процесс определения личности по образцу голоса путем сравнения образца с шаблонами, которые хранятся в базе. Системы распознавания могут быть разделены на текстонезависимые, в которых изначально не известен текст, произносимый пользователем, и текстозависимые, в которых, соответственно, данная информация известна. При текстозависимом распознавании могут использоваться как фиксированные фразы, так и фразы, сгенерированные системой и предложенные пользователю. Текстонезависимые системы предназначены обрабатывать произвольную речь.

Работа систем распознавания содержит два основных этапа: регистрация пользователей в системе и сам процесс распознавания и идентификации личности [3]. Регистрация пользователей проходит путем записи голоса пользователей. Образец голоса обрабатывается с целью извлечения признаков, которые могут быть использованы для распознавания. На основе извлеченных признаков строятся шаблоны голосов. Шаблон представляет собой некую структуру, позволяющую при данных признаках оценить

степень подобия либо сразу принять решение. Во время процесса идентификации происходит извлечение из предъявленного образца признаков, которые затем сравниваются с шаблонами зарегистрированных в системе пользователей.

В настоящее время все большую популярность набирают методы биометрической идентификации личности на основе глубоких нейронных сетей [4], так как они дают большую точность и скорость решения задач распознавания образов, идентификации и прогнозирования по сравнению с другими методами, используют небольшой объем памяти для функционирования системы, имеют сравнительно высокую производительность обучения и являются основой для более современных методов.

Для разработки модуля идентификации личности по голосу были выбраны две архитектуры нейронных сетей:

– многослойный персептрон, так как данная нейросеть является самой простой в реализации и имеет высокую скорость обучения;

– сверточная нейронная сеть, так как эта нейросеть является одной из популярных нейросетей в настоящее время и дает высокую точность распознавания и прогнозирования на относительно небольших наборах данных [5].

3. Разработка системы идентификации личности по голосу

В ходе исследования было реализовано два программных модуля, выполняющих одну и ту же задачу, но на двух различных нейросетевых архитектурах, таких как многослойный персептрон (MLP) и сверточная нейронная сеть (CNN). Основной целью данной работы является сравнение двух реализованных модулей между собой.

Методика идентификации личности по голосу работает в двух режимах: в режиме регистрации, то есть записи шаблонов голосов пользователей, обучении и занесения полученных шаблонов в базу данных (датасет), и режиме идентификации.

В подмодуле предварительной обработки акустический компонент преобразует сигнал в цифровую форму с параметром частоты дискретизации равным 22050 Гц. Выделение признаков речевого сигнала проводится в несколько этапов. Первым шагом являются разбиение всего сигнала на короткие фреймы (кадры) длиной 10 мс. Первичная предобработка данных включает в себя, помимо разметки, удаление пауз, шумов и тишины с помощью сторонней модели Voice Activity Detection. Это необходимо для того, чтобы записи, полученные в разное время, соответствовали друг другу независимо от сторонних факторов. Существует множество способов, при помощи которых можно уменьшить шумовые эффекты [6].

В данной методике для удаления шумов и тишины используется окно Хемминга, которое определяется следующей формулой:

$$\omega(n) = 0{,}53836 - 0{,}36164\cos\left(\frac{2\pi n}{N-1}\right), \quad (1)$$

где n – порядковый номер элемента в кадре; N – длина кадра.

Для преобразования полученных кадров в числовые характеристики в данной методике используются следующие функции выделения признаков речевого сигнала:

– mfcc, функция, предназначенная для преобразования сигналов временной области в сигнал частотной области;

– chroma, функция вычисления хромаграммы по форме волны;

– mel, функция вычисления спектограммы в масштабе mel (шкала Гц разбивается на ячейки, и каждая ячейка преобразовывается в соответствующую ячейку в шкале mel);

– contrast, функция вычисления спектрального анализа;

– tonnetz, функция вычисления средних значений тональных характеристик.

К основным характеристикам голоса, по которым производятся вычисления вы-

шеперечисленных функций, относятся: тембр, высота, тон, темп (скорость), акцент и громкость речи [7].

Каждый mfcc представляет собой массив длиной 40, chroma – 12, mel – 128, contrast – 7 и tonnetz – 6. В целом у нас получается 193 значения числовых характеристик для каждого голосового шаблона. Далее все эти числовые значения объединяются для каждого файла, чтобы в итоге получился один массив из 193 чисел для каждого файла.

