Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.934.2+534.784

Краснопрошин
Даниил Вадимович

# Система автоматического распознавания эмоций по речевому сигналу на основе нейросетевой модели

**АВТОРЕФЕРАТ**
на соискание степени магистра
по специальности 1-40 80 01 «Компьютерная инженерия (встраиваемые системы)»

_____
*(подпись магистранта)*

Научный руководитель
Вашкевич Максим Иосифович
*(фамилия, имя, отчество)*
Доктор технчиеских наук, профессор
*(ученая степень, ученое звание)*

_____
*(подпись научного руководителя)*

Минск 2024

# ВВЕДЕНИЕ

В современном обществе, на фоне стремительного развития технологий, проникновение компьютеров и искусственного интеллекта в различные сферы жизни становится все более заметным и важным. Развитие технологий играет ключевую роль в преобразовании и улучшении многих аспектов нашей повседневной жизни, оказывая значительное влияние на работу, образование, медицину, социальные отношения и даже наше самочувствие.

В этом контексте, одним из наиболее интересных и перспективных направлений исследований становится разработка систем и технологий, способных распознавать и интерпретировать эмоциональные состояния человека. Эмоции играют ключевую роль в человеческой коммуникации, влияя на наше поведение, принятие решений, адаптацию к окружающей среде и общение с другими людьми. Понимание эмоций часто является неотъемлемой частью успешного взаимодействия и достижения целей как в личной, так и в профессиональной сферах жизни.

Важно отметить, что существуют два основных подхода к решению задачи распознавания эмоций: с использованием классических алгоритмов машинного обучения и на основе нейросетевых моделей.

Первый подход, основанный на классических методах машинного обучения, обычно включает в себя разработку и использование статистических моделей, основанных на характеристиках и признаках, извлеченных из речи. Эти модели могут включать в себя такие методы, как случайные леса (Random Forest), линейный дискриминантный анализ (ЛДА), метод опорных векторов (МОВ) и др. Хотя эти методы могут быть эффективными в некоторых случаях и имеют некоторые преимущества, они часто требуют тщательного отбора признаков и не всегда могут обеспечивать высокую точность и обобщающую способность на новых данных.

Второй подход, который стал широко распространенным в последние годы, связан с использованием нейросетевых моделей для распознавания эмоций в речи. Эти модели обычно могут обучаться на больших объемах данных, извлекая сложные и абстрактные зависимости между входными данными и целевыми эмоциями. Среди них наиболее распространены классические полносвязные нейронные сети прямого распространения, сверточные нейронные сети, рекуррентные нейронные сети, а также более современные архитектуры, такие как трансформеры. Нейросетевые модели обладают высокой гибкостью и могут автоматически извлекать признаки из входных данных, что делает их привлекательным выбором для решения сложных задач распознавания эмоций.

Изучение систем автоматического распознавания эмоций, в том числе на основе нейросетевых моделей, является актуальной и важной научной задачей. Обработка эмоциональных состояний имеет широкий спектр применений, включая, но не ограничиваясь, робототехникой, медициной, образованием, психологией и многим другим. Например, в робототехнике системы распознавания эмоций могут помочь роботам лучше взаимодействовать с людьми, делая их более дружелюбными и понятными в общении. В медицине эмоции могут быть использованы для диагностики и мониторинга психических расстройств или для анализа эмоционального состояния пациентов в реальном времени. В образовании это может помочь улучшить процесс обучения, адаптируясь к индивидуальным потребностям студентов и обеспечивая более эффективное обучение. В психологии обаботка эмоций может быть использована для исследования человеческого поведения, восприятия и реакции на различные стимулы.

Таким образом, системы распознавания эмоций на основе классических алгоритмов машинного обучения, а также с использованием нейросетевых моделей открывают широкие перспективы для применения в различных сферах жизни, предоставляя новые возможности для повышения эффективности, комфорта и качества человеческого существования.

На сегодняшний день существует большое количество исследований посвященных проблеме распознавания человеческих эмоций в речи. Однако эта задача представляет собой одну из наиболее сложных и актуальных проблем в области искусственного интеллекта и обработки естественного языка. Несмотря на значительные достижения в области машинного обучения и нейронных сетей, полное и точное распознавание эмоций из речевых сигналов до сих пор остается сложной задачей. Существующие системы могут столкнуться с различными вызовами, включая вариабельность интонации и выражения, амбивалентность эмоций, а также культурные и индивидуальные различия в интерпретации эмоций. Более того, эмоции могут проявляться в различных контекстах и быть субъективно восприняты по-разному, что делает задачу распознавания эмоций еще более сложной. В свете этого, исследования в области разработки более точных, надежных и адаптивных систем распознавания эмоций в речи остаются актуальными и требуют дальнейшего изучения и инноваций.

# ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования**

Работа тесно связана с приоритетными направлениями научных исследований, такими как обработка естественного языка, машинное обучение и искусственный интеллект. В частности, она вносит существенный вклад в разработку технологий распознавания эмоций, имеющих широкий спектр практических применений в области интерфейсов человеко-машинного взаимодействия.

**Цели и задачи исследования**

Целью магистерской диссертации является разработка эффективной системы распознавания эмоций в речи на основе нейросетевых моделей, а также изучение методов извлечения и анализа речевых признаков для классификации эмоций.

Для достижения цели исследования необходимо решить следующие задачи:

– провести анализ существующих моделей и подходов к рапознаванию эмоций в речи;

– изучить типы и особенности извлечения речевых признаков;

– реализовать, протестировать и оценить классификаторы на основе различных алгоритмов, в том числе не относящихся к нейросетевым (классические алгоритмы машинного обучения);

– провести сравнительный анализ эффективности классификаторов;

– изучить и реализовать возможность повышения качества распознавания эмоций с помощью сокращения признакового пространства на основе LASSO регрессии;

– оценить эффективность примененного метода по сокращению признакового пространства.

**Объектом** исследования являются процессы и методы автоматического распознавания эмоций с применением нейросетевых моделей.

**Предметом** исследования являются алгоритмы и модели, основанные на нейронных сетях, используемые для анализа и интерпретации эмоциональных состояний в различных контекстах.

**Научная новизна**

Научная новизна диссертационной работы состоит в разработке эффективной системы распознавания эмоций в речи на основе нейросетевых моделей, а также разработки метода предназначенного для повышения качества распознавания эмоций за счет сокращения признакового пространства на основе LASSO регрессии. Данный метод позволяет использовать более простые модели

машинного обучения, такие как метод опорных векторов и линейный дискриминантный анализ для построения классификатора, при этом классификатор становится хорошо интерпретируемым и более экономичным с точки зрения потребления вычислительных ресурсов.

**Основные положения выносимые на защиту**

1. Эффективность нейросетевых моделей в распознавании эмоций в речи: исследование показывает, что нейросетевые модели демонстрируют высокую эффективность в распознавании эмоций в речи.

2. Значимость извлечения речевых признаков для классификации эмоций: работа подтверждает важность извлечения и анализа речевых признаков при решении задачи распознавания эмоций в речи. Различные методы извлечения признаков, такие как мел-частотные кепстральные коэффициенты, играют ключевую роль в достижении высокой точности классификации.

3. Сравнительный анализ эффективности классификаторов: были определены наиболее эффективные подходы к распознаванию эмоций в речи.

4. Метод сокращения признакового пространства на основе LASSO регрессии, что позволяет повысить качество распознавания эмоций и сделать модели более эффективными и интерпретируемыми.

**Личный вклад соискателя**

Все вошедшие в диссертационную работу результаты были получены лично автором. Изложенные в данной диссертационной работе результаты основываются на исследованиях автора, проводимых на кафедре электронных вычислительных систем БГУИР. Архитектура программной части адаптирована под конкретную задачу лично автором. В ходе работы были реализованы все этапы исследования, начиная от анализа литературы и реализации моделей, и завершая сравнительным анализом эффективности классификаторов. Проект также включал оригинальное исследование методов сокращения признакового пространства, что значительно расширило область исследования и повысило его научную ценность.

**Опубликованность результатов диссертации**

Результаты диссертационной работы были апробированы на различных научных конференциях и опубликованы в сборниках статей и тезисов.

# КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе проведен обзор существующих методов и подходов к распознаванию эмоций в речи. Целью этого обзора является изучение как классических методов, таких как анализ акустических признаков и использование словарей эмоций, так и современных подходов, основанных на глубоком обучении и нейросетевых моделях.

В начале были рассмотрены классические методы распознавания эмоций в речи, в том числе анализ акустических признаков.

Затем были изучены современные подходы, основанные на глубоком обучении и нейросетевых моделях. Были представлены основные концепции и принципы работы глубоких нейронных сетей, а также описаны методы их применения к задаче распознавания эмоций в речи.

