

РАЗРАБОТКА АЛГОРИТМА КЛАССИФИКАЦИИ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ СУБЪЕКТА НА ОСНОВАНИИ РЕЧЕВЫХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ПАКЕТА MATLAB

Семенюк Виктория Валерьевна

Аспирант, Южно-Российский государственный университет (НПИ) имени М.И. Платова, г. Новочеркасск
semenuk.viktoriya@gmail.com

Складчиков Максим Владимирович

Аспирант,
Донецкий национальный технический университет
maxsklad19981@yandex.ru

DEVELOPMENT OF AN ALGORITHM FOR CLASSIFYING THE EMOTIONAL STATE OF THE SUBJECT BASED ON SPEECH DATA USING THE MATLAB PACKAGE

**V. Semenyuk
M. Skladchikov**

Summary. The purpose of the work is to analyze current models, approaches and algorithms of image recognition systems. The implementation of a software package based on convolutional neural networks in order to improve the quality of pattern recognition and the construction of an optimal algorithm for recognizing objects based on the received speech data.

Computer pattern recognition is a rather complex task, which is successfully solved through the use of artificial neural networks. Automatic identification of images (text, sound, face, objects, etc.) using a computer is one of the most promising areas of development of artificial intelligence technologies, which allows us to give a key to understanding the features of human intelligence.

To classify the emotional state of the subject, an algorithm was created using convolutional neural networks. For training, a specialized data set was selected, presented in the form of audio recordings characterizing a separate emotional state of the subject. The input of the trained model received transformed speech data in the form of a spectrogram image containing characteristic signs of a certain emotion.

As a result of the study, an algorithm for classifying the emotional color of the subject based on speech data was developed. Experimental studies were carried out to assess the accuracy of the algorithm. This study is a logical continuation of the work in which a similar algorithm was also created [1]. However, due to the low accuracy of the work, it was necessary to improve it. As a result, it was decided to create a different neural network structure by means of the Matlab package. The proposed algorithm, due to its own universality, can be applied in various fields for the tasks of recognizing the emotional color of the subject.

Keywords: neural network, human emotion recognition, convolutional neural network, sound fingerprinting, Tenserflow, Matlab.

Аннотация. Целью работы является анализ текущих моделей, подходов и алгоритмов систем распознавания образов. Реализации программного комплекса на основании сверточных нейронных сетей с целью повышения качества распознавания образов и построение оптимального алгоритма для распознавания объектов на основе полученных речевых данных.

Компьютерное распознавание образов — является довольно сложной задачей, которая успешно решается за счет применения искусственных нейронных сетей. Автоматическая идентификация образов (текст, звук, лицо, предметы и т.д.) с помощью компьютера является одним из наиболее перспективных направлений развития технологий искусственного интеллекта, позволяющая дать ключ к пониманию особенностей работы человеческого интеллекта.

Для классификации эмоционального состояния субъекта был создан алгоритм с использованием сверточных нейронных сетей. Для обучения был выбран специализированный набор данных, представленный в виде аудиозаписей, характеризующих отдельное эмоциональное состояние субъекта. На вход обученной модели поступали преобразованные речевые данные в виде изображения спектрограммы, содержащие характерные признаки определенной эмоции.

В результате проведения исследования, был разработан алгоритм классификации эмоционального окраса субъекта на основе речевых данных. Были проведены экспериментальные исследования, позволяющие оценить точность работы алгоритма. Данное исследование является логическим продолжением работы, в которой также создавался подобный алгоритм [1]. Однако из-за малой точности работы, необходимо было его усовершенствовать. В результате, было решено создать иную структуру нейронной сети средствами пакета Matlab. Предложенный алгоритм, ввиду собственной универсальности, может применяться в различных сферах для задач распознавания эмоционального окраса субъекта.

Ключевые слова: нейронная сеть, распознавание эмоций человека, сверточная нейронная сеть, дактилоскопия звука, Tenserflow, Matlab.

Анализ задач распознавания образов

В современном мире, когда технический прогресс достиг определённого уровня, появляются новые возможности, что побуждает к созданию новых или усовершенствованию имеющихся систем. Для достижения требуемых эргономических параметров, остро стоит

задача создания алгоритмов, являющихся результатом объединения различных вариантов систем. Помимо задачи создания нового варианта структуры системы, необходимо уделить внимание вопросу самоорганизации.

