ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ, РЕЧИ, ТЕКСТА И РАСПОЗНАВАНИЕ ОБРАЗОВ

SIGNAL, IMAGE, SPEECH, TEXT PROCESSING AND PATTERN RECOGNITION

УДК 004.81
https://doi.org/10.37661/1816-0301-2022-19-4-53-68

Оригинальная статья
Original Paper

Разработка алгоритма распознавания эмоций человека с использованием сверточной нейронной сети на основе аудиоданных

В. В. Семенюк, М. В. Складчиков✉

Донецкий техникум промышленной автоматики имени А. В. Захарченко,
ул. Горького, 163, Донецк, 83000, Украина
✉*E-mail: maxsklad19981@yandex.ru*

Аннотация

Цели. Приведено описание и рассмотрен опыт создания алгоритма распознавания эмоционального состояния субъекта.

Методы. Использованы методы обработки изображений.

Результаты. Предложенный алгоритм позволяет распознавать эмоциональные состояния субъекта на основании звукового набора данных. Благодаря проведенному исследованию удалось улучшить точность работы алгоритма путем изменения подаваемого на вход нейронной сети набора данных.

Описаны этапы обучения сверточной нейронной сети на заранее заготовленном наборе звуковых данных, а также структура алгоритма. Для валидации нейронной сети был отобран иной, не участвующий в тренировке, набор аудиоданных. В результате проведения исследования построены графики, демонстрирующие точность работы предлагаемого метода.

После получения первоначальных данных сделан анализ возможностей улучшения алгоритма с точки зрения эргономики и точности его работы. Разработана стратегия, позволяющая добиться лучшего результата и получить более точный алгоритм. На основании заключений, изложенных в статье, приводится обоснование выбора представления набора данных и программного комплекса, необходимого для реализации программной части алгоритма.

Заключение. Предложенный алгоритм обладает высокой точностью и не требует больших вычислительных затрат.

Ключевые слова: нейронная сеть, распознавание эмоций человека, сверточная нейронная сеть, дактилоскопия звука, программная библиотека TensorFlow, нейросетевая библиотека Keras, пакет программ Matlab

Для цитирования. Семенюк, В. В. Разработка алгоритма распознавания эмоций человека с использованием сверточной нейронной сети на основе аудиоданных / В. В. Семенюк, М. В. Складчиков // Информатика. – 2022. − Т. 19, № 4. – С. 53–68. https://doi.org/10.37661/1816-0301-2022-19-4-53-68

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 08.08.2022
Подписана в печать | Accepted 08.09.2022
Опубликована | Published 29.12.2022

Algorithm development for recognizing human emotions using a convolutional neural network based on audio data

Viktoriya V. Semenuk, Maxim V. Skladchikov✉

Donetsk Technical School of Industrial Automation
after A. V. Zakharchenko,
st. Gorkogo, 163, Donetsk, 83000, Ukraine
✉*E-mail: maxsklad19981@yandex.ru*

Abstract

Objectives. This article provides a description and experience of creating the algorithm for recognizing the emotional state of the subject.

Methods. Image processing methods are used.

Results. The proposed algorithm makes it possible to recognize the emotional states of the subject on the basis of an audio data set. It was possible to improve the accuracy of the algorithm by changing the data set supplied to the input of the neural network.

The stages of training convolutional neural network on a pre-prepared set of audio data are described, and the structure of the algorithm is described. To validate the neural network different set of audio data, not participating in the training, was selected. As a result of the study, graphs were constructed demonstrating the accuracy of the proposed method.

After receiving the initial data of the study, the analysis of the possibilities for improving the algorithm in terms of ergonomics and accuracy of operation was also carried out. The strategy was developed to achieve a better result and obtain a more accurate algorithm. Based on the conclusions presented in the article, the rationale for choosing the representation of the data set and the software package necessary for the implementation of the software part of the algorithm is given.

Conclusion. The proposed algorithm has a high accuracy of operation and does not require large computational costs.

Keywords: neural network, human emotion recognition, convolutional neural network, sound fingerprinting, TensorFlow software library, Keras neural network library, Matlab software package

For citation. Semenuk V. V., Skladchikov M. V. *Algorithm development for recognizing human emotions using a convolutional neural network based on audio data*. Informatika [*Informatics*], 2022, vol. 19, no. 4, pp. 53−68 (In Russ.). https://doi.org/10.37661/1816-0301-2022-19-4-53-68

Conflict of interest. The authors declare of no conflict of interest.

Введение. Новизна предлагаемого метода заключается в высокой точности работы описываемого алгоритма по сравнению с имеющимися алгоритмами идентификации эмоций. Для достижения поставленной цели в качестве архитектуры была выбрана сверточная нейронная сеть. Использование разработанной структуры нейронной сети обусловлено высокой точностью и простотой распознавания изображений. Следует отметить, что большинство алгоритмов, идентифицирующих эмоции, формируют результат на основании видеоданных. Такой подход в первую очередь требует высококачественного регистрирующего оборудования. Кроме того, для работы необходимо производить сложные вычисления для регистрации активности лицевых мышц.

Мотивацией к разработке предложенного алгоритма и архитектуры нейронной сети стали исследование системы голосового управления и в целом классификация различных систем [1]. Как утверждают авторы доклада, точность распознавания звуковых данных при создании идеальных условий составила около 92 %.

Сверточные нейронные сети зачастую применяются для анализа изображений. Поэтому записанные данные звука преобразовывались в изображение с использованием технологии «дактилоскопия звука» [2]. Благодаря этому удалось снизить временные затраты, необходимые для обработки входных данных, что в свою очередь повысило эргономические свойства предлагаемого метода.

Каждый человек выражает эмоции при возникновении определенных внешних или внутренних возбудителей. Ввиду индивидуальности каждого субъекта, а также его психологического состояния, которое в разные периоды жизни может меняться, довольно сложно выделить единый способ оценки эмоций [3–7]. Это приводит к появлению большого количества подходов к их идентификации и классификации [8–24].

В современном мире существует большое количество алгоритмов и систем, способных анализировать эмоциональный окрас человека [13, 15]. В век цифровизации и развития искусственного интеллекта в когнитивистике наиболее актуальным подходом к анализу эмоций человека является технология SER (Speech Emotion Recognition). Этот подход основывается на обработке и анализе звуковых сигналов. Голос как набор данных для идентификации отдельных типов эмоций является наиболее информативным, что позволяет качественно классифицировать эмоции человека по сравнению с другими подходами к оценке эмоционального окраса человека.

Проанализировав существующие решения в области распознавания эмоционального состояния человека по голосу, можно выделить ряд проблем [9, 10, 12–15, 17, 18, 20, 21, 24]:

1. Жесткая взаимосвязь точности и количества идентифицируемых эмоций снижает сферу применения алгоритмов.

2. Для увеличения точности работы может использоваться более громоздкий математический аппарат, который не всегда обоснован с точки зрения эргономики работы алгоритма.

3. Для сокращения затрачиваемых ресурсов анализируется малый участок идентифицируемого набора данных, на основании которого делается вывод об адекватности эмоционального состояния. Это в свою очередь увеличивает риск ошибочного заключения на ином участке распознавания.

4. Использование исключительно статистических методов машинного обучения для распознавания эмоций имеет ряд существенных недостатков и не позволяет с необходимой точностью обеспечивать управление информационными потоками, развитие, перестроение и увеличение набора данных. Также подобные системы тяжело синтезировать и обеспечивать их взаимодействие на различных уровнях между собой.

Сказанное выше обусловливает цель исследования, которая заключается в построении системы, удовлетворяющей решению выявленных проблем.

Методики оценки эмоционального состояния. Современные модели оценки эмоционального состояния человека можно разделить на два основных вида: дискретные и многомерные.