Алгоритм работы программного модуля, реализованного на архитектуре MLP, приведен на рисунке 1.

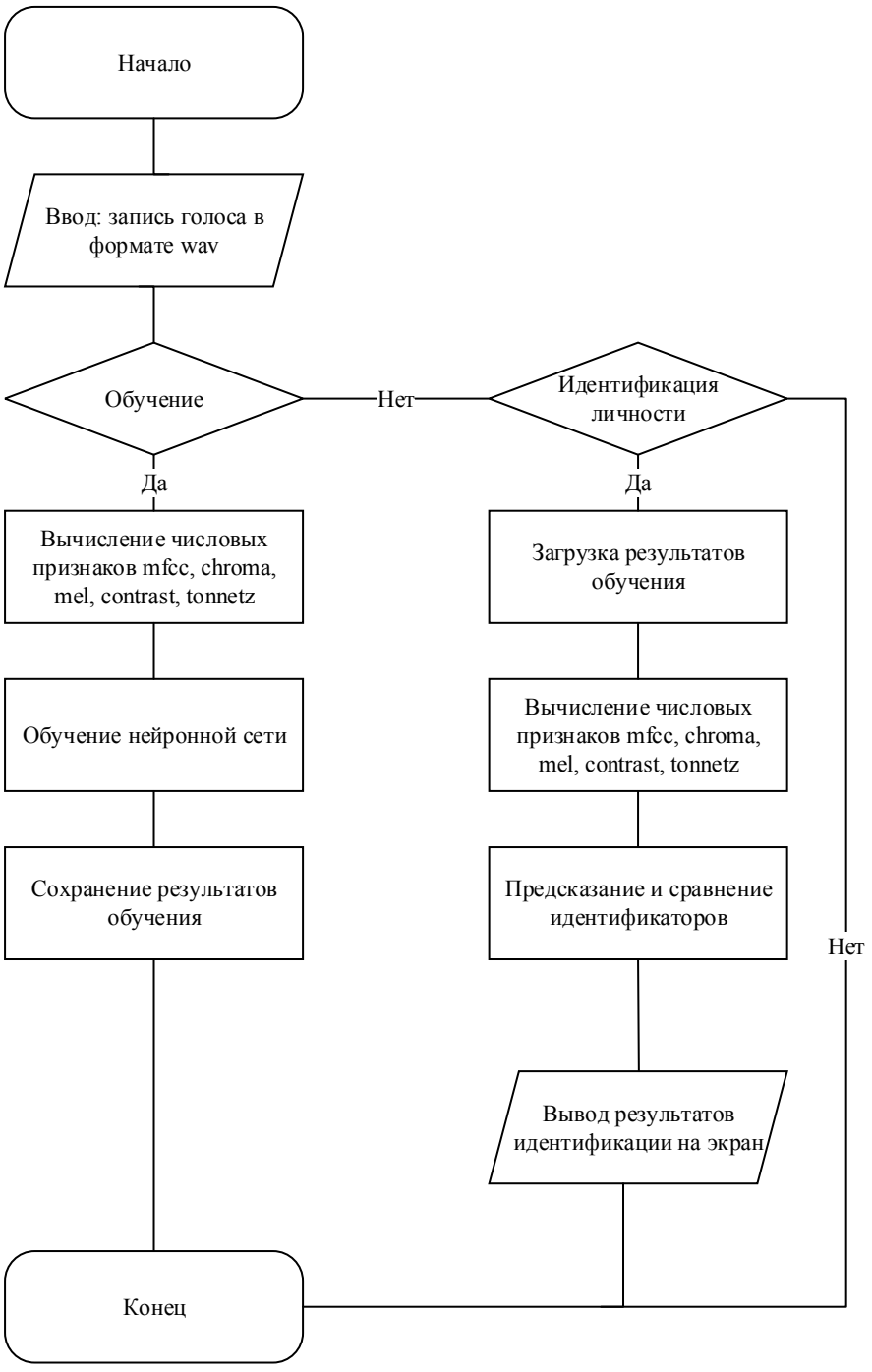

Рис. 1. Алгоритм работы методики голосовой идентификации на архитектуре MLP

Fig. 1. The operation algorithm of the voice identification technique on the MLP architecture

В программном модуле, реализованном на архитектуре многослойного персептрона, имеющем два скрытых слоя relu и softmax, в процессе обучения в нейронную сеть входные данные подаются не в виде исходного сигнала, а в виде массива числовых признаков, которые вычисляются в подмодуле предварительной обработки. Представление сигнала в виде числовых характеристик дает существенный прирост в качестве и скорости обучения нейронной сети. Закончив обучение, нейронная сеть сохраняет свой процесс и при необходимости переходит в режим идентификации личности.