Такие системы, в настоящее время, являются наиболее актуальными. Это связано, в первую очередь,

с принципом работы описанных систем. В зависимости от информации, которая поступает на вход данной системы, алгоритм подстраивается под входные данные. В том случае, когда на входе появляется новый набор данных, не участвующий в обучении, алгоритм запоминает её. Тем самым удаётся накапливать опыт, содержащийся в наборе данных. Данная система наиболее актуальна в том случае, когда имеется потребность работы с большим количеством данных (изображение, звук) [2]. Системы распознавания образов применяются во множестве сфер и позволяют решить множество практических задач:

1. Робототехника [3, 4].

Ярким примером является разработка корпорации Microsoft — Kinect. Данный контроллер изначально был разработан для игровой индустрии. Его основная задача заключалась в распознавании образов человека, а именно — Skeleton Tracking. Данная технология позволяет отслеживать перемещение тела человека и воссоздавать его движения за счёт управления виртуальным аватаром. Отслеживание производится на основании работы двух камер. Одна из них — цветная камера. Её основная задача — получение кадров и подача полученных изображений на алгоритм сравнения. Вторая камера — инфракрасная. Её основная задача заключается в построении карты глубины. В результате, на вход подсистемы сравнения поступает два изображения — цветное и бинаризованное. На основании совместного наложения каждого из них друг на друга, представляется возможность найти человека в области видимости контроллера. После, информация поступает на нейросетевой алгоритм, основная задача которого, заключается в разбиении изображения на области, а также, сегментировании его на составные части тела человека. Выходом системы являются точки, отвечающие каждому суставу отслеживаемого человека в пределах видимости контроллера (максимально количество отслеживаемых суставов — 21). На основании построения карты глубины, а также, стереотриангуляции, возможно отслеживание координат суставов в 3-хмерной плоскости.

По мере продвижения данного продукта в массы, разработкой заинтересовались компании, занимающиеся созданием робототехнических систем. В результате, корпорация Microsoft перевела свой вектор внимания с игровой индустрии на робототехнику. В настоящее время, имеется целое семейство датчиков, которые позволяют создавать собственные алгоритмы, позволяющие обучить нейронную сеть на идентификацию жестов.

Особого внимания заслуживают антропоморфные роботы. Человекоподобные роботы повсеместно применяются в различных сферах деятельности. Антропоморфные роботы могут полностью либо частично

заменить труд человека. Такая замена человеческого труда повышает финансово-экономический потенциал предприятия. В том случае, когда подобные роботизированные системы имеют непосредственный контакт с человеком, например, в медицине, то без использования алгоритмов распознавания образов или эмоций не обойтись.

Используя возможности датчика, был разработан алгоритм формирования траектории для роботизированной системы с использованием контроллера Kinect [5].

2. Системы безопасности [6, 7].

3. Медицина [8].

4. Системы управления и безопасности автомобиля [9].

Ввиду сложной природы образов, подлежащих идентификации, разрабатываемые системы опираются на определённые «триггеры» или классы, которые описывают распознаваемый объект. На основании увеличения спроса на системы идентификации, увеличилась и сложность разрабатываемых систем. Одним из вариантов подобных систем являются алгоритмы распознавания эмоций человека.

Эмоции человека — это реакция организма на внутренние или внешние триггеры. Как утверждает физиолог Анохин П.К., способность выражать эмоции, является результатом эволюционного становления современного человека. Он рассматривал способность выражения эмоций с точки зрения адаптации первобытного человека к жёстким условиям внешнего мира. В современном же мире, эмоциональный окрас субъекта позволяет наладить межличностные отношения и укрепить небывальную связь в обществе.

Если рассматривать эмоции человека, как объект для идентификации, то разрабатываемая система должна учитывать множество факторов. Ввиду индивидуальности каждого субъекта, а также его психологического состояния, которое в период жизни может изменяться, довольно сложно выделить единый способ оценки эмоций. При попытке построить информационную модель системы идентификации, имитирующую психологическое поведение человека, то полностью и достаточно точно описать внутреннюю структуру не удастся. В науке подобные системы принято называть «чёрный ящик». При описании задачи идентификации эмоций, как последовательного алгоритма, мы в состоянии проанализировать внешнее воздействие, на основании которых можем предположить результат действия субъекта. Однако полностью восстановить внутреннюю структуру и модель поведения человека, которая необходима для построения системы распознавания эмоций, практически невозможно. Описанная проблематика анали-