Дискретная модель подразумевает упрощенную оценку эмоций человека. При использовании такой модели результат оценки эмоционального состояния субъекта основывается лишь на базовых (первичных) факторах. Это приводит к уменьшению точности работы системы, что сильно сказывается на объективности вывода алгоритма.

Многомерная модель предполагает более глубокий анализ эмоционального состояния субъекта. С ее помощью оцениваются не только базовые параметры, влияющие на состояние человека, но и косвенные (пульс, изменение цвета лица и т. д.). Многомерная модель позволяет приближенно имитировать работу мозга при оценке эмоционального состояния человека. Применение такой модели дает возможность получить более точный результат, однако при этом возрастают сложность и громоздкость разрабатываемого алгоритма.

Вариантом использования описанных моделей являются аффективные вычисления. Данная область знаний базируется на множестве научных дисциплин (информатике, когнитивистике,

психологии и т. д.). Ее основная задача – анализ и интерпретация эмоций человека. В зависимости от сферы применения аффективные вычисления могут использоваться для идентификации эмоций человека (методики распознавания эмоций), а также для создания или симулирования его эмоциональных состояний (робототехника). Для реализации систем первого типа имеются различные человекомашинные интерфейсы. При идентификации эмоций в «статике» можно использовать изображение. Такие системы просты в построении и имеют достаточную точность, необходимую для получения качественного результата работы алгоритма, однако они не позволяют идентифицировать интенсивность эмоционального состояния субъекта.

Целью настоящего исследования является разработка алгоритма идентификации эмоций человека на основании набора звуковых данных. Для классификации и распознавания эмоций человека применялись аудиозаписи, которые содержали характерные признаки той или иной эмоции. Для реализации процесса распознавания вместо привычной рекуррентной нейронной сети использовалась сверточная нейронная сеть, которая была оптимизирована специально для работы с аудиозаписями и текстом в качестве эксперимента, так как возможно, что за счет использования сверточной нейронной сети можно получить более быстрый результат. Ввиду нестандартного выбора типа нейронной сети, применяемой в исследовании, необходимо было преобразовать входной информационный поток в изображение. Для этого использовалась технология audio fingerprint, основная задача которой – преобразование входного аудиопотока данных в спектрограмму.

Проведение сравнительного эксперимента. Для проведения сравнительного эксперимента и оценки зависимости точности работы алгоритма распознавания эмоций человека от количества идентифицируемых эмоций было принято решение разделить обучение нейронной сети на две независимые модели: на три класса эмоций (позитивные, нейтральные и негативные) и с целью более точной классификации – на восемь классов (счастье, агрессия, спокойствие, отвращение, удивление, нейтральное состояние, печаль, страх). Выбор конкретного набора эмоциональных состояний обусловлен возможностями используемой нейросетевой библиотеки Keras.

На рис. 1 изображено дерево разбиения эмоционального состояния на классы.

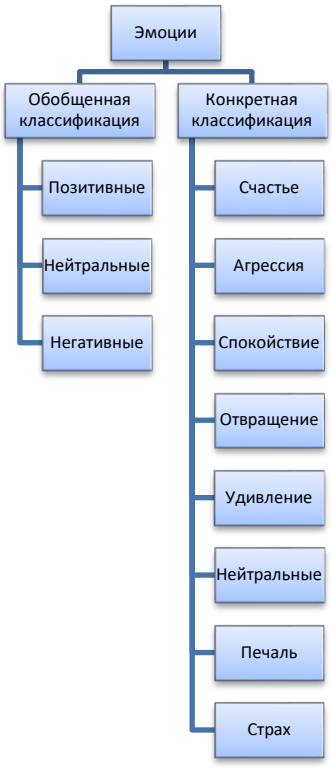


Рис. 1. Иерархия классов эмоций

Fig. 1. Hierarchy of emotion classes

Для обучения нейронной сети в соответствии с иерархией классов (рис. 1) был отобран набор аудиофайлов. Для каждого эмоционального состояния нейронная сеть выделяла признаки на определенном количестве аудиофайлов, отображающих одну и ту же эмоцию. В табл. 1–4 представлены данные о количестве аудиофайлов для каждого класса.

Таблица 1
Количество аудиофайлов для каждого класса обучающей выборки

Table 1
Number of audiofiles for each class of training sample

Класс *Class*	Количество аудиофайлов *Number of images*
Агрессия	665
Спокойствие	299
Отвращение	519
Страх	665
Счастье	665
Нейтральные	244
Печаль	346
Удивление	200
Всего	3603

Таблица 2
Количество аудиофайлов для каждого класса тестовой выборки

Table 2
Number of audiofiles for each class of test set

Класс *Class*	Количество аудиофайлов *Number of images*
Агрессия	167
Спокойствие	75
Отвращение	130
Страх	167
Счастье	167
Нейтральные	113
Печаль	87
Удивление	50
Всего	956

Таблица 3
Количество аудиофайлов для обобщенных классов обучающей выборки

Table 3
The number of audiofiles for generalized classes of training sample

Класс *Class*	Количество аудиофайлов *Number of images*
Позитивные	865
Нейтральные	543
Негативные	2195
Всего	3603

Таблица 4
Количество аудиофайлов для обобщенных классов тестовой выборки

Table 4
Number of audiofiles for generalized classes of test sample

Класс *Class*	Количество аудиофайлов *Number of images*
Позитивные	217
Нейтральные	188
Негативные	551
Всего	956

В табл. 5 приведены результаты сравнительного анализа систем распознавания эмоций и их характеристики [25].

Для разрабатываемого метода идентификации эмоций был использован специальный алгоритм, принимающий на вход набор аудиофайлов. В результате работы данного алгоритма сформировался соответствующий набор спектрограмм.

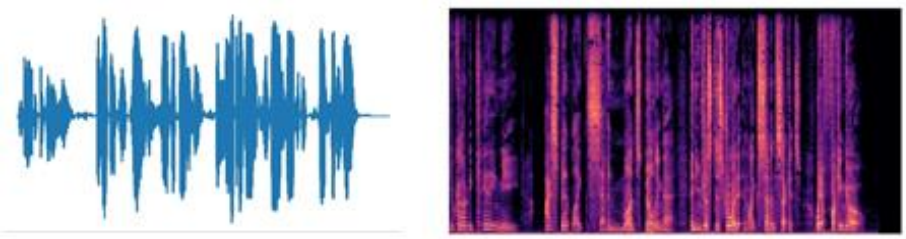
Задачей классификации эмоций занимаются уже давно. Данный факт обусловлен низкой точностью работы применяемых алгоритмов, что требует дальнейших исследований в этой области. Большая часть работ, которые были изучены для разработки стратегии исследования, базировалась на классификации эмоций с помощью видеопотока данных. В качестве аттракторов использовались опорные точки лица. Основная задача нейронной сети – построение карты точек лица. В результате удается сформировать данные, необходимые для тренировки. В связи с необходимостью сложных вычислений для такой задачи используются сверточные нейронные сети. На вход нейронной сети подается набор анализируемых изображений и на основании геометрических параметров лица происходит сегментация отдельных его зон. Далее полученная информация используется для выделения ключевых признаков, на основании которых

и происходит в дальнейшем классификация эмоций. С помощью специализированных данных DataSet, создаваемых в идеальных условиях и применяемых в задачах обучения нейронных сетей, удается разработать оптимальный алгоритм для распознавания [26, 27].