Следующим этапом был анализ сильных и слабых сторон каждого из рассмотренных методов. Преимущества классических методов включают их простоту и интерпретируемость, в то время как недостатками являются их ограниченная способность к обработке сложных структур и неоднородных данных. Современные подходы, такие как нейросетевые модели, обладают большей гибкостью и способностью к адаптации, но могут потребовать больших объемов данных и ресурсов для обучения.

В заключительной части главы были выделены перспективные направления для дальнейших исследований в области распознавания эмоций в речи. Это включает в себя разработку более эффективных и точных нейросетевых моделей, улучшение методов извлечения признаков и применение технологий глубокого обучения для анализа более сложных структур речевых данных.

В второй главе был проведен анализ алгоритма извлечения мел-частотных кепстральных коэффициентов (МЧКК) из речевых данных и их последующее использование для классификации эмоций.

Были изучены основные концепции и принципы МЧКК, включая их предназначение для представления спектральных характеристик звуков и их применение в области обработки речи. Далее были описаны методы извлечения МЧКК из аудиосигналов, включая предварительную обработку звуковых данных, вычисление спектрограммы и применение дискретного косинусного преобразования.

Расчет МЧКК относится к методам кратковременного анализа речевого сигнала, которые предполагают разбиение сигнала на фреймы (короткие сегменты). Считается, что в интервале от 10 до 30 мс голосовой сигнал можно считать стационарным. На рисунке 1 представлена схема вычисления МЧКК.

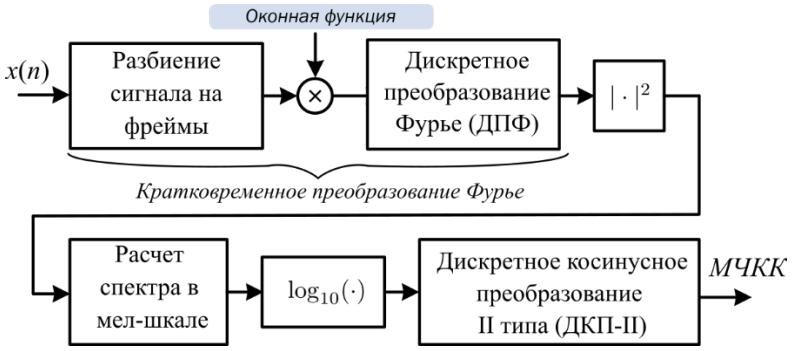



**Рисунок 1. Схема вычисления мел-частотных кепстральных коэффициентов**

Согласно рисунку 1, процесс извлечения МЧКК включает следующие шаги:

1) вычисление кратковременного преобразования Фурье (КВПФ) и нахождение квадрата модуля КВПФ для получения спектрограммы сигнала;

2) вычисление мел-спектрограммы (энергия сигнала из шкалы герц переводится в мел-шкалу, отражающую свойства человеческого слуха);

3) взятие логарифма от энергии сигнала в мел-частотных полосах;

4) применение декоррелирующего преобразования, в качестве которого используется дискретное косинусное преобразование II типа (ДКП-II).

В качестве иллюстрации на рисунке 2 показан пример речевого сигнала, выражающего эмоцию гнева.

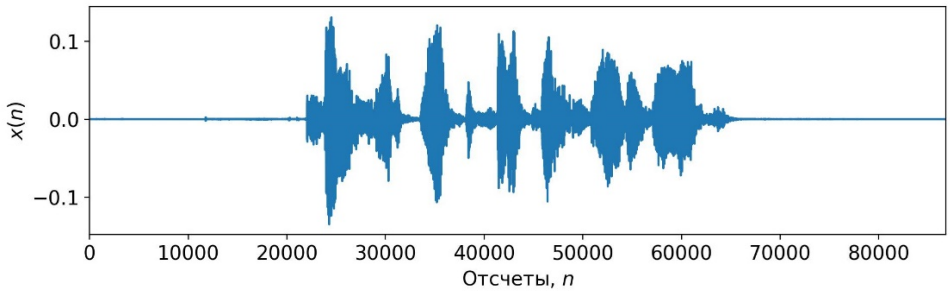
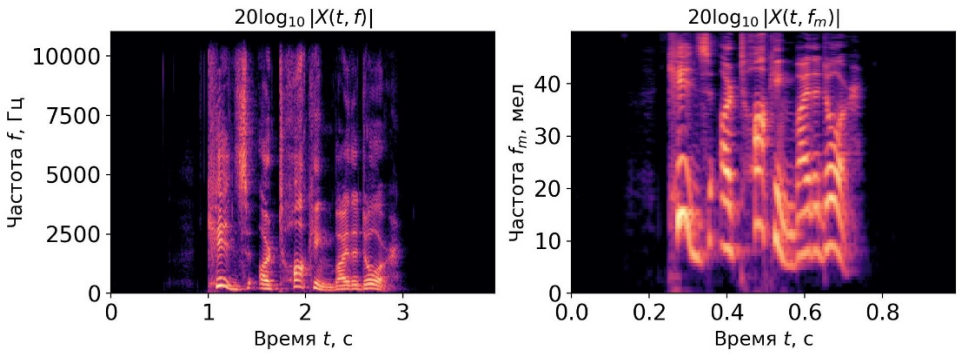



