25. *Klevtsov S.I.* Opredelenie kharaktera izmeneniy parametra na osnove analiza dinamiki formy sovokupnosti ego znacheniy v real'nom vremeni [Determining the nature of parameter changes based on the analysis of the dynamics of the shape of a set of its values in real time], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences]. – 2020. – № 3. – S. 46-55.
26. *Klevtsov S.I.* Opredelenie momenta skachkoobraznogo izmeneniya bystroperemennoy fizicheskoy velichiny v real'nom vremeni s ispol'zovaniem diagramm Puankare [Determination of the moment of a sudden change in a rapidly varying physical quantity in real time using Poincare diagrams], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2012, No. 5, pp. 108-113.

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Кравченко.

Клевцов Сергей Иванович – Южный федеральный университет; e-mail: sergkmps@mail.ru; г. Таганрог, Россия; тел.. 88634328025; к.т.н.; доцент.

Klevtsov Sergey Ivanovich – Southern Federal University; e-mail: sergkmps@mail.ru; Taganrog, Russia; phone: +78634328025; cand. of eng. sc.; associate professor.

УДК 004.021 DOI 10.18522/2311-3103-2024-4-91-100

А.А. Егорчев, Д.М. Пашин, Н.А. Сарамбаев, А.Ф. Фахрутдинов

АНАЛИЗ СИСТЕМ ОПРЕДЕЛЕНИЯ И КЛАССИФИКАЦИИ ЭМОЦИЙ ЧЕЛОВЕКА ПО ДАННЫМ ЗВУКОВОГО ПОТОКА

В современной быстро меняющейся и требовательной рабочей среде способность быстро и точно оценить эмоциональное состояние сотрудника имеет решающее значение для защиты человеческих жизней и снижения материальных рисков. Эмоциональное благополучие играет важную роль в обеспечении безопасности на рабочем месте, производительности труда и общего психического здоровья. Поэтому разработка эффективных инструментов для мониторинга негативных эмоций и реагирования на них является актуальной задачей современности. Целью данного исследования является разработка алгоритма, способного классифицировать эмоции, используя аудиоданные, записанные смартфоном пользователя. Такой инструмент особенно полезен, если интегрирован в более широкую систему мониторинга здоровья, позволяющую оценивать показатели здоровья человека в режиме реального времени с помощью неинвазивных методов. В этой статье представлено новое решение, которое использует акустические сигналы, улавливаемые микрофоном смартфона, для обнаружения и классификации эмоций пользователя. Используя сверточные нейронные сети (CNNS), тип алгоритма глубокого обучения, известного своей эффективностью при обработке аудио- и визуальных данных, предлагаемая система может определять эмоциональное состояние пользователя. Модель CNN обучена распознавать признаки в аудиоданных, соответствующие различным эмоциональным проявлениям, фокусируясь на обнаружении негативных эмоций, таких как, гнев или печаль. Результаты исследования демонстрируют эффективность системы: частота ошибок при определении негативных эмоций составляет 19,5% для ложноположительных результатов (ошибки I рода) и 20,1% для ложноотрицательных результатов (ошибки II рода). Эти показатели указывают на ее потенциал для практического применения в реальных условиях. Внедряя это решение в существующие системы биомедицинского мониторинга, организации могут расширить свои возможности по мониторингу эмоционального благополучия сотрудников, потенциально предотвращая негативные последствия, такие как несчастные случаи на производстве или кризисы психического здоровья. Интеграция распознавания эмоций с помощью смартфонов в системы мониторинга состояния здоровья представляет собой значительный прогресс в области неинвазивного биомедицинского мониторинга, использующего повсеместное присутствие смартфонов и возможности машинного обучения.

Система неинвазивного мониторинга; машинное обучение; биомедицинский мониторинг; смартфонная сенсорика; анализ акустического сигнала; распознавание эмоций.