Таблица 5
Программные пакеты для распознавания эмоций на основании видеоданных
Table 5
Software packages for emotion recognition based on video data

Программный пакет *Software package*	Количество эмоций *Number of emotions*	Способы поиска решения *Ways to find a solution*	Методы классификации *Classification methods*
Compound emotion	7	Распознавание эмоций с помощью фильтра Габора	*k*-ближайших соседей. Дискриминантный анализ Кернела
EmotioNet	23	Вычисление евклидова расстояния между нормализованными ориентирами. Вычисление угла между ключевыми точками. Использование фильтра Габора	Дискриминантный анализ Кернела
Real-time mobile	7	Активная форма модели. Смещение между ориентирами	Метод опорных векторов
Local region specific feature	7	Функция извлечения локального двоичного шаблона (LBP). Геометрическая нормализация центра	Метод опорных векторов

Ввиду того что работы по распознаванию эмоций зачастую базируются на видеоданных, было решено разработать модель нейронной сети, позволяющую классифицировать эмоциональные состояния на основании речевого набора данных. На вход алгоритма преобразования поступал аудиофайл, характеризующий определенную эмоцию (рис. 2, *a*). Сигнал анализировался, и в соответствии со спектральной плотностью мощности формировалось изображение на выходе (рис. 2, *b*).



a) *b)*

Рис. 2. Частотная синусоида (*a*) и спектрограмма (*b*)
Fig. 2. Frequency sinusoid (a) and spectrogram (b)

Для реализации разрабатываемого алгоритма было решено создать структуру нейронной сети, включающую следующие слои:

Conv2D – три сверточных слоя;
MaxPooling2D – слой выделения признаков;
Dropout – два слоя коррекции (вносят случайную величину в веса нейронов);
Flatten – конвертер из сверточной структуры в многослойную;
Dense – два слоя многоосной нейронной сети.

В соответствии со структурой нейронной сети построена ее концептуальная модель (рис. 3).

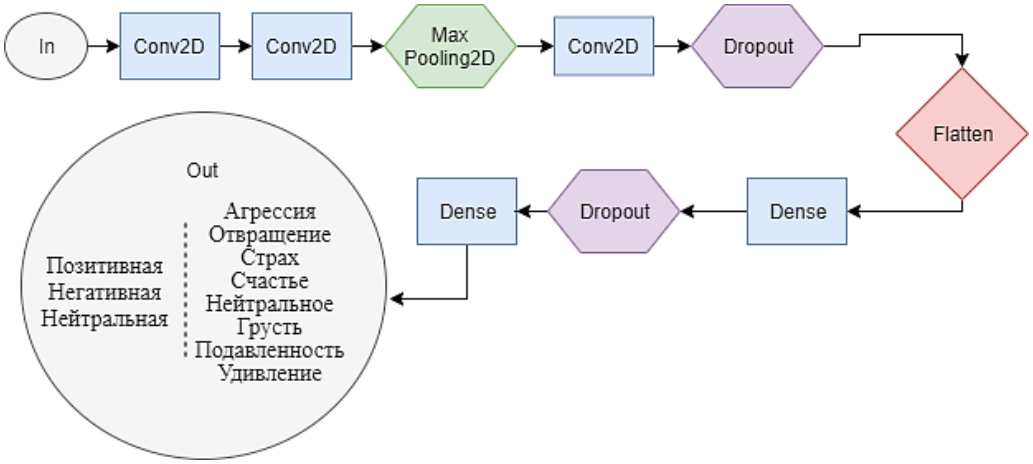


Рис. 3. Концептуальная модель нейронной сети

Fig. 3. Conceptual model of neural network

В итоге созданы две модели для каждого из подходов с различным количеством выходных нейронов. Модели для обоих подходов имеют одинаковую структуру.

Для разработанного проекта были созданы две модели сверточной нейронной сети: для восьми (подробных) и трех (общих) классов. На рис. 4 показана модель для общих классов эмоций. Обе модели имеют полностью одинаковые структуры за исключением выхода, количество нейронов на котором должно быть равно количеству классов.

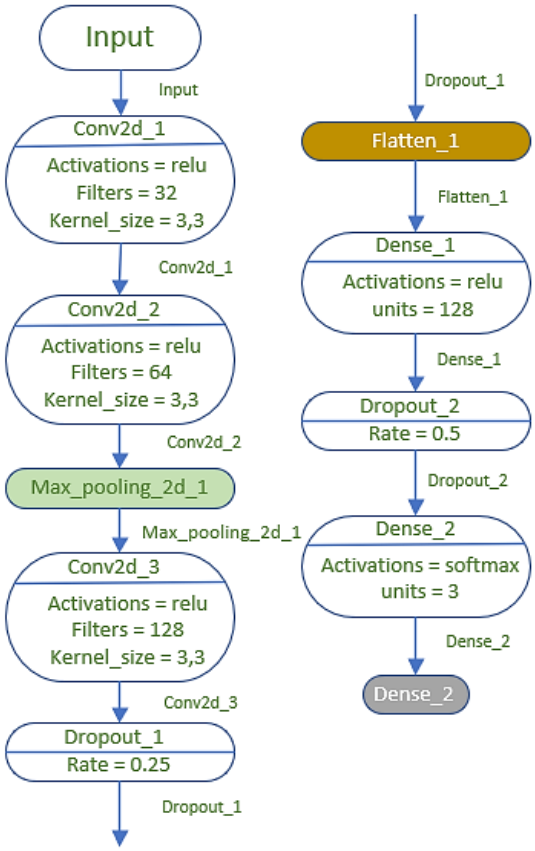


Рис. 4. Модель для трех классов

Fig. 4. Model for three classes

Эксперимент был реализован с помощью следующего набора инструментов:

TensorFlow – открытой программной библиотеки для машинного обучения;

Keras – нейросетевой библиотеки;

Librosa – аудиобиблиотеки для анализа звуковых сигналов;

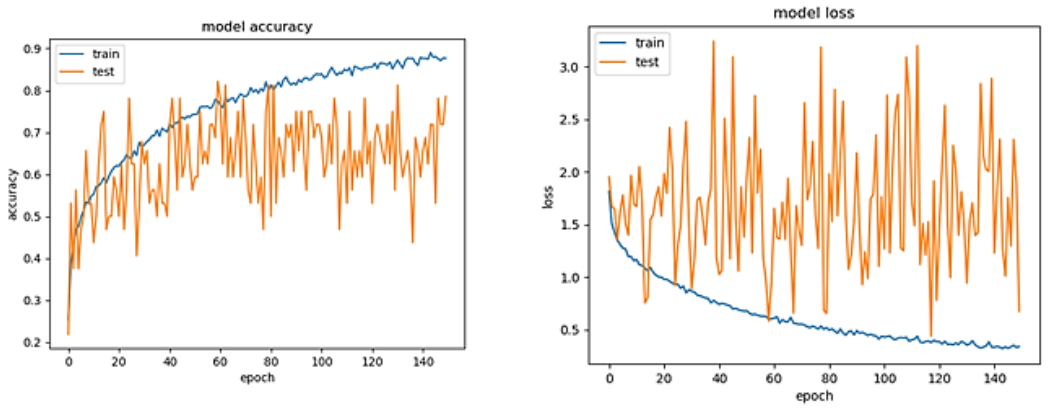
PyAudio – модуля с кроссплатформенной библиотекой PortAudio, позволяющего проигрывать и записывать звуки;

Pillow – библиотеки для работы с изображениями;

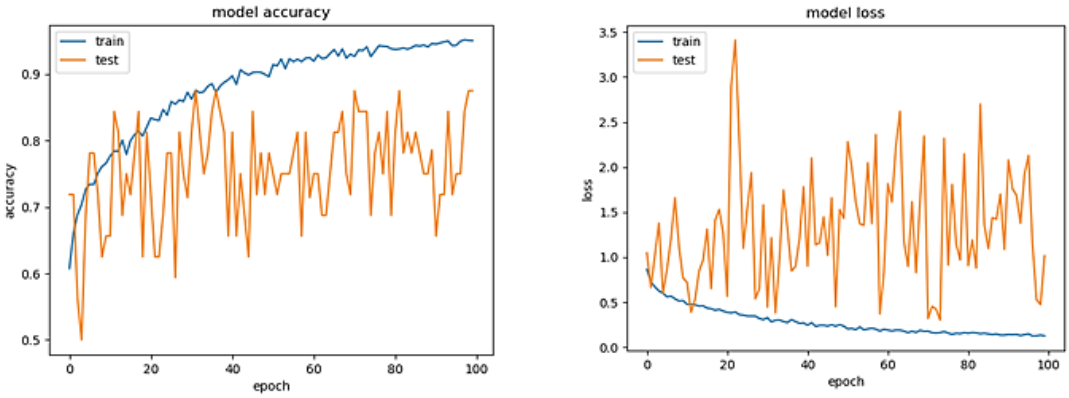
NumPy – библиотеки языка программирования Python для реализации вычислительных алгоритмов, оптимизированной для работы с многомерными массивами (для ускорения вычислительных процессов);

SciPy – библиотеки языка программирования Python для выполнения научных и инженерных расчетов.