В программном модуле, реализованном на архитектуре сверточной нейронной сети, используется дополнительная функция для создания изображений со спектрограммами для каждого аудиофайла. Данная операция выполняется после вычисления числовых параметров в подмодуле предварительной обработки. Спектрограмма, представленная на рисунке 2, дает большой объем информации, в том числе о характеристиках личности говорящего, и динамически показывает характеристики изменения спектра сигнала [8]. Хотя речь представляет собой изменяющийся во времени сигнал со сложными корреляциями в различных временных масштабах, спектрограмма обеспечивает хорошее обучение нейронной сети и дает высокие значения точности в процессе идентификации личности.

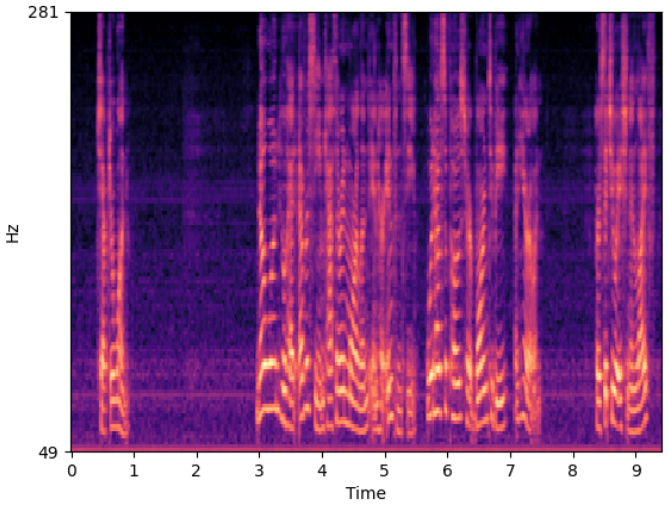

Рис. 2. Пример изображения спектограммы аудиофайла

Fig. 2. An example of an audio file spectrogram image

Сверточная нейронная сеть, используемая в реализации программного модуля, имеет вход Conv2D с MaxPooling2D и пять скрытых слоев: три Conv2D с их соответствующими MaxPooling2D, слой fully connect и выходной слой output [9, 10]. На вход модуля для обучения подается не сам аудиофайл, а массив изображений, полученный в ходе преобразования аудиофайла в спектрограммы. Процесс работы модуля, реализованного на CNN, представлен на рисунке 3.

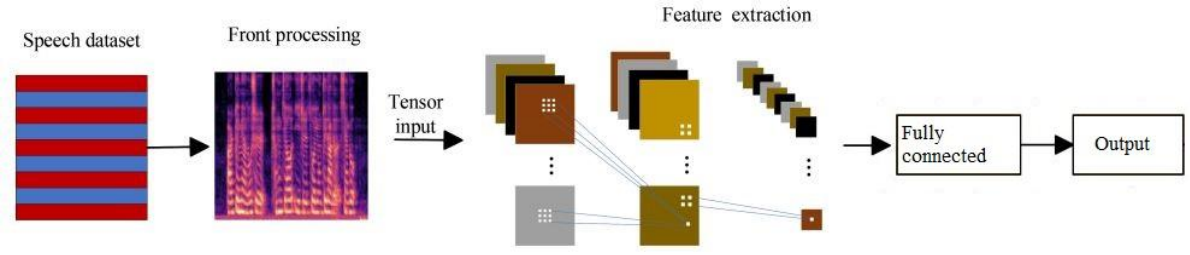

Рис. 3. Процесс работы модуля, реализованного на CNN

Fig. 3. The workflow of the module implemented on CNN

При идентификации личности на вход модели поступает ранее записанный аудиофайл. Стоит отметить, что звук, записанный с микрофона, не поступает сразу же в модуль распознавания личности, а предварительно сохраняется в отдельный аудиофайл в соответствующем формате для дальнейшего распознавания. Далее происходит загрузка ранее полученных результатов обучения. Последним шагом в идентификации личности является процесс предсказания идентификатора и вывод ФИО пользователя на экран при положительном исходе работы модуля или сообщения «Личность не идентифицирована!» при отрицательном результате.