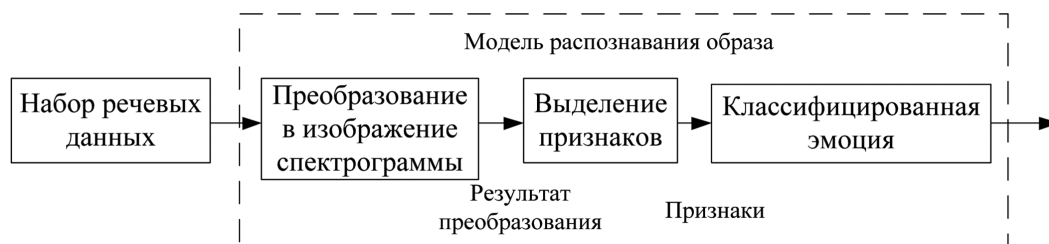


Рис. 1. Структура распознавания

за связана в первую очередь с различными моделями поведения и различной реакцией особи на внешние возмущение. В связи с увеличением сфер применения, направленных на решение сложных технических задач, которые в свою очередь требуют внедрения систем идентификации эмоций (робототехника, системы безопасности и так далее), необходимо систематизировать существующие подходы к решению поставленной задачи, а также, разработать новый алгоритм для классификации эмоций человека [10–14].

Разработка алгоритма классификации эмоций

В ходе проведения исследования, было решено создать две модели нейронной сети, позволяющие идентифицировать:

1. Три класса эмоций: позитивные, негативные и нейтральные.
2. Семь классов эмоций: агрессия, отвращение, страх, счастье, нейтральное состояние, печаль, удивление.

Основная структура распознавания эмоций человека представлена на рис. 1.

Рассмотрим данную структуру более подробно.

Основной целью преобразования входных аудиоданных является построение частотной синусоиды, которая в последующем будет представлена в виде изображения спектрограммы.

Пусть имеется входной набор данных, представленный в виде аудиоволны (рис. 2).



Рис. 2. Пример аудиосигнала

В любой момент времени можно получить данные об амплитуде сигнала и его гармонической составляющей. На основании амплитудных и гармонических колебаний, представляется возможным выделить акустический спектр тона. При его анализе возможно выделить несущую (ключевую) частоту, характерную для отдельного эмоционального состояния человека.

Подобное исследование уже проводилось [1]. Обучение проводилось на основании речевой базы данных. Перед тренировкой нейронной сети, речевые данные подвергались преобразованию в спектрограмму. Однако точность работы алгоритма, полученная в результате обучения нейронной сети, являлась неудовлетворительной, что оставляло вопрос улучшения алгоритма открытым. На рис. 3 представлены гистограммы, отражающие точность работы алгоритма.

Из рис. 3 имеем, что точность распознавания для 3 эмоций:

1. Negative — 83,5%.
2. Neutral — 48,7%.
3. Positive — 59,3%.

При распознавании 8 классов эмоций:

1. Anger — 62,1%.
2. Calm — 73,3%.
3. Disgust — 69,8%.
4. Fear — 76,8%.
5. Happy — 70,3%.
6. Neutral — 20,05%.
7. Sadness — 9,4%.
8. Surprise — 38,7%.

Для улучшения работоспособности алгоритма, было решено заменить метод извлечения характеристик для обучения нейронной сети. В предыдущем исследовании для построения спектрограмм, использовалось оконное преобразование Фурье. Однако проведя более глубокое исследование, было обнаружено, что использование мел-спектрограмм позволит значительно увеличить точность работы алгоритма.

Для проведения данного исследования, было решено выбрать программную среду Matlab. Данный выбор можно обосновать тем, что в Matlab имеется мощный инструмент Deep Network Designer, позволяющий визуализировать