Результаты первоначального исследования. Графики точности и ошибки для восьми классов эмоций показаны на рис. 5, *a*, для трех классов – на рис. 5, *b*.



a)



b)

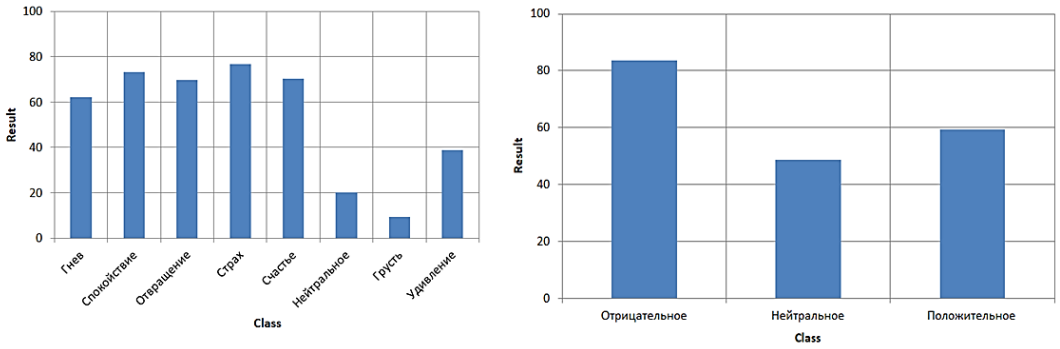
Рис. 5. График точности и график ошибок для восьми классов эмоций (*a*), для трех классов (*b*)

Fig. 5. Accuracy plot and error plot for eight emotion classes (a), for three classes (b)

Модели имеют идентичные структуры и состоят из практически одинаковых наборов файлов: train – обеспечивает процесс обучения нейронной сети, test – необходим для выполнения тестирования на наборе данных, single_test – представляет собой рабочий исполняемый файл, используемый для анализа эмоционального состояния. Принципиальное различие моделей заключается только в выходах, количество нейронов на выходе соответствует количеству классов.

Тестирование выполнялось в несколько этапов: классификация групп (тестирование набора данных) по классам, процентная оценка и тестирование каждого файла в отдельности.

Тестирование набора данных осуществлялось с помощью тестов по 100 файлов в каждой категории. На рис. 6, *a* показаны результаты тестирования для восьми классов эмоций, на рис. 6, *b* – для трех классов.



a) *b)*

Рис. 6. Точность классификации для восьми классов эмоций (*a*), для трех классов (*b*)

Fig. 6. Classification accuracy for eight emotion classes (a), for three classes (b)

При тестировании отдельных образцов для трех классов в качестве теста были взяты следующие эмоции: агрессия, счастье, отвращение, страх, нейтральное состояние, грусть и удивление. Результаты исследований показали, что корректно распознаются не все эмоции. Нейронная сеть хорошо распознает агрессию, счастье, отвращение, страх и грусть. Удивление и нейтральное состояние распознаются некорректно, нейронная сеть относит их к негативным эмоциям.

При тестировании отдельных образцов для восьми классов в качестве теста были использованы такие же параметры. Результаты исследований показали, что 100%-го попадания в класс не было. Нейронная сеть хорошо распознает счастье, отвращение, грусть и удивление, вместо класса «агрессия» получен результат «страх», вместо «нейтральное состояние» – «счастье», а вместо «страх» – «отвращение».

Поиск решения для повышения точности алгоритма. На рис. 6 показано, что тестирование не является однозначным, так как человеческие эмоции определяются сложно. Набор данных не может показать определенный результат, поэтому был выполнен отдельный анализ для каждого тестового образца. Определение обобщенных классов дало более точный результат, чем определение конкретных классов. Стоит заметить, что из-за меньшего количества классов скорость работы программы значительно больше при классификации трех классов, чем при классификации восьми классов, поэтому программа будет требовать для работы меньше процессорного времени.

Согласно рис. 6 получены следующие показатели точности, %:

1. Распознавание трех эмоций:

отрицательная – 83,5;

нейтральная – 48,7;

положительная – 59,3.

2. Распознавание восьми эмоций:

гнев – 62,1;

спокойствие – 73,3;

отвращение – 69,8;

страх – 76,8;

счастье – 70,3;

нейтральное – 20,05;

грусть – 9,4;

удивление – 38,7.

Видно, что точность мала, алгоритм с полученными результатами будет иметь высокую погрешность. Низкая точность может быть обусловлена ошибками первого и второго рода при распознавании.

Следующим этапом исследования стал поиск путей для увеличения точности работы алгоритма. Для этого сначала потребовалась проверка работоспособности готовых нейронных сетей с имеющимся набором данных. Для тестирования алгоритма было решено перейти в среду для разработки Matlab. Были выбраны нейронные сети GoogleNet и Rasnet-50. Они зарекомендовали себя как одни из самых лучших структур, применяемых при классификации изображений. Для проведения исследования эти структуры нейронных сетей были переобучены на имеющийся набор данных. По итогам обучения обе сети дали очень похожие результаты (рис. 7).