**Рисунок. 2. Представление речевого сигнала, выражающего гнев**

На рисунке 3 показан результат вычисления КВПФ и мел-спектрограммы сигнала, представленного на рисунке 1. На рисунке 4 изображена временная последовательность МЧКК, рассчитанная для сигнала на рисунке 2.




*a* *b*

**Рисунок 3. Речевой сигнал, выражающий гнев: *a* – спектрограмма кратковременного преобразования Фурье; *b* – мел-спектрограмма**

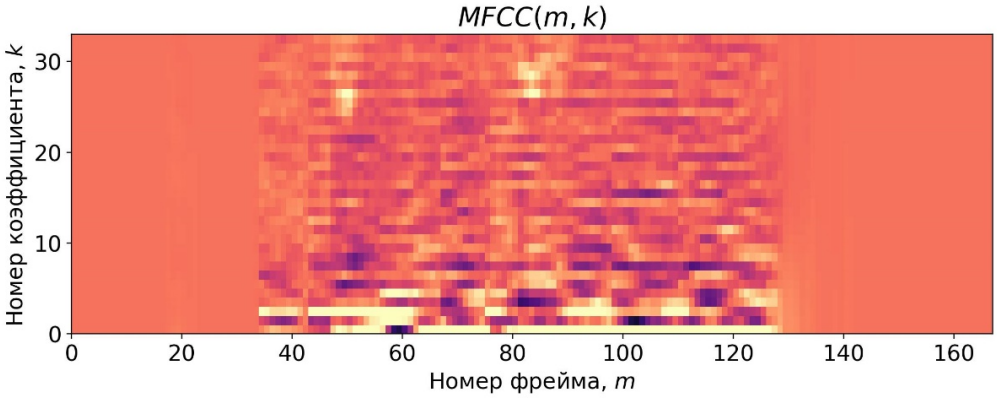



**Рисунок 4. Временная последовательность мел-частотных кепстральных коэффициентов**

Затем был проанализирован процесс использования МЧКК для классификации эмоций. Были рассмотрены различные аспекты этого процесса, включая методы нормализации признаков для обеспечения сопоставимости данных, выбор оптимальных параметров извлечения МЧКК и анализ влияния МЧКК на производительность классификаторов.