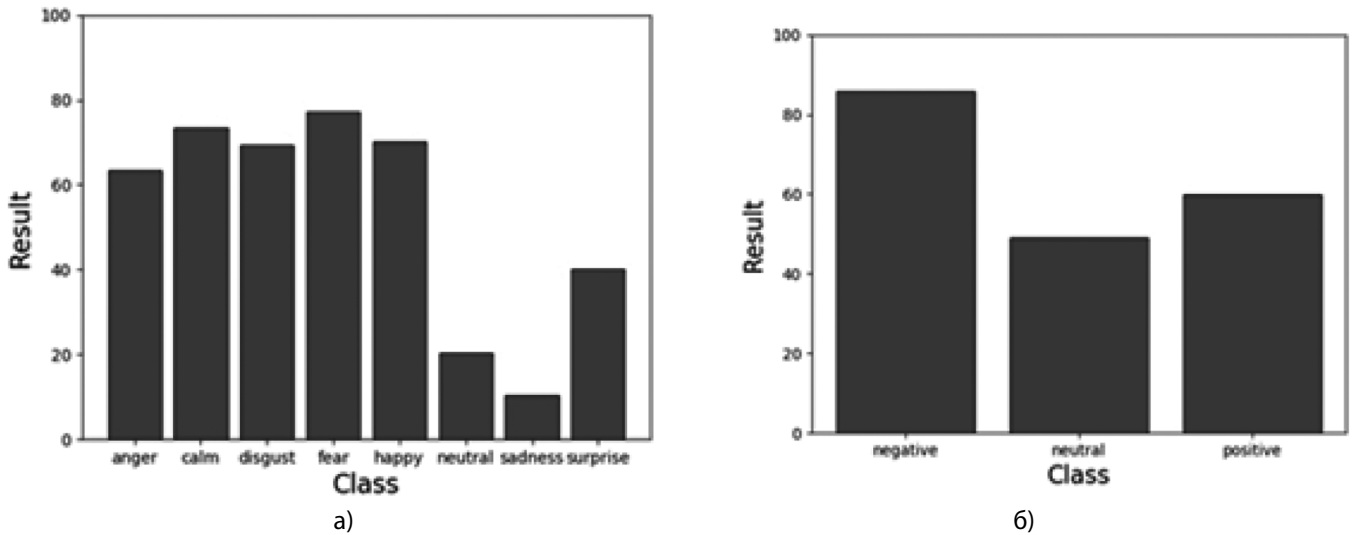


Рис. 3. Результаты тестирования для 8 классов (а) и результаты тестирования для 3 классов (б)