4. Результаты тестирования

Для проведения экспериментов по качеству и точности идентификации личности были записаны 7 шаблонов голоса одного человека для обучения нейронной сети в идеальных условиях, то есть при отсутствии различного рода помех, неисправностей записывающей системы и искажений голоса. Для тестирования были записаны 7 образцов голоса в различных окружающих условиях и состояниях здоровья человека. Результат идентификации личности в ходе проведения экспериментов на двух различных модулях представлен в таблице 1.

Таблица 1

Результат идентификации личности в ходе проведения экспериментов

Table 1. The result of personal identification in the course of experiments

№ п/п	Условия тестирования	MLP	CNN
1	Здоровье человека в норме, помехи окружающей среды отсутствуют	+	+
2	Здоровье человека в норме, эксперимент проводился в условиях уличного шума	+	+
3	Здоровье человека в норме, эксперимент проводился около оживленной проезжей дороги	-	+
4	Здоровье человека в норме, эксперимент проводился вблизи шумной стройки	-	-
5	В голосе присутствует хрипота из-за болезни человека, помехи окружающей среды отсутствуют	+	+
6	Осипший голос из-за болезни, эксперимент проводился около оживленной проезжей дороги	-	-
7	Незначительная картавость речи, помехи окружающей среды отсутствуют	+	+

В таблице 1 используются следующие обозначения: «+» – положительный результат идентификации; «-» – отрицательный результат идентификации.

Для анализа эффективности работы реализованных модулей идентификации личности по голосу необходимо сравнить их между собой по трем основным параметрам:

- точность идентификации личности;
- скорость идентификации;
- вычислительная нагрузка на аппаратуру.

Запуски программных кодов производились на одинаковом наборе данных. Для наглядности результаты представлены в виде графиков и диаграмм, в которых номер нейросетевых архитектур расположен по горизонтали, а результат сравнения – по вертикали. Результат представлен в десятибалльной системе, где 10 – худший показатель, а 0 – лучший.

Точность идентификации личности по голосу является наиболее важным показателем, так как от этого параметра зависит целесообразность применения методики.

Показатели точности идентификации личности по голосу в модуле, реализованном на архитектуре MLP, составляют 93,1%. График точности представлен на рисунке 4.