A.A. Egorchev, D.M. Pashin, N.A. Sarambaev, A.F. Fakhrutdinov

EMOTION DETECTION AND CLASSIFICATION SYSTEM BASED ON SOUND FLOW DATA

In today's rapidly changing and demanding work environment, the ability to quickly and accurately assess an employee's emotional state is crucial to protecting human lives and reducing material risks. Emotional well-being plays an important role in workplace safety, productivity, and overall mental health. Therefore, the development of effective tools for monitoring negative emotions and responding to them is an urgent task of our time. The purpose of this study is to develop an algorithm capable of classifying emotions using audio data recorded by a user's smartphone. Such a tool is especially useful if integrated into a broader health monitoring system that allows you to evaluate human health indicators in real time using non-invasive methods. This article presents a new solution that uses acoustic signals picked up by a smartphone microphone to detect and classify user emotions. Using convolutional neural networks (CNNS), a type of deep learning algorithm known for its effectiveness in processing audio and visual data, the proposed system can determine the user's emotional state. The CNN model is trained to recognize patterns in audio data corresponding to various emotional manifestations, focusing on detecting negative emotions such as anger or sadness. The results of the study demonstrate the effectiveness of the system: the error rate in determining negative emotions is 19.5% for false positive results (errors of the first kind) and 20.1% for false negative results (errors of the second kind). These indicators indicate its potential for practical application in real conditions. By integrating this solution into existing biomedical monitoring systems, organizations can expand their ability to monitor the emotional well-being of employees, potentially preventing negative consequences such as industrial accidents or mental health crises. The integration of emotion recognition using smartphones into health monitoring systems represents significant progress in the field of non-invasive biomedical monitoring, using the ubiquitous presence of smartphones and machine learning capabilities.

Noninvasive monitoring system; machine learning; biomedical monitoring; smartphone sensors; acoustic signal analysis; emotion recognition.

Введение. В современном развивающемся мире имеется огромное количество рабочих должностей для разного уровня специалистов. Очевидно, что стресс оказывает негативное влияние на деятельность работника, а в конечном счете на производительность компании. В бизнесе и менеджменте эмоциональный интеллект играет важную роль в формировании лидерства и взаимоотношений с коллегами и подчиненными. В образовании эмоции позволяют создавать позитивную образовательную среду и эмоционально поддерживать учащихся. В творческих профессиях эмоции напрямую влияют на производимый продукт. Более того негативные эмоции могут быть предвестниками конфликтов. Следовательно, существует необходимость в разработке автоматизированной системы, которая позволит своевременно выявлять людей с негативными эмоциями для снижения негативного их влияния на рабочую среду, а также оказания своевременной помощи, например, для предотвращения суицидов. В данной работе предложено решение на основе анализа алгоритмов определения эмоций из категории алгоритмов искусственного интеллекта, который служит для распознавания одной из трех эмоций: положительной, нейтральной или негативной в речевом сигнале. Данное решение предназначено для применения в рамках большой системы мониторинга состояния здоровья человека на предприятиях.

Основная часть. Тема классификации эмоций является довольно популярной и имеет множество работ, авторы которых предлагают свои подходы к определению. В основном решения базируются на использовании алгоритмов машинного обучения. Машинное обучении может использоваться для распознавания эмоций в аудиоданных следующим образом:

- Построение табличной структуры данных: строится таблица данных со значениями эмоций для каждой аудиозаписи.
- Предварительная обработка: используя инструменты OpenSMILE [1], извлекаются статистические признаки из аудиосигнала

- Обучение модели: с использованием CNN и построенной таблицы, модель обучается на распознавание эмоций.
- Тестирование модели: модель тестируется на способность распознавания эмоций с использованием тестовой выборки.
- Использование модели: обученная модель может использоваться для распознавания эмоций в новых аудиосигналах.

В решении Siddhant Mulajkar [2] автор предлагает модель распознавания стресса, основанную на глубоком обучении с использованием речевых сигналов. Предложенный алгоритм сначала извлекает мел-кепстральные коэффициенты из предварительно обработанных речевых данных, а после предсказывает состояние стресса, деля данные признаки на две группы (стресс/отсутствие стресса), опираясь на данных сверточной нейронной сети (CNN). Для построения данного решения использовался только один датасет (RAVDESS), так как весь алгоритм по извлечению признаков и построению табличной структуры данных строился, опираясь на идентификаторы аудиофайлов. Следовательно, для других датасетов, с другими идентификаторами этот алгоритм не подходит. Датасет RAVDESS наиболее распространенный и часто используется другими авторами. Опираясь на идентификаторы имен файлов, датасет разделён на табличную структуру данных, в отдельных столбцах которого: путь к файлу, источник, номер голоса, пол, интенсивность, повторение, эмоция. В датасете имеются записи как мужских, так и женских голосов, а так как они различаются по тембру и высоте голоса, автором было предложено решение классифицировать эмоции только для мужских голосов. После обучения модели, на основании тестовой выборки представлена матрица ошибок по 5-ти эмоциям (рис. 1). Данное решение было протестировано и получена матрица ошибок (рис. 2), не совпадающая с матрицей, заявленной автором. От данного решения было решено отказаться, так как для него подходит только один датасет. Для использования этого алгоритма оказалось необходимым полностью переписать решение, чтобы можно было использовать признаки с других датасетов.