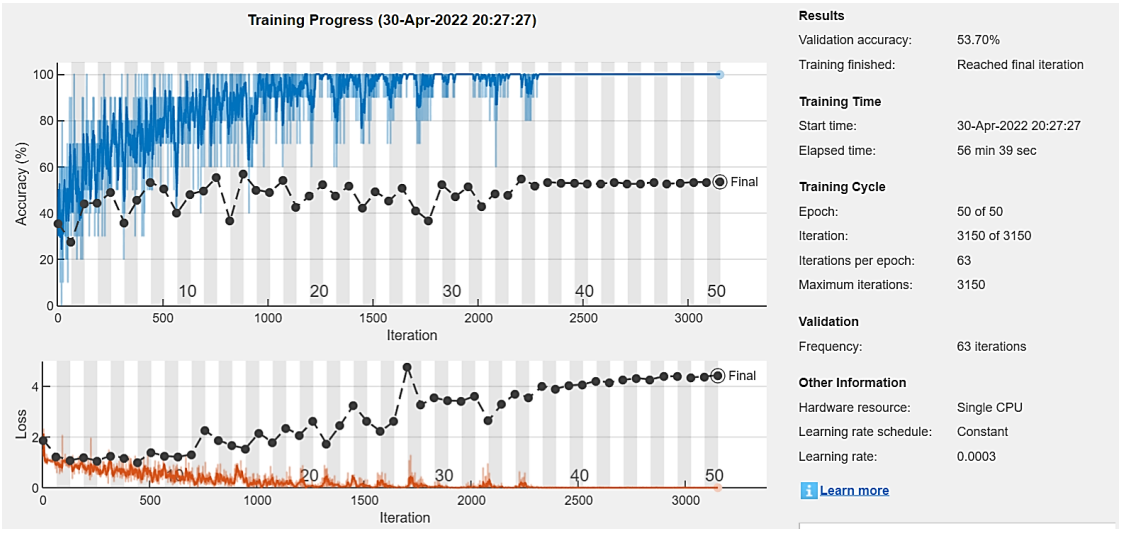
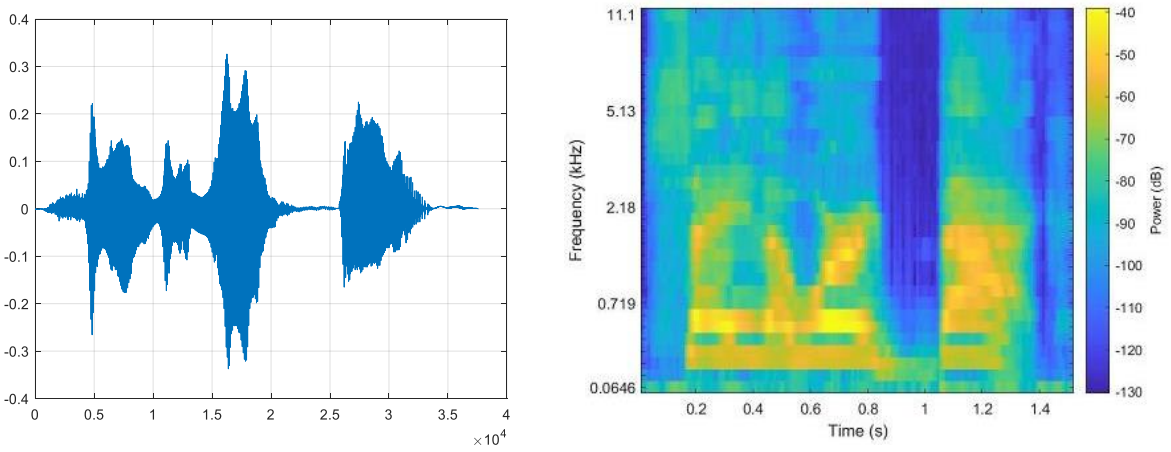


Рис. 7. Результаты обучения готовых структур нейронных сетей

Fig. 7. Results of training of ready-made structures of neural networks

В результате исследования была установлена жесткая зависимость точности работы алгоритма и плохого качества входных данных. Следовательно, необходимо было изменить входной набор данных. Для повышения точности было решено использовать mel-спектрограммы. На рис. 8 показан аудиосигнал и соответствующая ему mel-спектрограмма.

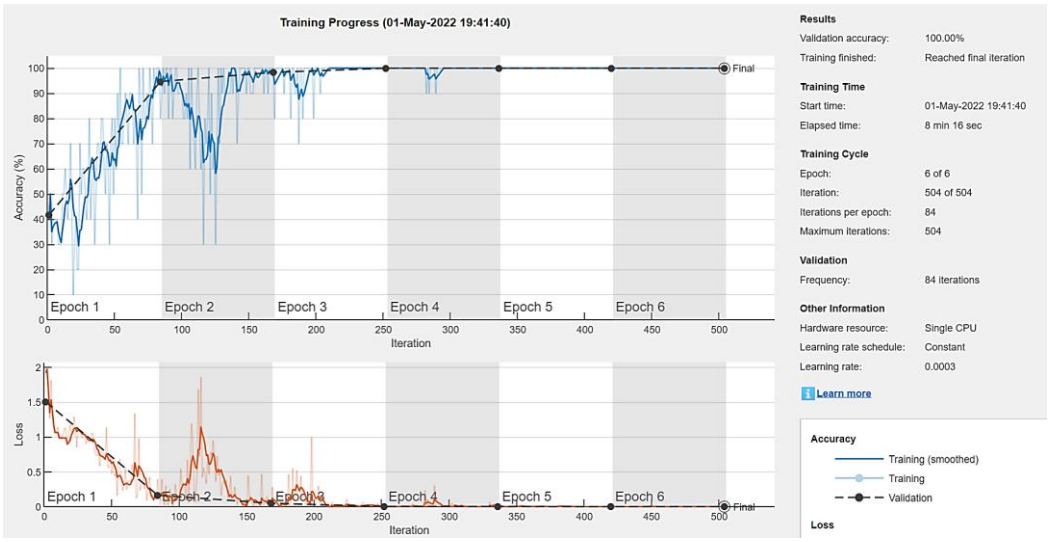


a) *b)*

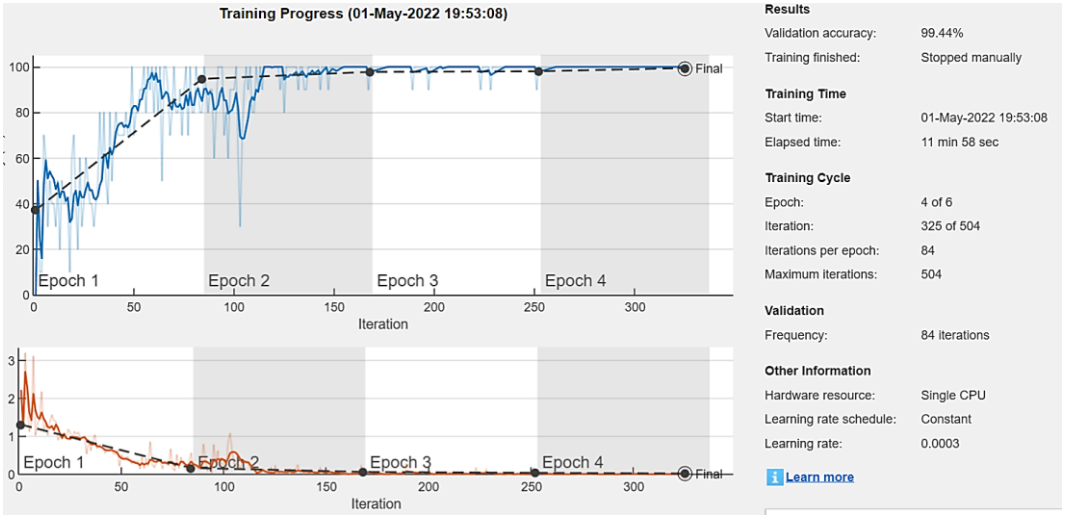
Рис. 8. График аудиосигнала (*a*), MFCC-спектрограмма (*b*)

Fig. 8. Audio signal plot (a), MFCC spectrogram (b)

Все входные данные, полностью преобразованные к виду MFCC-спектрограммы, были поданы вновь на готовые нейронные сети. В результате удалось достичь 100 %-й точности распознавания трех эмоций на тренировочном наборе данных при использовании RasNet-50 (рис. 9, *a*) и 99,44 %-й точности при использовании GoogleNet (рис. 9, *b*).



a)



b)

Рис. 9. Графики обучения нейронных сетей Rasnet-50 (*a*) и GoogleNet (*b*)

Fig. 9. Training graphs for Rasnet-50 (a) and GoogleNet (b) neural networks

Разработка усовершенствованной модели распознавания и анализ полученных данных. После окончания текущего исследования было решено разработать новую структуру нейронной сети, состоящую из 24 слоев, с использованием программного пакета Matlab (рис. 10).

На вход разработанной нейронной сети подавались MFCC-спектрограммы. Полученные данные были разделены следующим образом: 60 % для обучения, 20 % для валидации, 20 % для проверки точности работы (эти данные не использовались при обучении). На рис. 11, *a* приведены графики обучения нейронной сети для распознавания трех эмоций (точность на обучающем наборе данных составила 100 %), на рис. 11, *b* – для распознавания семи эмоций (точность на обучающем наборе данных составила 99,82 %).