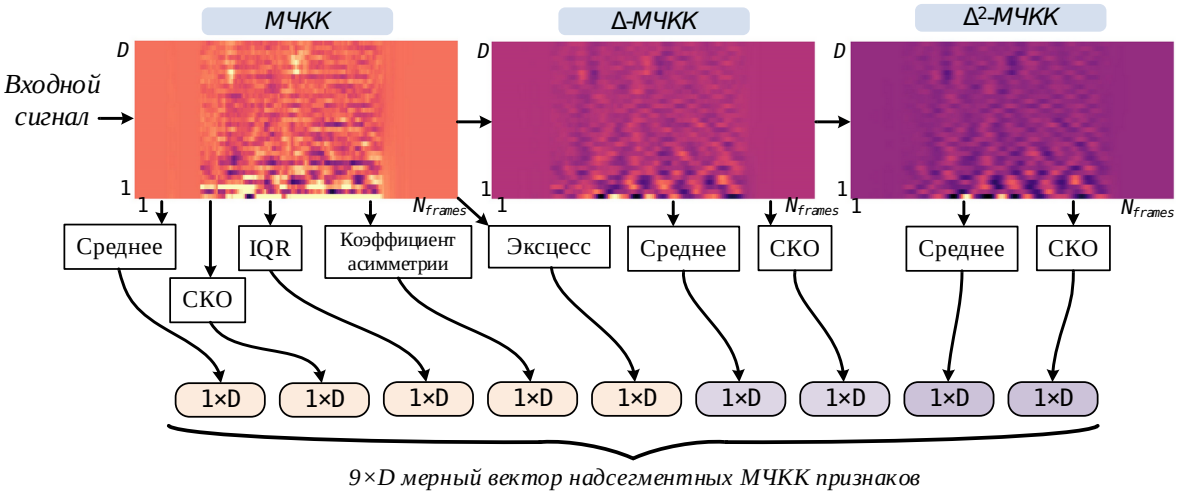



**Рисунок. 5. Схема формирования вектора признаков**

В третьей главе выполнена реализация и анализ классификатора для задачи распознавания эмоций в речи. Основная задача состояла в разработке эффективного классификатора, способного достаточно точно определять эмоциональное состояние говорящего на основе извлеченных признаков, включая мел-частотные кепстральные коэффициенты (МЧКК), их первые и вторые производные, а также другие релевантные характеристики.

Перым этапом была реализация классификатора на основе извлеченных из аудиосигнала признаков. Затем были проведены эксперименты по тренировке классификатора на обучающем наборе данных и последующей его оценке на тестовом наборе.

Далее был проведен сравнительный анализ эффективности классификаторов эмоций в речи на основе метода опорных векторов (МОВ), линейного дискриминантного анализа (ЛДА) и полносвязных нейронных сетей, которые были реализованы в предыдущих разделах этой главы.

В качестве исходного датасета использовался набор данных Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS).

Набор данных RAVDESS содержит записи 24 актеров, каждый из которых произносил 60 речевых высказываний с различными эмоциональными состояниями: нейтральность, спокойствие, счастье, грусть, гнев, страх, удивление и отвращение. Эмоциональные состояния озвучивались на двух уровнях эмоциональной громкости (нормальном и повышенном). В рамках экспериментов использовалась только часть базы RAVDESS, содержащая 1440 файлов в формате wav, то есть по 60 записей на каждого из 24 актеров.

Для каждого классификатора были проведены этапы обучения и тестирования на данном наборе данных. Для оценки производительности классификаторов использовалась метрика UAR. Далее в таблице 1 представлены результаты сравнения производительности различных классификаторов.

Таблица 1**.** Результирующий UAR для классификаторов на основе метода опорных векторов (RBF ядро), линейного дискриминантного анализа и полносвязной нейронной сети

| Метод классификации / Classification method | UAR |
|---|---|
| Метод опорных векторов (RBF ядро) | 0,482 |
| Линейный дискриминантный анализ | 0,460 |
| Полносвязная НН прямого распространения | **0,485** |

По результатам сравнительного анализа эффективности классификаторов эмоций в речи на основе метода опорных векторов (ядро на основе радиальной базисной функции (РБФ)), линейного дискриминантного анализа и полносвязных нейронных сетей прямого распространения, видно, что значения метрики UAR для всех трех методов находятся примерно на одном уровне.

При этом, важно отметить, что использование более простых моделей, таких как, например, линейный дискриминантный анализ, может иметь ряд преимуществ. Во-первых, такие модели обладают более высокой интерпретируемостью, что позволяет легче понять, какие признаки влияют на прогнозы модели. Это особенно важно в контексте задач, где необходимо понимать причинно-следственные связи между входными данными и прогнозами модели, например, в медицинских и психологических исследованиях.

Во-вторых, более простые модели зачастую требуют меньше вычислительных ресурсов для обучения и применения, что делает их более доступными и экономически эффективными для использования в реальных приложениях. Это особенно актуально в случаях, когда имеются ограничения по вычислительной мощности или бюджету проекта. Таким образом, при выборе модели классификации для задачи распознавания эмоций в речи необходимо учитывать как ее эффективность, так и другие факторы, такие как интерпретируемость и вычислительные расходы.

В четвертой главе было выполнено исследование и реализация метода повышения качества распознавания эмоций в речи с использованием метода по сокращению признакового пространства на основе LASSO регрессии.

Основной задачей было определить, насколько эффективным может быть применение данного метода для улучшения производительности классификатора в задаче распознавания эмоций.