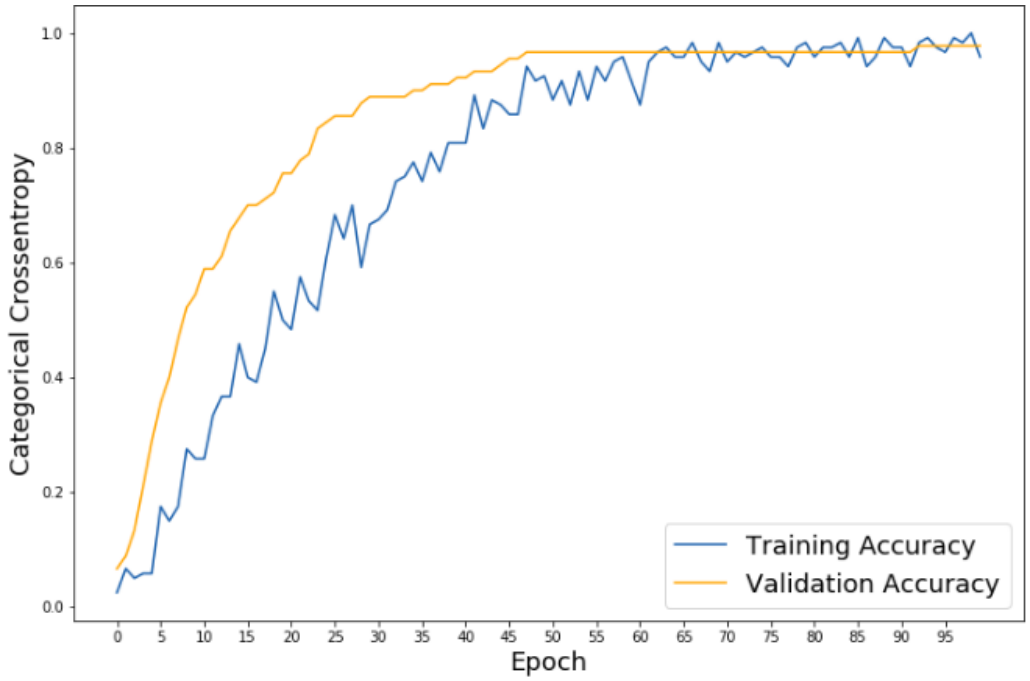

Рис. 4. Точность идентификации личности методики MLP

Fig. 4. Personal identification accuracy of the MLP technique

Показатели точности идентификации личности по голосу в модуле, реализованном на архитектуре CNN, составляют 98,7%. График точности представлен на рисунке 5.

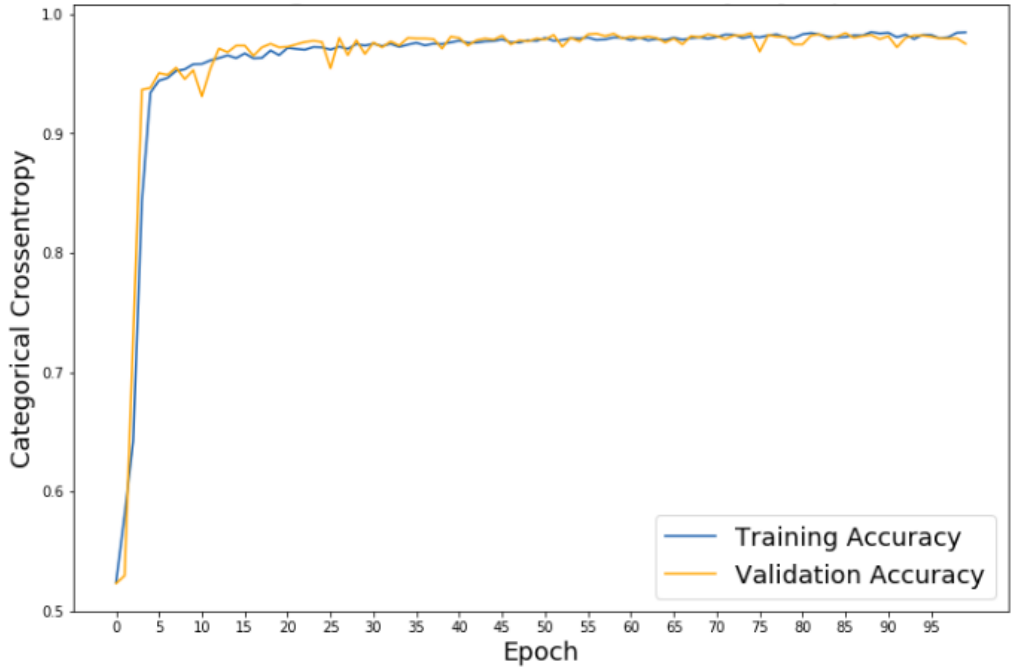

Рис. 5. Точность идентификации личности методики CNN

Fig. 5. Personal identification accuracy of the CNN technique

Скорость идентификации личности показывает, какое минимальное количество времени потребуется для процесса распознавания личности по голосу. Диаграмма скорости идентификации личности реализованных методик представлена на рисунке 6.

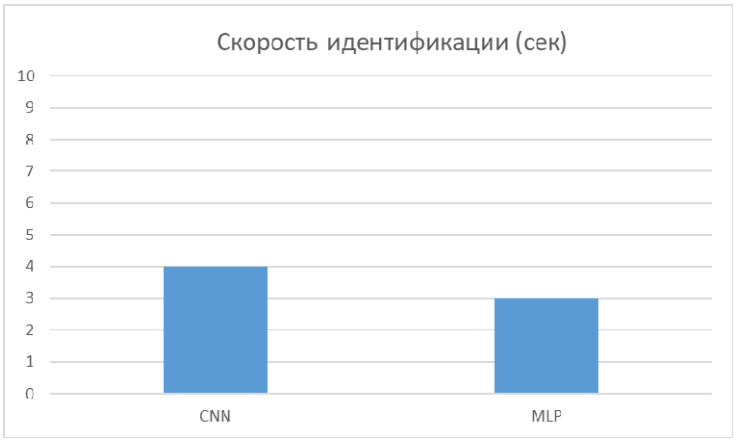

Рис. 6. Сравнительный анализ скорости идентификации

Fig. 6. Comparative analysis of identification speed

Сравнивая результаты работ методик по влиянию на вычислительные мощности, было обнаружено, что методика на основе MLP меньше всего потребляет системных ресурсов. Результат сравнения представлен на рисунке 7.