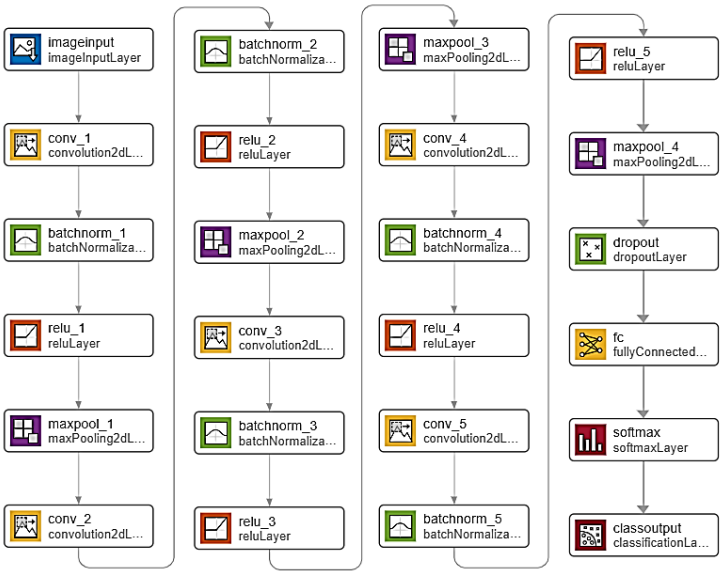
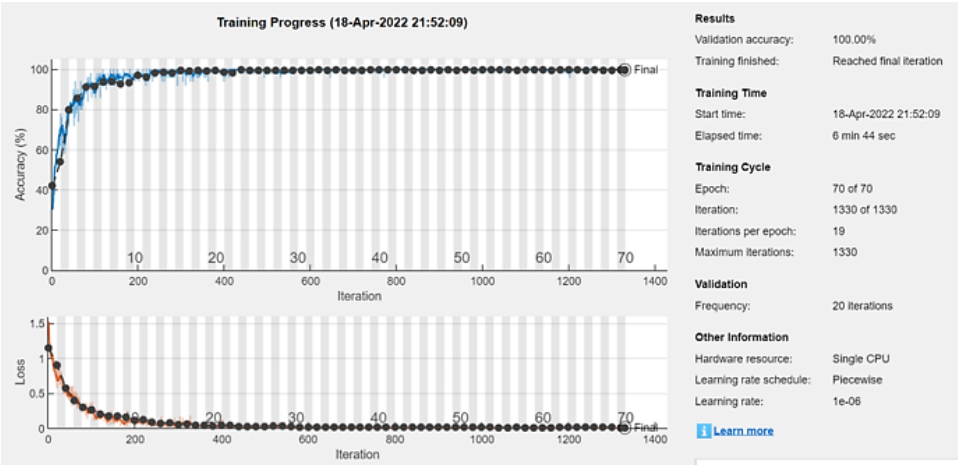
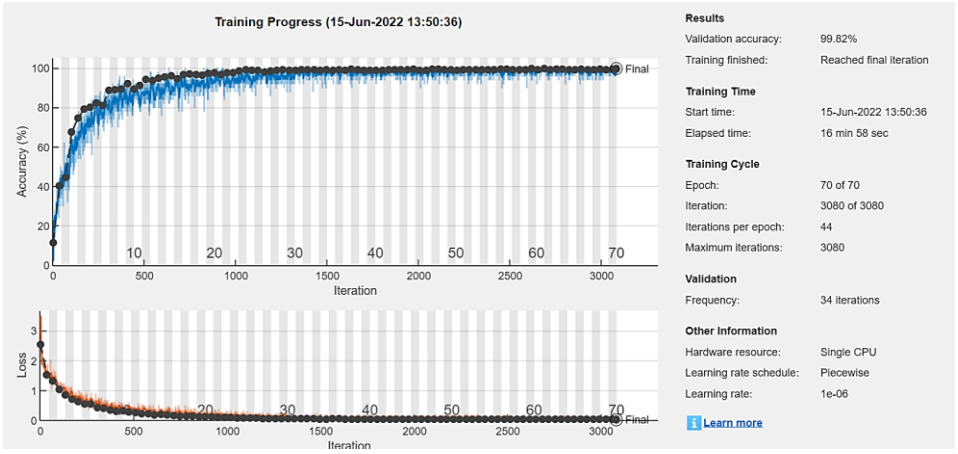


Рис. 10. Структура нейронной сети
Fig. 10. Neural network structure



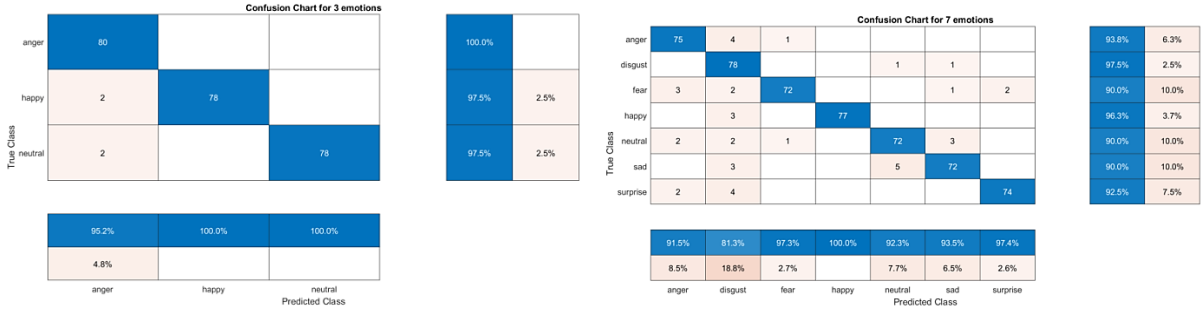
a)



b)

Рис. 11. Результат обучения нейронной сети для трех эмоций (*a*), для семи эмоций (*b*)
Fig. 11. The result of neural network training for three emotions (a), for seven emotions (b)

После обучения для проверки работоспособности на вход нейронной сети подавались данные, не участвовавшие в обучении. Для трех эмоций (рис. 12, *a*) точность составила 98,33 %, а для семи эмоций (рис. 12, *b*) – 92,87 %.



a) *b)*

Рис. 12. Точность работы алгоритма на наборе, не участвовавшем в обучении:
a) для трех эмоций; *b)* для семи эмоций

Fig. 12. The accuracy of the algorithm on the set that did not participate in training:
a) for three emotions; b) for seven emotions

Заключение. Авторами был разработан алгоритм, на вход которого подавались спектрограммы, полученные в результате оконного преобразования Фурье. Однако точность полученного алгоритма была слишком мала, чтобы завершить на этом процесс исследования. Для повышения точности алгоритма было решено провести исследование, позволяющее определить зависимость результата от входных параметров или структуры нейронной сети. Изменение структуры нейронной сети при неизменных входных данных не повысило точность алгоритма, что навело на мысль о необходимости изменения вида входных данных. Для этого входной набор данных был преобразован к виду MFCC, что в последующем показало зависимость точности работы алгоритма от входного набора данных.

Для разработки новой структуры нейронной сети был использован программный пакет Matlab, с помощью которого удалось получить высокие эргономические параметры исследуемой области. В качестве актуальности данной тематики можно отметить высокую точность работы предложенного алгоритма по сравнению с имеющимися на текущий момент вариантами.

Вклад авторов. *В. В. Семенюк* осуществила постановку задачи, определила направление и цель исследования, разработала структуру нейронной сети с использованием языка программирования Python. *М. В. Складчиков* проанализировал полученные данные, разработал концепцию, позволяющую улучшить точность алгоритма путем выявления факторов, влияющих на ее показатели, разработал архитектуру нейронной сети в программе Matlab и провел соответствующие эксперименты, осуществил научное редактирование статьи.