Далее на рисунке 6 представлен процесс разработки системы распознавания эмоций с учетом использования метода понижения размерности признакового пространства.

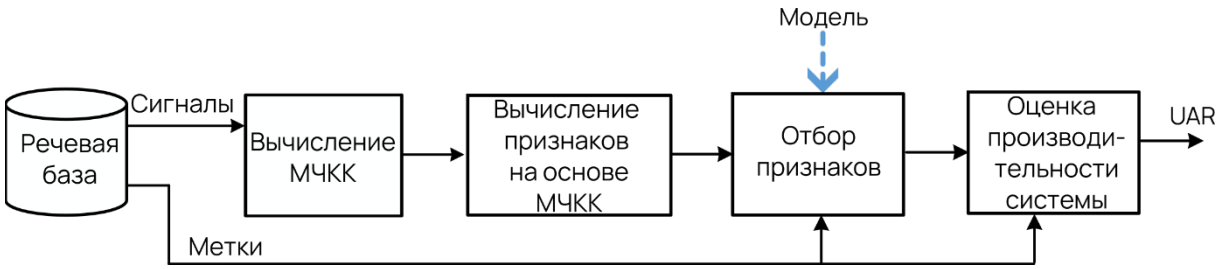



**Рисунок. 6. Процесс разработки системы распознавания эмоций по речи**

Результаты исследования показали, что применение метода LASSO регрессии позволило улучшить производительность классификатора по сравнению с использованием полного признакового пространства. Это подтверждает эффективность метода сокращения признакового пространства на основе LASSO регрессии в задаче распознавания эмоций в речи и может быть полезным для дальнейших исследований в этой области.

# ЗАКЛЮЧЕНИЕ

В рамках данной диссертационной работы были рассмотрены различные аспекты распознавания эмоций в речи с использованием нейросетевых моделей. Целью исследования было разработать эффективную систему распознавания эмоций в речи и изучить методы извлечения и анализа речевых признаков для классификации эмоций.

Для достижения цели исследования был поставлен и решен ряд задач. Во-первых, был проведен обширный анализ существующих моделей и подходов к распознаванию эмоций в речи. Это позволило понять текущее состояние этой научной области, а также определить перспективные направления для дальнейших исследований.

Во-вторых, были изучены типы и особенности извлечения речевых признаков. Это включало в себя анализ алгоритмов извлечения признаков, таких как мел-частотные кепстральные коэффициенты (МЧКК), и исследование их эффективности для классификации эмоций.

Далее были реализованы и протестированы классификаторы с использованием различных алгоритмов машинного обучения. В рамках данной работы были использованы метод опорных векторов, линейный дискриминантный анализ и полносвязные нейронные сети.

Для обучения, оценки и проведения сравнительного анализа эффективности классификаторов были использованы данные из набора Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS).

Наконец, была изучена и реализована возможность повышения качества распознавания эмоций с помощью сокращения признакового пространства на основе LASSO регрессии, что позволило улучшить производительность классификатора.

Таким образом, результаты данного исследования могут быть использованы в различных областях, таких как робототехника, медицина, образование и психология, где эффективное распознавание эмоций в речи имеет большое значение. Дальнейшие исследования могут включать в себя улучшение существующих моделей, а также исследование новых подходов и методов для более точного распознавания эмоций в речи.

# СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1 Краснопрошин Д. В. Распознавание эмоций с использованием кепстрального представления речевого сигнала и метода опорных векторов // 59-я научная конференция аспирантов, магистрантов и студентов БГУИР. – 2023.

2 Krasnoproshin D.V., Vashkevich M.I. Speech emotion recognition using SVM classifier with suprasegmental MFCC features // Pattern recognition and information processing. Minsk, Belarus – 2023.

3 Краснопрошин Д. В., Вашкевич М.И. Метод распознавания эмоций в речевом сигнале с использованием машины опорных векторов и надсегментных акустических признаков // Доклады БГУИР. – 2024.

4 Краснопрошин Д. В., Вашкевич М.И. Метод понижения размерности пространства признаков на основа LASSO-регрессии для задачи распознавания эмоций в речи // 26-я Международно-техническая конференция «Цифровая обработка сигналов и ее применение DSPA 2024». – 2024.

5 Краснопрошин Д. В. Применение полносвязных нейронных сетей в задачах распознавания эмоций в речи // 60-я научная конференция аспирантов, магистрантов и студентов БГУИР. – 2024.