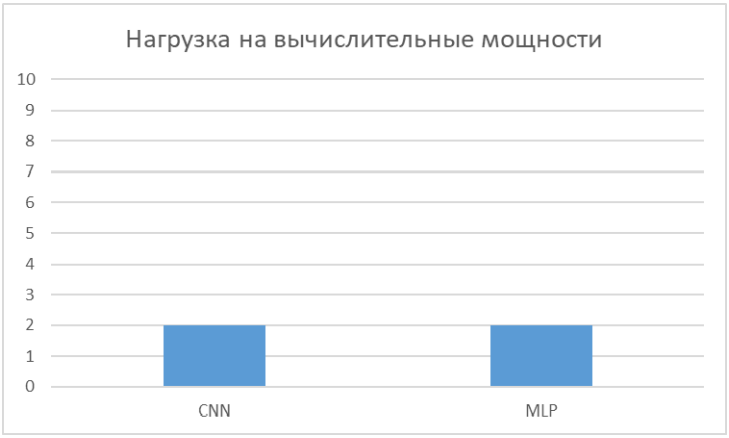

Рис. 7. Сравнительный анализ нагрузки на системные ресурсы

Fig. 7. Comparative analysis of the load on system resources

В результате проделанной работы можно сделать выводы, что использование глубокого нейронного многослойного персептрона позволяет сократить время выполнения процедуры идентификации. Это связано с тем, что после вычисления числовых параметров аудиозаписи массив входных данных сразу передается на обучение нейронной сети. В то время как в модуле, реализованном на архитектуре CNN, после вычисления массива числовых параметров происходит представление данного массива значений в массив изображений, то есть выполняется дополнительная функция преобразования входного сигнала. В свою очередь модуль, реализованный на архитектуре CNN, уже на 10 эпохе дает практически стопроцентную вероятность точной идентификации, в то время как MLP достигает этих значений только к 50 эпохе. Сравнивая результаты работы модулей по влиянию на вычислительную мощность системы, было выявлено, что они потребляют одинаково малое количество системных ресурсов.

5. Заключение

Запуски программных кодов производились на одинаковом наборе данных. Точность идентификации личности по голосу является наиболее важным показателем, так как от этого параметра зависит целесообразность применения методики [11].

Показатели точности идентификации личности по голосу в модуле, реализованном на архитектуре MLP, составляют 93,1%.

Показатели точности идентификации личности по голосу в модуле, реализованном на архитектуре CNN, составляют 98,7%.

В результате проделанной работы можно сделать вывод, что сверточные нейронные сети позволяют достичь более высоких результатов по точности идентификации личности при равных прочих аспектах. Однако в модуле, реализованном на архитектуре CNN, после вычисления массива числовых параметров происходят представления данного массива значений в массив изображений, то есть выполняется дополнительная функция преобразования входного сигнала, что делает обучение модели и последующее применение для идентификации личности более ресурсоемким.

Примечания

1. Биометрия от «А» до «Я» полное руководство биометрической идентификации и аутентификации // Интемс. URL: https://securityrussia.com/blog/biometriya.html (дата обращения: 01.04.2022).

2. Chastikova V.A., Zherlitsyn S.A., Volya Yu.I. Development of a personal identification technique for automation systems // Iop Conference Series: Materials Science and Engineering. 2021. Vol. 1047, No. 3. P. 12138.

3. Первушин Е.А. Обзор основных методов распознавания дикторов // Математические структуры и моделирование. 2011. № 24. С. 41–54.

4. Частикова В.А., Жерлицын С.А., Воля Ю.И. Нейросетевой подход к решению задачи построения фоторобота по словесному описанию // Известия Волгоградского государственного технического университета. 2018. № 8. С. 63-67.

5. Analysis of training of deep neural networks with heterogeneous architecture while detecting malicious network traffic / V.A. Chastikova, S.A. Zherlitsyn, Yu.I Volya., V.V. Sotnicov // IOP Conference Series: materials Science and Engineering. 2021. Vol. 1047. P. 12135.