Список использованных источников

1. Mesaros, A. Acoustic scene classification: Overviews of DCASE 2017 challenge entries / A. Mesaros, T. Heittola, T. Virtanen // 16th Intern. Workshop on Acoustic Signal Enhancement (IWAENC 2018), Tokyo, Japan, 17–20 Sept. 2018. – Tokyo, 2018. – P. 411–415.

2. Haitsma, J. A highly robust audio fingerprinting system / J. Haitsma, T. Kalker // 3rd Intern. Conf. on Music Information Retrieval, Paris, France, 13–17 Oct. 2002. – Paris, 2002. – P. 107–115.

3. Ильин, Е. П. Эмоции и чувства / Е. П. Ильин. – СПб. : Питер, 2001. – 752 с.

4. Изард, К. Э. Психология эмоций / К. Э. Изард. – СПб. : Питер, 2012. – 464 с.

5. Карелина, И. О. Развитие понимания эмоций в период дошкольного детства: психологический ракурс : монография / И. О. Карелина. – Прага : Vědecko vydavatelské centrum «Sociosféra-CZ», 2017. – 178 с.

6. Орехова, О. А. Цветовая диагностика эмоций. Типология развития : монография / О. А. Орехова. – СПб. : Речь; М. : Сфера, 2008. – 176 с.

7. Шаповал, Ж. Я. Распознавание эмоций человека по изображению как часть автоматизированного переводчика языка жестов / Ж. Я. Шаповал // Молодежный научно-технический вестник. – 2017. – № 7. – С. 55.

8. Голубинский, А. Н. Выявление эмоционального состояния человека по речевому сигналу на основе вейвлет-анализа / А. Н. Голубинский // Вестник Воронежского института МВД России. – 2011. – № 3. – С. 144–153.

9. Сидоров, К. И. Автоматическое распознавание эмоций человека на основе реконструкций аттракторов образцов речи / К. И. Сидоров, Н. Н. Филатова // Программные системы и вычислительные методы. – 2012. – № 1. – С. 67–79.

10. Галичий, Д. А. Распознавание эмоций человека при помощи современных методов глубокого обучения / Д. А. Галичий, Г. И. Афанасьев, Ю. Г. Нестеров // E-SCIO. – 2021. – Т. 5, № 56. – С. 316–329.

11. Бредихин, А. И. Применение вейвлетов в задаче распознавания эмоций человека по его речи / А. И. Бредихин // Сборник избранных статей научной сессии ТУСУР. – 2018. – № 1–3. – С. 115–119.

12. Рюмина, Е. В. Аналитический обзор методов распознавания эмоций по выражениям лица человека / Е. В. Рюмина, А. А. Карпов // Научно-технический вестник информационных технологий, механики и оптики. – 2020. – Т. 20, № 2. – С. 163–176.

13. Dvoinikova, A. Emotion recognition and sentiment analysis of extemporaneus speech transcriptions in Russian / A. Dvoinikova, O. Verkholyak, A. Karpov // Lectures notes in computer science. – 2020. – Vol. 12335. – P. 136–144. https://doi.org/10.1007/978-3-030-60276-5_14

14. Devi, J. S. Speaker emotion recognition based on speech feateres and classification techniques / J. S. Devi, S. Yarrammelle, S. P. Nandyala // Intern. J. of Image, Graphics, and Signal Processing. – 2014. – Vol. 6, no. 7. – P. 61–77. https://doi.org/10.5815/ijigsp.2014.07.08

15. Speech emotion recognition based on an improved brain emotion learning model / Z. I. Liu [et al.] // Neurocomputing. – 2018. – Vol. 309. – P. 145–156. https://doi.org/10.1016/ j.neucom.2018.05.005

16. Shirami, A. Speech emotion recognition based on SVM as both features selector and classifier / A. Shirami, A. R. N. Nilchi // Intern. J. of Image, Graphics, and Signal Processing. – 2016. – Vol. 8, no. 4. – P. 39–45. https://doi.org/10.5815/ijigsp.2016.04.05

17. Assuncao, G. Intermediary fuzzyfication in speech emotion recognition / G. Assuncao, P. Menezes // IEEE Intern. Conf. on Fuzzy System, Glasgow, United Kingdom, 19–24 July 2020. – Glasgow, 2020. – P. 9177699. https://doi.org/10.1109/FUZZ48607.2020.9177699

18. Zisad, S. N. Speech emotion recognition in neurological disorders using convolutional neural network / S. N. Zisad, M. S. Hossain, K. Andersson // Lecture Notes in Computer Science. – 2020. – Vol. 12241. – P. 287–296. https://doi.org/10.1007/978-3-030-59277-6_26

19. Werner, S. Speech emotion recognition: hemans vs machines / S. Werner, G. K. Petrenko // Discourse. – 2019. – Vol. 5, no. 5. – P. 136–152. https://doi.org/10.32603/2412-8562-2019-5-5-136-152

20. Muppidi, A. Speech emotion recognition using quaternion convolutional neural networks / A. Muppidi, M. Radfar // IEEE Intern. Conf. of Acoustics, Speech and Signal Processing-Proceedings, Toronto, ON, Canada, 6–11 June 2021. – Toronto, 2021. – P. 6309–6313. https://doi.org/10.1109/ICASSP39728.2021.9414248

21. Zheng, W. Multi-scale discrepancy adversarial network for crosscorpus speech emotion recognition / W. Zheng, Y. Zong // Virtual Reality and Intelligent Hardware. – 2021. – Vol. 3, no. 1. – P. 65–75. https://doi.org/10.1016/j.vrih.2020.11.006

22. Hazjan, V. Context-independent multilingual emotion recognition from speech signals / V. Hazjan, Z. Kacic // Intern. J. of Speech Technology. – 2003. – Vol. 6, no. 3. – P. 311–320.

23. Zhang, C. Autoencoder with emotion embedding for speech emotion recognition / C. Zhang, L. Xue // IEEE Access. – 2021. – Vol. 9. – P. 51231–51241. https://doi.org/10.1109/ACCESS.2021.3069818

24. Kanwal, S. Speech emotion recognition using clustering based GA-optimized feature set / S. Kanwal, S. Asghar // IEEE Access. – 2021. – Vol. 9. – P. 125830–125842. https://doi.org/10.1109/ACCESS.2021.3111659

25. Byoung, C. K. A brief review of facial emotion recognition based on visual information / C. K. Byoung // Sensors. – 2018. – Vol. 18, iss. 2. – P. 401. https://doi.org/10.3390/s18020401

26. Audio-visual emotion recognition using deep transfer learning and multiple temporal models / X. Ouyang [et al.] // ICMI '17 : Proc. of the 19th ACM Intern. Conf. on Multimodal Interaction, Glasgow, United Kingdom, 13–17 November 2017. – Glasgow, 2017. – P. 577–582. https://doi.org/10.1145/3136755.3143012

27. Hassani, B. Facial expression recognition using enhanced deep 3D convolutional neural networks / B. Hassani, M. H. Mahoor // 2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017. – Honolulu, 2017. – P. 1955–1962. https://doi.org/10.1109/CVPRW.2017.282

References

1. Mesaros A., Heittola T., Virtanen T. Acoustic scene classification: Overviews of DCASE 2017 challenge entries. *16th International Workshop on Acoustic Signal Enhancement (IWAENC 2018), Tokyo, Japan, 17–20 September 2018*. Tokyo, 2018, pp. 411–415.

2. Haitsma J., Kalker T. A highly robust audio fingerprinting system. *3rd International Conference on Music Information Retrieval, Paris, France, 13–17 October 2002*. Paris, 2002, pp. 107–115.