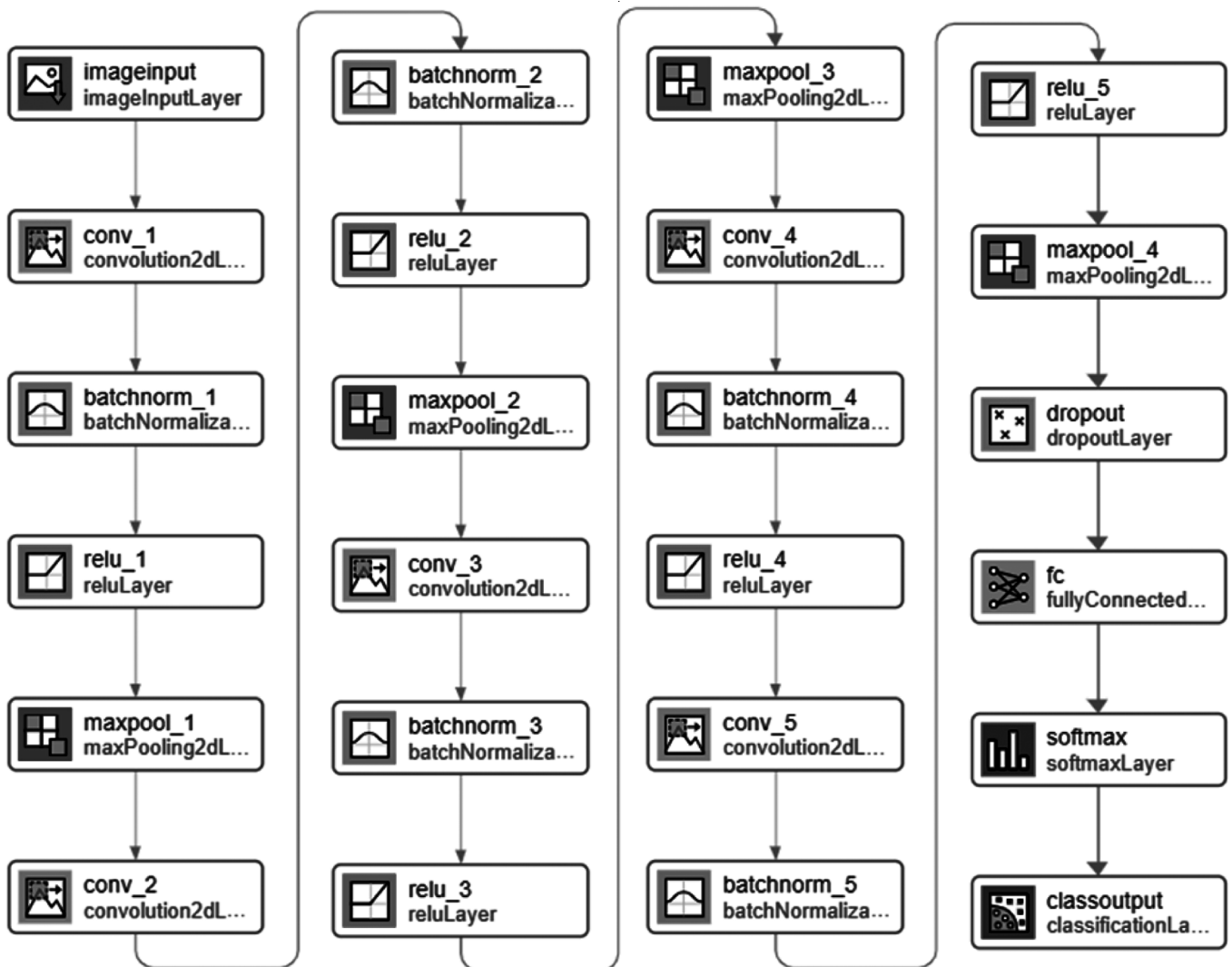


Рис. 4. Структура нейронной сети

зировать процесс создания нейронной сети. Также, имеется возможность параллельной работы в среде Matlab, что уменьшает время, затрачиваемое на обучение.

Для обучения использовался набор речевых данных по 400 аудиозаписей для каждой эмоции. Было решено разделить данные в следующем процентном соотноше-