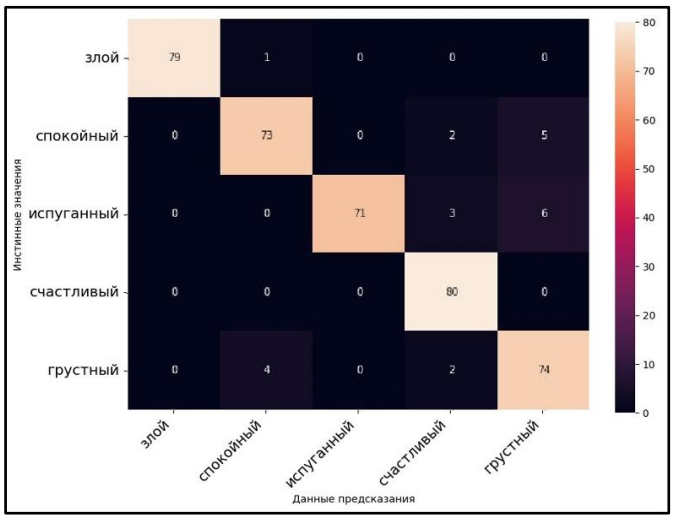

Рис. 1. Матрица ошибок для определения эмоций для лиц мужского пола для решения [2]

Решение на платформе Kaggle от Eu Jin Lok [3], состоит из пяти частей (разбиение на табличную структуру данных, вычисление признаков, построение модели, проверка на записанных данных, проверка на измененных данных). Данное решение написано на языке Python и в нём автором представлен классификатор эмоций по голосу человека. Решение было предложено с целью упростить работу колл-центров. На рис. 3 продемонстрирована матрица ошибок для данного решения.

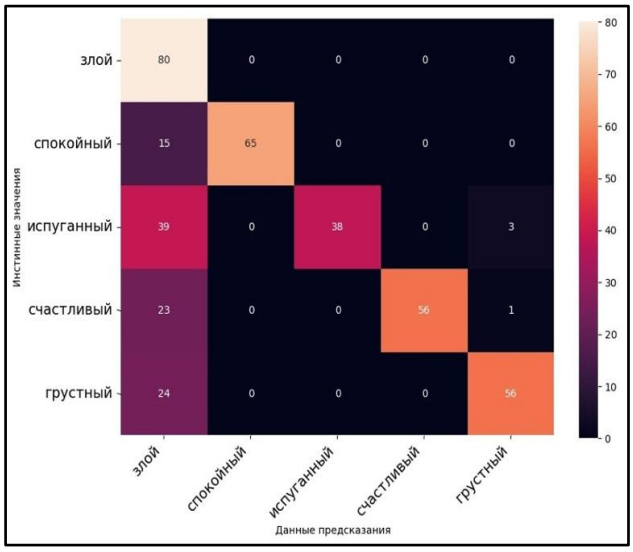

Рис. 2. Матрица ошибок для определения эмоций для лиц женского пола для решения [2]

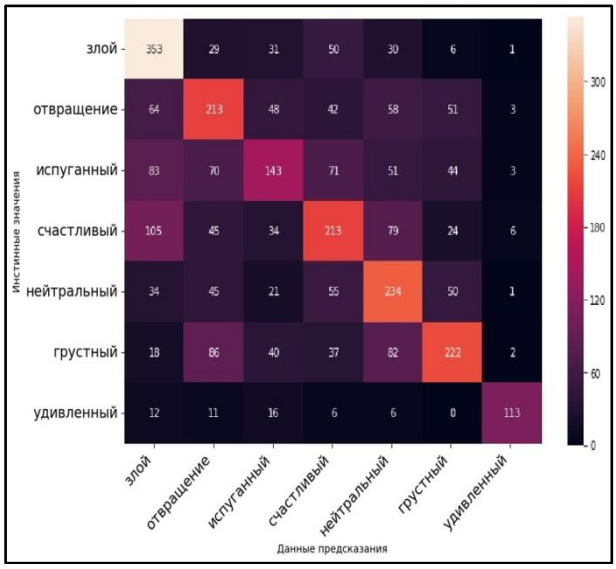

Рис. 3. Матрица ошибок для определения эмоций для лиц мужского и женского пола для решения [3]

Датасет в машинном обучении – это набор связанных между собой данных. Датасет может включать в себя текст, изображения, аудио и другие типы данных. Он может быть собран из различных источников, таких как базы данных, веб-сайты, социальные сети и т.д. Датасет может быть использован для обучения модели, а также для оценки ее качества и настройки параметров. Машинное обучение нацелено на создание систем, способных получать знания из данных, способных с помощью обучения улучшать показатели своей работы [4].