3. Ilin E. P. Jemocii i chuvstva. *Emotions and Feelings*. Saint Petersburg, Piter, 2001, 752 p. (In Russ.).

4. Izard K. E. Psihologija jemocij. *Psychology of Emotions*. Saint Petersburg, Piter, 2012, 464 p. (In Russ.).

5. Karelina I. O. Razvitie ponimanija jemocij v period doshkol'nogo detstva: psihologicheskij rakurs. *Developing an Understanding of Emotions during Preschool Childhood: A Psychological Perspective*, Prague, Vědecko vydavatelské centrum "Sociosféra-CZ", 2017, 178 p. (In Russ.).

6. Orehova O. A. Cvetovaja diagnostika jemocij. Tipologija razvitija. Monografija. *Color Diagnostics of Emotions. Typology of Development. Monograph*. Saint Petersburg, Sphere, 2008, 176 p. (In Russ.).

7. Shapoval J. A. *Recognition of Human Emotions by image as part of an automated sign language translator*. Molodezhnyj nauchno-tekhnicheskij vestnik [*Youth Scientific and Technical Bulletin*], 2017, no. 7, p. 55 (In Russ.).

8. Golubinskij A. N. *Identification of a person's emotional state by a speech signal based on a Wavelet analysis*. Vestnik Voronezhskogo instituta Ministerstva vnutrennih del Rossii [*Bulletin of the Voronezh Institute of the Ministry of Internal Affairs of Russia*], 2011, no. 3, pp. 144–153 (In Russ.).

9. Sidorov K. I., Filatova N. N. *Automatic recognition of human emotions based on reconstructions of attractors of speech samples*. Programmnye sistemy i vychislitel'nye metody [*Software systems and computational methods*], 2012, no. 1, pp. 67–79 (In Russ.).

10. Galichij D. A., Afanaciev G. I., Nesterov U. G. Recognition of human emotions using modern methods of deep learning. *E-SCIO*, 2021, vol. 5, no. 56, pp. 316–329 (In Russ.).

11. Bredihin A. I. *The use of wavelets in the task of recognizing a person's emotions by his speech*. Sbornik izbrannyh statej nauchnoj sessii Tomskogo gosudarstvennogo universiteta sistem upravlenija i radiojelektroniki [*Collection of selected articles of the scientific session of Tomsk State University of Control Systems and Radioelectronics*], 2018, no. 1–3, pp. 115–119 (In Russ.).

12. Rumina E. V., Karpov A. A. *Analytical review of emotion recognition methods based on human facial expressions*. Nauchno-tekhnicheskij vestnik informacionnyh tekhnologij, mekhaniki i optiki [*Scientific and Technical Bulletin of Information Technologies, Mechanics and Optics*], 2020, vol. 20, no. 2, pp. 163–176 (In Russ.). https://doi.org/10.17586/2226-1494-2020-20-2-163-176

13. Dvoinikova A., Verkholyak O., Karpov A. Emotion recognition and sentiment analysis of extemporaneus speech transcriptions in Russian. *Lectures Notes in Computer Science*, 2020, vol. 12335, pp. 136–144. https://doi.org/10.1007/978-3-030-60276-5_14

14. Devi J. S., Yarrammelle S., Nandyala S. P. Speaker emotion recognition based on speech feateres and classification techniques. *International Journal of Image, Graphics, and Signal Processing*, 2014, vol. 6, no. 7, pp. 61–77. https://doi.org/10.5815/ijigsp.2014.07.08

15. Liu Z. I., Xie Q., Wu M., Cao W. H., Mao J. W., Mei Y. Speech emotion recognition based on an improved brain emotion learning model. *Neurocomputing*, 2018, vol. 309, pp. 145–156. https://doi.org/10.1016/j.neucom.2018.05.005

16. Shirami A., Nilchi A. R. N. Speech emotion recognition based on SVM as both features selector and classifier. *International Journal of Image, Graphics, and Signal Processing*, 2016, vol. 8, no. 4, pp. 39–45. https://doi.org/10.5815/ijigsp.2016.04.05

17. Assuncao G., Menezes P. Intermediary fuzzyfication in speech emotion recognition. *IEEE International Conference on Fuzzy System, Glasgow, United Kingdom, 19–24 July 2020*. Glasgow, 2020, p. 9177699. https://doi.org/10.1109/FUZZ48607.2020.9177699

18. Zisad S. N., Hossain M. S., Andersson K. Speech emotion recognition in neurological disorders using convolutional neural network. *Lecture Notes in Computer Science*, 2020, vol. 12241, pp. 287–296. https://doi.org/10.1007/978-3-030-59277-6_26

19. Werner S., Petrenko G. K. Speech emotion recognition: hemans vs machines. *Discourse*, 2019, vol. 5, no. 5, pp. 136–152. https://doi.org/10.32603/2412-8562-2019-5-5-136-152

20. Muppidi A., Radfar M. Speech emotion recognition using quaternion convolutional neural networks. *IEEE International Conference of Acoustics, Speech and Signal Processing-Proceedings, Toronto, ON, Canada, 6–11 June 2021*. Toronto, 2021, pp. 6309–6313. https://doi.org/10.1109/ICASSP39728.2021.9414248

21. Zheng W., Zong Y. Multi-scale discrepancy adversarial network for crosscorpus speech emotion recognition. *Virtual Reality and Intelligent Hardware*, 2021, vol. 3, no. 1, pp. 65–75. https://doi.org/10.1016/j.vrih.2020.11.006

22. Hazjan V., Kacic Z. Context-independent multilingual emotion recognition from speech signals. *International Journal of Speech Technology*, 2003, vol. 6, no. 3, pp. 311–320.

23. Zhang C., Xue L. Autoencoder with emotion embedding for speech emotion recognition. *IEEE Access*, 2021, vol. 9, pp. 51231–51241. https://doi.org/10.1109/ACCESS.2021.3069818

24. Kanwal S., Asghar S. Speech emotion recognition using clustering based GA-optimized feature set. *IEEE Access*, 2021, vol. 9, pp. 125830–125842. https://doi.org/10.1109/ACCESS.2021.3111659

25. Byoung C. K. A brief review of facial emotion recognition based on visual information. *Sensors*, 2018, vol. 18, iss. 2, pp. 401. https://doi.org/10.3390/s18020401

26. Ouyang X., Kawaai S., Goh E. G. H., Shen S., Ding W., …, D.-Y. Huang. Audio-visual emotion recognition using deep transfer learning and multiple temporal models. *ICMI '17 : Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, United Kingdom, 13–17 November 2017*. Glasgow, 2017, pp. 577–582. https://doi.org/10.1145/3136755. 3143012

27. Hassani B., Mahoor M. H. Facial expression recognition using enhanced deep 3D convolutional neural networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017.* Honolulu, 2017, pp. 1955–1962. https://doi.org/10.1109/CVPRW.2017.282

Информация об авторах

Семенюк Виктория Валерьевна, магистр технических наук, преподаватель специальных дисциплин, Донецкий техникум промышленной автоматики имени А. В. Захарченко.
E-mail: semenuk.viktoriya@gmail.com

Складчиков Максим Владимирович, магистр технических наук, преподаватель специальных дисциплин, Донецкий техникум промышленной автоматики имени А. В. Захарченко.
E-mail: maxsklad19981@yandex.ru

Information about the authors

Viktoriya V. Semenuk, M. Sc. (Eng.), Teacher of Special Disciplines, Donetsk Technical School of Industrial Automation after A. V. Zakharchenko.
E-mail: semenuk.viktoriya@gmail.com

Maxim V. Skladchikov, M. Sc. (Eng.), Teacher of Special Disciplines, Donetsk Technical School of Industrial Automation after A. V. Zakharchenko.
E-mail: maxsklad19981@yandex.ru