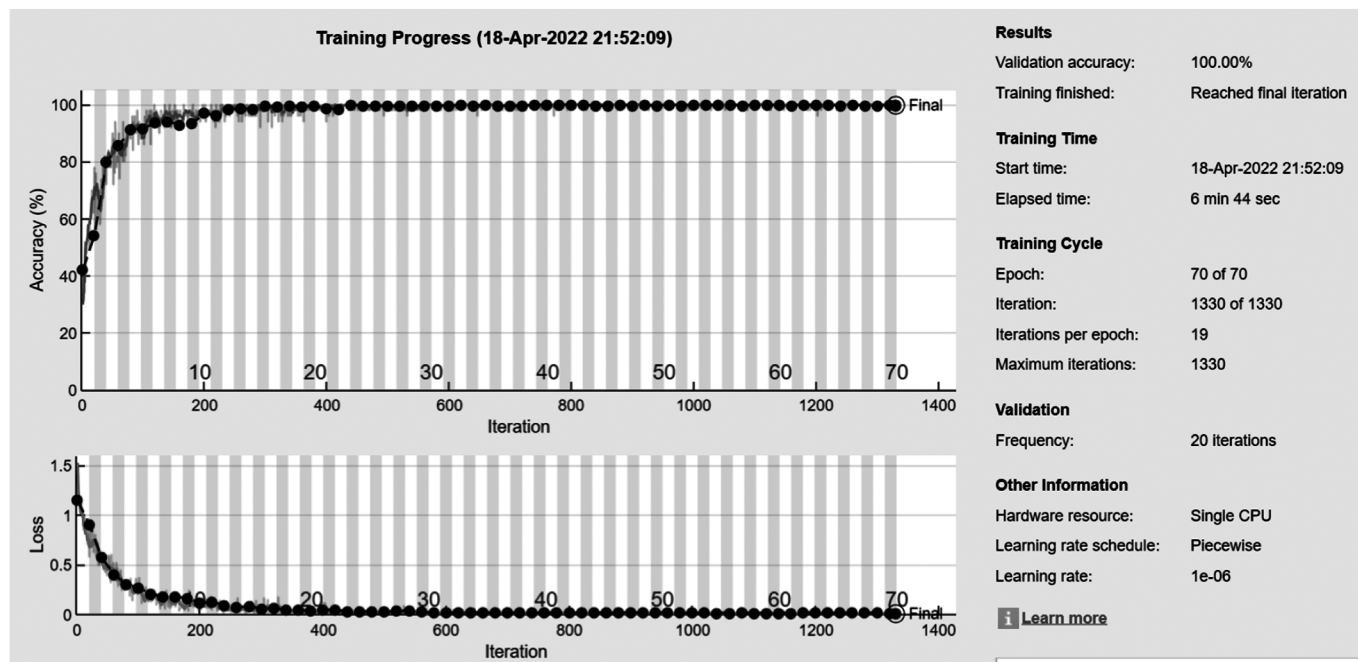


Рис. 5. Результат обучения модели нейронной сети для 3 эмоций

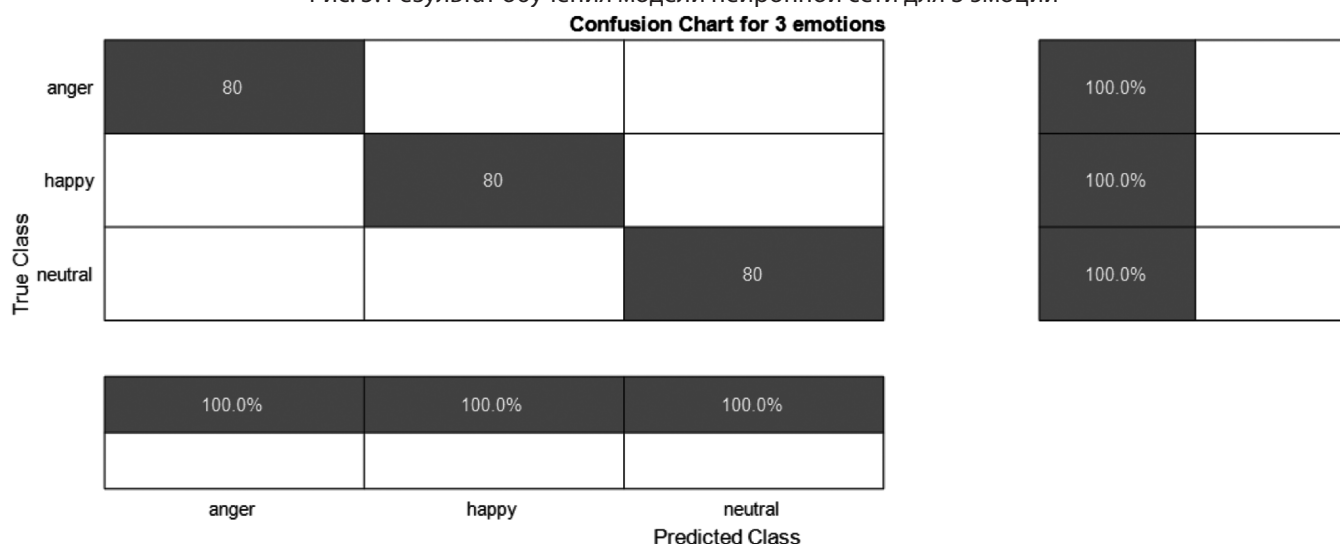


Рис. 6. Результат распознавания эмоций обученной нейронной сетью

нии: 80 % для обучения (320 аудиозаписей), 20 % для валидации (80 аудиозаписей).

В Matlab была разработана структура нейронной сети, состоящая из 24 слоёв (рис. 4).

Для обучения нейронной сети был использован алгоритм оптимизации Адама. В сравнении с остальными алгоритмами оптимизации (стохастический градиентный спуск и RMSProb), выбранный даёт максимальную точность и скорость обучения. Для тренировки нейронной сети, было задано 70 эпох обучения. В результате, процесс обучения был завершён за 6 минут и 44 секунды, причём точность обучения составила 100 % (рис. 5).

На рис. 6 представлены результаты тестирования нейронной сети для 3 эмоций.

Следующим этапом научного исследования являлось использование описанной выше структуры нейронной сети (см. рис. 4), но обучить её на распознавание большего количества эмоций. На рис. 7 представлен результат обучения для 7 эмоциональных состояний.

На обучение было затрачено около 37 минут и точность распознавания составляет 94.11%. На рис. 8 представлены результаты тестирования нейронной сети для 7 эмоций.