Для проверки готовых решений и обучения модели, был произведен поиск данных. Критерии, по которым были отобраны датасеты:

- Четкость аудиозаписей (без шума).
- Сбалансированность по количеству голосов (акцент, национальность).

- Объем датасета.
- Язык.

Были отобраны наиболее распространенные датасеты, которые находятся в открытом доступе:

- SAVEE [5]. Используемый датасет, составленный на базе записей четырех носителей английского языка, которые являются аспирантами и исследователями Университета Суррея в возрасте от 27 до 31 года, включает в себя семь видов эмоций: злость, отвращение, испуг, страх, счастье, печаль, удивление и нейтральное состояние.
- RAVDESS [6]. Данный датасет содержит 1440 звуковых файлов голосов 12 мужчин и 12 женщин, где для каждого человека имеется 60 записей. Файлы записаны для семи эмоциональных состояний человека: грусть, злость, радость, спокойствие, страх, удивление и отвращение. В каждой записи выражение воспроизводится на одном из двух уровнях эмоциональной интенсивности: нормальный и сильный.
- CREMA-D [7]. Этот датасет из 7442 клипов от 91-го человека. Клипы были сняты 48 актерами мужского пола и 43 актрисами женского пола, в возрасте от 20 до 73 лет. Для записи датасета люди говорили одно из двенадцати предложенных им предложений для шести различных эмоциональных состояний: гнев, радость, страх, отвращение, грусть и нейтральное, и 4 различных уровней эмоциональной интенсивности: низкий, средний, высокий и неопределенный.

Извлечение признаков в машинном обучении – это процесс преобразования исходных данных в набор признаков, который может быть использован в качестве входных данных модели машинного обучения. Процесс может включать в себя операции, такие как сжатие, нормализация или кодирование данных, а также использование различных алгоритмов для извлечения признаков, таких как преобразование Фурье или мел-кепстральное преобразование. Извлечение признаков является важной частью процесса машинного обучения, поскольку качество и количество признаков, используемых для обучения модели, может существенно влиять на точность и обобщающую способность модели.

Для этих целей нужно преобразовать их в нужный формат. MFCC (Мел-кепстральные коэффициенты) используются в области обработки звуковых сигналов, чтобы улучшить эффективность анализа и распознавания звуков. Они представляют собой признаки, которые отражают частотные характеристики звука, учитывая чувствительность человеческого уха к различным частотам. Они также позволяют уменьшить избыточность информации в звуковых сигналах и оптимизировать анализ звуковых признаков для задач распознавания эмоций. Человеческих слух воспринимает высоты звука не линейно, по отношению к частоте, и описывается следующим образом:

$$M(f) = 1125 \ln\left(1 + \frac{f}{700}\right). \quad (1)$$

График зависимости мел-шкалы от частоты представлен на рис. 4.

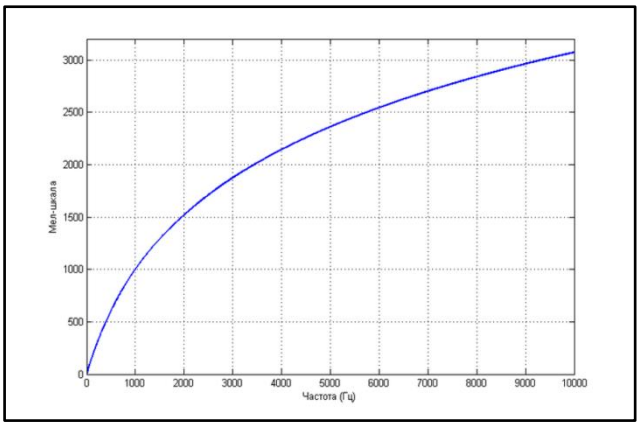

Рис. 4. График зависимости мел-шкалы от частоты

Подобные единицы измерения часто используют, так как они позволяют приблизиться к механизмам человеческого восприятия, которое пока что является лидирующей гипотезой в рамках известных систем распознавания речи [8].