6. Bosi M., Goldberg R.E. Deconvolutive Short-Time Fourier Transform Spectrogram. Springer, 2010. 434 p.

7. Частикова В.А., Титова А.А., Войлова Д.А. Аналитический обзор методов идентификации личности на основе биометрических характеристик // Вестник Адыгейского государственного университета. Сер.: Естественно-математические и технические науки. 2022. Вып. 1 (296). С. 92–112. URL: http://vestnik.adygnet.ru

8. Lu W., Zhang Q. New directions in cryptography // IEEE Signal Processing Letters. 2009. Vol. 16, No. 7. P. 576–579.

9. Chastikova V.A., Sotnicov V.V. Method of analyzing computer traffic based on recurrent neural networks // Journal of Physics: Conference Series: International Conference “High-Tech and Innovations in Research and Manufacturing”, HIRM 2019. 2019. P. 012133.

10. Нейросетевая технология обнаружения аномального сетевого трафика / В.А. Частикова, С.А. Жерлицын, Ю.И. Воля, В.В. Сотников // Прикаспийский журнал: управление и высокие технологии. 2020. № 1. С. 20–32.

11. Сравнительный анализ некоторых алгоритмов роевого интеллекта при обнаружении сетевых атак нейросетевыми методами / В.А. Частикова, М.П. Малыхина, С.А. Жерлицын, Ю.И. Воля // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2017. № 129. С. 106–115.

References

1. Biometrics from “A” to “Z” a complete guide to biometric identification and authentication. // Intems. URL: https://securityrussia.com/blog/biometriya.html (access date: 01/04/2022).

2. Chastikova V.A., Zherlitsyn S.A., Volya Yu.I. Development of a personal identification technique for automation systems // Iop Conference Series: Materials Science and Engineering. 2021. Vol. 1047, No. 3. P. 12138.

3. Pervushin E.A. Overview of the main methods of speaker recognition // Mathematical Structures and Modelling. 2011. No. 24. P. 41–54.

4. Chastikova V.A., Zherlitsyn S.A., Volya Yu.I. Neural network approach to the identikit

compilationproblem through verbal description // News of the Volgograd State Technical University 2018. No. 8. P. 63–67.

5. Analysis of training of deep neural networks with heterogeneous architecture while detecting malicious network traffic / V.A. Chastikova, S.A. Zherlitsyn, Yu.I Volya., V.V. Sotnicov // IOP Conference Series: materials Science and Engineering. 2021. Vol. 1047. P. 12135.

6. Bosi M., Goldberg R.E. Deconvolutive Short-Time Fourier Transform Spectrogram. Springer, 2010. 434 p.

7. Chastikova V.A., Titova A.A., Voylova D.O. Analytical review of personal identification methods based on biometric characteristics // The Bulletin of the Adyghe State University. Ser.: Natural-Mathematical and Technical Sciences. 2022. Iss. 1 (296). P. 92–112. URL: http://vestnik.adygnet.ru

8. Lu W., Zhang Q. New directions in cryptography // IEEE Signal Processing Letters. 2009. Vol. 16, No. 7. P. 576–579.

9. Chastikova V.A., Sotnicov V.V. Method of analyzing computer traffic based on recurrent neural networks // Journal of Physics: Conference Series: International Conference “High-Tech and Innovations in Research and Manufacturing”, HIRM 2019. 2019. P. 012133.

10. Neural network technology for detecting anomalous network traffic / V.A. Chastikova, S.A. Zherlitsyn, Yu.I. Volya, V.V. Sotnikov // Caspian Magazine: Management and High Technologies. 2020. No. 1. P. 20–32.

11. Comparative analysis of some swarm intelligence algorithms with detection of network attacks using neural network methods / V.A. Chastikova, M.P. Malykhina, S.A. Zherlitsyn, Yu.I. Volya // Polythematic Network Electronic Scientific Magazine of the Kuban State Agrarian University. 2017. No. 129. P. 106–115.

Авторы заявляют об отсутствии конфликта интересов.

The authors declare no conflicts of interests.

Статья поступила в редакцию 21.12.2022; *одобрена после рецензирования 20.01.2023*; *принята к публикации 21.01.2023*.

The article was submitted 21.12.2022; *approved after reviewing 20.01.2023*; *accepted for publication 21.01.2023*.