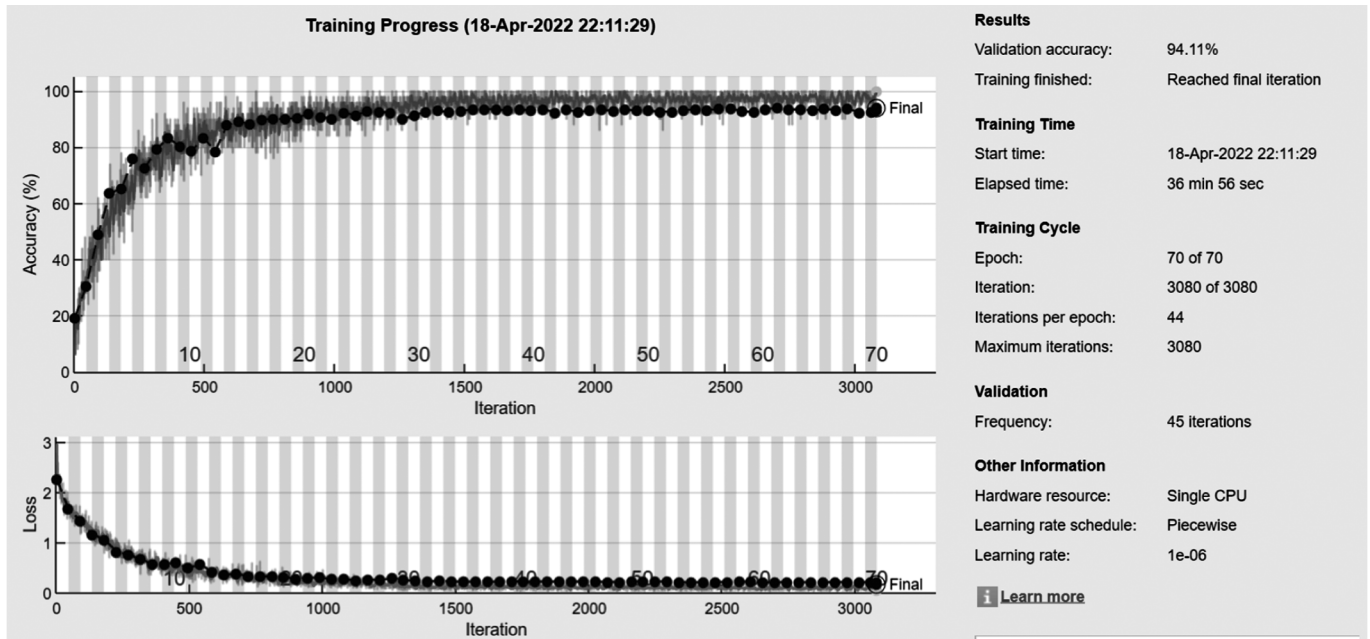


Рис. 7. Результат обучения модели нейронной сети для 7 эмоций

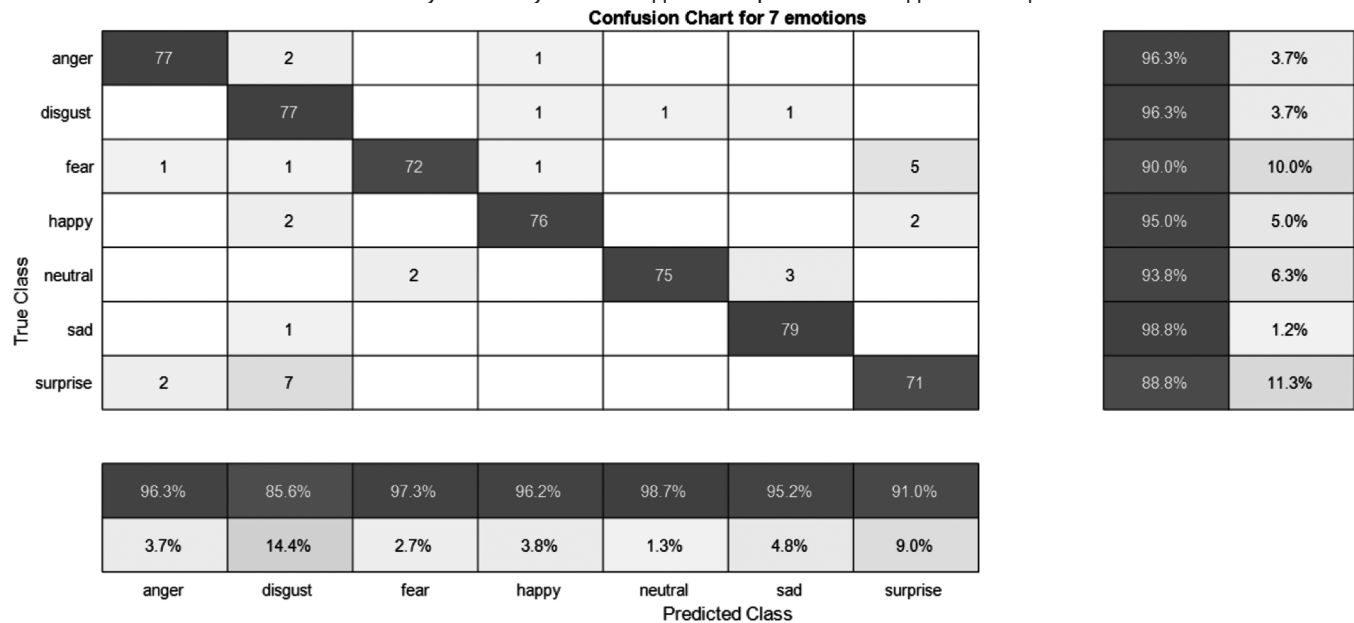


Рис. 8. Результат распознавания 7 эмоций

Заключение

В результате проведённого исследования и полученных результатов, видно, что алгоритм функционирующий и может применяться в различных сферах деятельности в соответствие с поставленной задачей.