Для извлечения звуковых признаков была использована библиотека openSMILE [1]. OpenSMILE (Open-Source Speech and Music Interpretation by Large-Space Extraction) является библиотекой, написанной на языке программирования C, предназначенная для извлечения признаков из аудио файлов. Используя данную библиотеку, можно извлечь такие признаки, как мел-кепстральные коэффициенты, коэффициенты энергии, аудиоспектральные коэффициенты и другие. Этот инструмент может быть использован в различных задачах аудио анализа, таких как распознавание эмоций или голосовой идентификации [9]. В рамках системы по распознаванию негативного и позитивного эмоционального состояния необходимо было разработать решение, способное классифицировать получаемые данные только на три класса, а существующие датасеты содержат от шести до восьми различных эмоций. Появилась необходимость в реорганизации найденных датасетов. Каждый из датасетов имеет свой идентификатор имени файла. В датасете SAVEE аудиофайлы названы таким образом, 'DC_d03.wav'. Первые 2 буквы префикса имени файла представляют инициалы говорящего. Буквы-префиксы описывают классы эмоций следующим образом:

- ‘a’ – anger(злость);
- ‘d’ – disgust(отвращение);
- ‘f’ – fear(страх);
- ‘h’ – happiness(радость);
- ‘n’ – neutral(нейтральное);
- ‘sa’ – sadness(грусть);
- ‘su’ – surprise(удивление).

Числа в конце названия файлов обозначают номер прочитанного предложения.

В датасете RAVDESS, каждый файл именовался следующим образом:

‘03-01-01-01-01-01-01.wav’, где каждый номер имеет следующие обозначения:

- Модальность (01 = аудио с видео, 02 = только видео, 03 = только аудио).
- Речевой канал (01 = речь, 02 = песня).
- Эмоция (01 = нейтральное, 02 = спокойный, 03 = радостный, 04 = грустный, 05 = злой, 06 = напуганный, 07 = раздражение, 08 = удивленный).
- Эмоциональная напряженность (01 = нормальное, 02 = сильное).
- Высказывание (01 = “Kids are talking by the door”, 02 = “Dogs are sitting by the door”).
- Повторение (01 = первое повторение, 02 = второе повторение).
- Актер (с 1-24, нечетные номера – мужчины, нечетные номера – женщины).

В датасете CREMA-D актеры и эмоции, как и все предыдущие датасеты, помечены в самом имени файла аудио. Пример файла из этого датасета: ‘1001_IEO_ANG_MD.wav’. Для данного датасета имеется файл, формата .csv, с идентификаторами распределения по эмоциям и актерам.

Датасеты были преобразованы в табличные структуры данных, которые были объединены для дальнейшего извлечения признаков.

Основная часть. Формула для вычисления мел-кепстральных коэффициентов состоит из нескольких шагов:

1. Преобразование Фурье для получения спектра частот.
2. Применение мел-фильтров для получения мел-спектра.
3. Вычисление логарифма мел-спектра.
4. Применение обратного преобразования Фурье для получения мел-кепстральных коэффициентов.

Мел-кепстральные коэффициенты часто используются для распознавания эмоций в аудио-сигнале, так как они могут дать информацию о частотной характеристике звука, которая может быть связана с эмоциональным состоянием человека. Например, низкие частоты могут быть связаны с тревожностью или гневом, в то время как высокие частоты

могут быть связаны с радостью или удовольствием. Использование мел-кепстральных коэффициентов позволяет извлечь эту информацию из аудио-сигнала и использовать ее для обучения модели распознавания эмоций.

CNN (Convolutional Neural Network – сверточные нейронные сети) модель для аудиосигнала обычно состоит из нескольких слоев:

- Слой входных данных: здесь аудиосигнал преобразуется в массив данных, который может быть обработан моделью.
- Слои свертки: это основа CNN модели. Они используются для извлечения признаков из входных данных.
- Слои пуллинга: это слои, которые используются для уменьшения размерности данных и уменьшения шума. Они также помогают увеличить обобщающую способность модели.
- Полносвязные слои: это слои, которые используются для классификации данных. Они используются для анализа извлеченных признаков и делают предсказание на основе этих признаков. Для распознавания эмоций по голосу часто используются нейронные сети с одним или несколькими сверточными слоями (CNN), которые способны извлекать важные признаки из аудиосигнала и передавать их в полносвязные слои для классификации.

С помощью TensorFlow Keras на языке Python с применением PyQt6 была разработана итоговая модель определения эмоций. На вход алгоритма подаются аудиоданные длительностью в 2.5 секунда и с частотой дискретизации 44100 Гц. В качестве датасетов использовались представленные ранее датасеты. Структура модели представлена на рис. 5.