Ввиду открытости, синергетические алгоритмы позволяют решать большое количество проблем за счёт возможности беспрепятственного «поглощения» информационных потоков извне, что в последующем влечёт за собой изменение внутренней структуры. Также, стоит отметить, что данные системы являются нелинейными, что позволяет существенно сузить порог нечувствитель-

ности разрабатываемой системы. Это в особенности позволяет решить вопрос нестабильности, а также, изменчивости информационных потоков.

При соблюдении определённых условий, имеется возможность объединения нескольких подобных подходов. Основная задача при таком объединении — поиск способа предотвращения коллизий при совместной их работе. Удачный синтез развивающихся систем можно обеспечить на основе закономерностей и принципов, которые лежат в основе процессов самоорганизации этих систем, а именно, за счёт диссипативного свойства описываемых алгоритмов.

В дальнейшем планируется проведение исследования, по созданию более универсального синергетического алгоритма, позволяющего снизить время обуче-

ния до минимума, при возможном увеличении точности создаваемого алгоритма.

ЛИТЕРАТУРА

1. Семенюк В.В. «Повышение качества компьютерного распознавания эмоционального состояния человека по голосу» // Научно-практическая конференция «Научные революции: Сущность и роль в развитии науки и техники» — НАУЧНО-ИЗДАТЕЛЬСКИЙ ЦЕНТР «АЭТЕРНА» Челябинск, 2021 — 142 с.
2. Клименко Н.С. Разработка структуры текстонезависимой системы идентификации диктора // Искусственный интеллект. 2012. № 4. С. 161–171.
3. Топ-5 сфер применения систем распознавания объектов [Электрон. ресурс] <https://habr.com/ru/company/toshibarus/blog/433544/> (Дата обращения: 10.03.2022).
4. Шиллер А.В. Выражение модернизируемых эффектов эмоций у искусственных агентов как визуальный язык // Праксема. Проблемы визуальной семиотики. 2019. № 4 (22). С. 223–243.
5. Складчиков М.В., Розкаряка П.И. Разработка алгоритма планирования траектории перемещения для роботизированной системы с использованием контроллера Kinect // Материалы 7-й международной научно-практической конференции «Инновационные перспективы Донбасса». Т. 2: Перспективы развития электротехнических, электромеханических и энергосберегающих систем. — Донецк: ДонНТУ. 2021. С. 65–77.
6. Рюмина Е.В., Карпов А.А. Аналитический обзор методов распознавания эмоций по выражениям лица человека // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 2. С. 163–176.
7. Карпичников А.П., Ляшева С.А., Шлеймович М.П. Обнаружение и сопровождение людей в интеллектуальных детекторах внештатных ситуаций // Вестник казанского технологического университета. 2014. С. 351–356.
8. Sucksmith E, Allison C, Baron-Cohen S, Chakrabarti B, Hoekstra R.A. Empathy and emotion recognition in people with autism, first-degree relatives, and controls // *Neuropsychologia*. 2013. V. 5. p. 98–105.
9. Ахметшин Р.И., Кирпичников А.П., Шлеймович М.П. Распознавание эмоций человека на изображениях // Вестник технологического университета. 2015. № 18 (11). С. 160–163.
10. Хрусталёв В.О., Зубков А.В. Разработка технологии распознавания эмоций человека на основе изображений, поступающих с веб-камеры // XXIV региональная конференция молодых учёных и исследователей волгоградской области. 2020. С. 223–224.
11. Курицкий В.Ю., Садов С.В. Нейросетевой алгоритм распознавания эмоций по изображению лица // Компьютерные технологии и анализ данных (СТДА'2020). 2020. С. 245–248.
12. Хнюнин М.В., Гпнелина Н.Д. Перспективы применения личных мобильных устройств для распознавания эмоций человека по фотографиям лица // Интеллектуальный потенциал Сибири. 2021. С. 179–182.
13. Артемьева Ж.Г., Крушная Н.А. Распознавание эмоций у других людей детьми с нарушениями опорно-двигательного аппарата // Человеческий фактор: социальный психолог. 2020. № 1 (39). С. 288–294.
14. Сидоров К.И., Филатова Н.Н. Автоматическое распознавание эмоций человека на основе реконструкций аттракторов образцов речи // Программные системы и вычислительные методы. 2012. № 1. С. 67–79.

© Семенюк Виктория Валерьевна (semenuk.viktoriya@gmail.com); Складчиков Максим Владимирович (maxsklad19981@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»