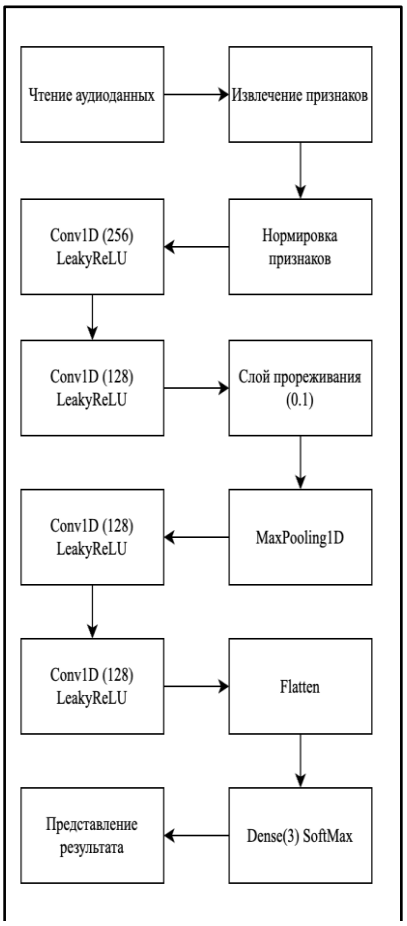

Рис. 5. Структура алгоритма определения эмоций

Результаты испытаний. Для определения качества разработанного решения проведены ряд испытаний. В качестве испытуемых приглашено 20 человек, которые являются студентами и сотрудниками Казанского Федерального Университета. Для испытаний использовалась клиент-серверная система мониторинга состояния здоровья человека. В качестве сервера выступал персональный компьютер Forsite(Intel® Core™ i9-9920X, NVIDIA Quadro RTX 6000; ОП 64 Гб; SSD 1Тб; клавиатура; мышь; монитор). Серверный персональный компьютер работал с СУБД PostgreSQL, в таблица которого выполнялись записи результатов измерений испытуемых посредством выполнения SQL запросов, получаемых от клиентского устройства. В качестве клиентских устройств использовались следующие модели смартфонов: Huawei nova 8i, Samsung A8 (2018), Xiaomi Mi 9 Lite, Redmi Note 9S, POCO X3 Pro, Samsung SM-A515F. Клиентские устройства, которыми пользовались испытуемые содержали установленное приложение. Каждый испытуемый проводил не менее 10 испытаний для проверки каждого типа классификации. Результаты испытаний представлены в табл. 1.

Таблица 1

Результаты испытания классификации эмоций.

Тип эмоции	Ошибки 1 рода	Ошибки 2 рода
Негативная	19.5 %	20.1 %
Положительная	22.4 %	31.5 %
Нейтральная	40.6 %	53.3 %

Обсуждение. В результате проведенных испытаний можно сделать вывод о том, что представленная модель позволяет практически с 20% ошибкой определять негативные эмоции. Это в принципе является более приоритетным классом для определения в рамках рассматриваемого применения решения. Однако является непригодной для определения других типов эмоций. Основными причинами таких результатов может являться небогатая обучающая выборка. Потенциально можно улучшить результаты, расширив количество рассматриваемых признаков, проведя балансировку обучающей выборки, а также увеличив размер ее размер.

Выводы. Рассмотрев тему классификации эмоций, разработан алгоритм, которая решает данную задачу используя алгоритмы машинного обучения, которую можно интегрировать в систему биомедицинского мониторинга. Испытания показали применимость полученной модели и алгоритма для определения негативных эмоций 19.5% ошибки первого рода и 20.1% ошибки второго рода. На тестовых записях положительных и нейтральных эмоций система показала большую ошибку, что делает решение непригодным для определения именно этих типов эмоций. Решение имеет потенциал для улучшения точности распознавания, путем увеличения размера обучающей выборки и расширения наборов признаков.

Благодарность. Данная работа выполнена в Казанском Федеральном Университете в рамках программы “Приоритет-2030”.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Opensmile 2.4.2 // PyPI - Указатель пакетов Python'a. – URL: https://pypi.org/project/opensmile/ (дата обращения: 12.12.2022).
2. Sentiment-Predictor-for-Stress-Detection-using-Voice // GitHub. – URL: https://github.com/sidmulajkar/sentiment-predictor-for-stress-detection (дата обращения: 10.10.2022).
3. Audio Emotion | Part 1 - Explore data // Kaggle. – URL: https://www.kaggle.com/code/ejlok1/audio-emotion-part-1-explore-data (дата обращения: 10.10.2022).
4. *Воронина В.В., Михеев А.В., Ярушкина Н.Г., Святов К.В.* Теория и практика машинного обучения: учеб. пособие. – Ульяновск: УлГТУ, 2017. – 290 с.
5. Surrey Audio-Visual Expressed Emotion (SAVEE) Database // The Centre for Vision, Speech and Signal Processing. – URL: http://kahlan.eps.surrey.ac.uk/savee/ (дата обращения: 29.11.2022).

6. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English // PLOS. – URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0196391 (дата обращения: 28.11.2022).
7. CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset // The National Center for Biotechnology Information. – URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313618/ (дата обращения: 30.11.2022).
8. Лекция 1. Первичный анализ речевых сигналов // Alpha Cephei speech Recognition. – URL: https://alphacephei.com/ru/research (дата обращения: 12.12.2022).
9. *Панфилов И.А., Алексеев М.С., Сивцова Е.И.* Извлечение признаков голосового корсета // Цифровая трансформация экономических систем: проблемы и перспективы (экопром-2022). – СПб.: ПОЛИТЕХ-ПРЕСС, 2022. – С. 794-796.
10. *El Ayadi M., Kamel M.S., Karray F.* Survey on speech emotion recognition: Features, classification schemes, and databases // Pattern recognition. – 2011. – Vol. 44, No. 3. – P. 572-587.
11. *Abbaschian B.J., Sierra-Sosa D., Elmaghraby A.* Deep learning techniques for speech emotion recognition, from databases to models // Sensors. – 2021. – Vol. 21, No. 4. – P. 1249.
12. *Swain M., Routray A., Kabisatpathy P.* Databases, features and classifiers for speech emotion recognition: a review // International Journal of Speech Technology. – 2018. – Vol. 21. – P. 93-120.
13. *Chen L., Su W., Feng Y., Wu M., She J., Hirota K.* Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction // Information Sciences. – 2020. – Vol. 509. – P. 150-163.
14. *Gokilavani M., Katakam, H. Basheer.* Ravdness, crema-d, tess based algorithm for emotion recognition using speech // 2022 4th International conference on smart systems and inventive technology (ICSSIT). – IEEE, 2022. – P. 1625-1631.
15. *Zielonka M., Piastowski A., Czyżewski A., Nadachowski P., Operlejn M., Kaczor K.* Recognition of emotions in speech using convolutional neural networks on different datasets // Electronics. – 2022. – Vol. 11, No. 22. – P. 3831.
16. *Xie Y., Liang R., Liang Z., Huang C., Zou C., Schuller B.* Speech emotion classification using attention-based LSTM // IEEE/ACM Transactions on Audio, Speech, and Language Processing. – 2019. – Vol. 27, No. 11. – P. 1675-1685.
17. *Zhao J., Mao X., Chen L.* Speech emotion recognition using deep 1D & 2D CNN LSTM networks // Biomedical signal processing and control. – 2019. – Vol. 47. – P. 312-323.
18. *Wang J., Xue M., Culhane R., Diao E., Ding J., Tarokh V.* Speech emotion recognition with dual-sequence LSTM architecture // ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2020. – P. 6474-6478.
19. *Huang Z., Dong M., Mao Q., Zhan Y.* Speech emotion recognition using CNN // Proceedings of the 22nd ACM international conference on Multimedia. – 2014. – P. 801-804.
20. *Anvarjon T., Mustaqeem, Kwon S.* Deep-net: A lightweight CNN-based speech emotion recognition system using deep frequency features // Sensors. – 2020. – Vol. 20, No. 18. – P. 5212.

REFERENCES

1. Opensmile 2.4.2, *PyPI - Ukazatel' paketov Python'a* [PyPI - Python Package Index]. Available at: https://pypi.org/project/opensmile/ (accessed 12 December 2022).
2. Sentiment-Predictor-for-Stress-Detection-using-Voice // GitHub. – URL: https://github.com/sidmulajkar/sentiment-predictor-for-stress-detection (accessed 10 October 2022).
3. Audio Emotion | Part 1 - Explore data, *Kaggle*. Available at: https://www.kaggle.com/code/ejlok1/audio-emotion-part-1-explore-data (accessed 10 October 2022).
4. *Voronina V.V., Mikheev A.V., Yarushkina N.G., Svyatov K.V.* Teoriya i praktika mashinnogo obucheniya: ucheb. posobie [Theory and practice of machine learning: a textbook]. Ul'yanovsk: UlGTU, 2017, 290 p.
5. Surrey Audio-Visual Expressed Emotion (SAVEE) Database, *The Centre for Vision, Speech and Signal Processing. – URL: http://kahlan.eps.surrey.ac.uk/savee/* (accessed 29 November 2022).
6. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English, *PLOS*. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0196391 (accessed 28 November 2022).
7. CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset, *The National Center for Biotechnology Information*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4313618/ (accessed 30 November 2022).
8. Lektsiya 1. Pervichnyy analiz rechevykh signalov [Primary analysis of speech signals], *Alpha Cephei speech Recognition,* Available at: https://alphacephei.com/ru/research (accessed 12 December 2022).

9. *Panfilov I.A., Alekseev M.S., Sivtsova E.I.* Izvlechenie priznakov golosovogo korseta [Extraction of signs of a vocal corset], *Tsifrovaya transformatsiya ekonomicheskikh sistem: problemy i perspektivy (ekoprom-2022)* [Digital transformation of economic systems: problems and prospects (ecoprom-2022)]. Saint Petersburg: POLITEKH-PRESS, 2022, pp. 794-796.
10. *El Ayadi M., Kamel M.S., Karray F.* Survey on speech emotion recognition: Features, classification schemes, and databases, *Pattern recognition,* 2011, Vol. 44, No. 3, pp. 572-587.
11. *Abbaschian B.J., Sierra-Sosa D., Elmaghraby A.* Deep learning techniques for speech emotion recognition, from databases to models, *Sensors,* 2021, Vol. 21, No. 4, pp. 1249.
12. *Swain M., Routray A., Kabisatpathy P.* Databases, features and classifiers for speech emotion recognition: a review, *International Journal of Speech Technology,* 2018, Vol. 21, pp. 93-120.
13. *Chen L., Su W., Feng Y., Wu M., She J., Hirota K.* Two-layer fuzzy multiple random forest for speech emotion recognition in human-robot interaction, *Information Sciences,* 2020, Vol. 509, pp. 150-163.
14. *Gokilavani M., Katakam, H. Basheer.* Ravdness, crema-d, tess based algorithm for emotion recognition using speech, *2022 4th International conference on smart systems and inventive technology (ICSSIT).* IEEE, 2022, pp. 1625-1631.
15. *Zielonka M., Piastowski A., Czyżewski A., Nadachowski P., Operlejn M., Kaczor K.* Recognition of emotions in speech using convolutional neural networks on different datasets, *Electronics,* 2022, Vol. 11, No. 22, pp. 3831.
16. *Xie Y., Liang R., Liang Z., Huang C., Zou C., Schuller B.* Speech emotion classification using attention-based LSTM, *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* 2019, Vol. 27, No. 11, pp. 1675-1685.
17. *Zhao J., Mao X., Chen L.* Speech emotion recognition using deep 1D & 2D CNN LSTM networks, *Biomedical signal processing and control,* 2019, Vol. 47, pp. 312-323.
18. *Wang J., Xue M., Culhane R., Diao E., Ding J., Tarokh V.* Speech emotion recognition with dual-sequence LSTM architecture, *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6474-6478.
19. *Huang Z., Dong M., Mao Q., Zhan Y.* Speech emotion recognition using CNN, *Proceedings of the 22nd ACM international conference on Multimedia,* 2014, pp. 801-804.
20. *Anvarjon T., Mustaqeem, Kwon S.* Deep-net: A lightweight CNN-based speech emotion recognition system using deep frequency features, *Sensors,* 2020, Vol. 20, No. 18, pp. 5212.

Статью рекомендовал к опубликованию д.т.н., профессор А.В. Боженюк.

Егорчев Антон Александрович – Казанский федеральный университет; e-mail: anton@egorchev.ru; Казань, Россия; директор института вычислительной математики и информационных технологий; к.т.н.

Пашин Дмитрий Михайлович – e-mail: dmitry.m.pashin@gmail.com; проректор по цифровой трансформации и инновационной деятельности; д.т.н.

Сарамбаев Никита Андреевич – e-mail: sarambaev@gmail.com; тел.: +79179029352; аспирант.

Фахрутдинов Адель Фердинандович – e-mail: timvaz@yandex.ru; тел.: +79872394153; аспирант.

Egorchev Anton Alexandrovich – Kazan Federal University; e-mail: anton@egorchev.ru; Kazan, Russia; director of Institute Computational Mathematics and IT; cand. of eng. sc.

Pashin Dmitry Mikhailovich – e-mail: dmitry.kfu@ya.ru; Vice-Rector for Digital Transformation and Innovation; dr. of eng. sc.

Sarambaev Nikita Andreevich – e-mail: sarambaev@gmail.com; phone: +79179029352; graduate student.

Fakhrutdinov Adel Ferdinandovich – e-mail: timvaz@yandex.ru; phone: +79872394153; graduate student.