

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

УДК 004.891 + 612.78

Жолуд
Елизавета Игоревна

Методы и алгоритмы акустического анализа голосового
сигнала для систем медицинской диагностики

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 7-06-0611-05 «Компьютерная инженерия»

Научный руководитель
Вашкевич Максим Иосифович
профессор, доктор технических наук

Минск 2025

ВВЕДЕНИЕ

Анализ акустических сигналов человеческого голоса имеет следующие цели:

1. Разработка технологий хранения, кодирования, передачи и синтеза речевых сигналов.
2. Изучение процесса речеобразования и восприятия речи через анализ речевого сигнала.
3. Получение информации о здоровье и самочувствии человека через анализ его голоса.

В рамках магистерской рассматриваются методы анализа и классификации голосового сигнала и в последствии разрабатывается приложение, реализующее систему диагностики заболеваний на основе анализа голосового сигнала. Приложение выполняет обработку голосового сигнала с помощью алгоритмов извлечения акустических характеристик. Приложение вычисляет значения джиттера, шиммера и отношения гармоник к шуму (HNR). Полученные значения джиттера, шиммера и HNR используются для предсказания о наличии или отсутствии заболевания у пациента. Результаты предсказания отображаются в приложении в удобной для пользователя форме.

Для точного акустического анализа голосового сигнала необходимо учитывать физиологические механизмы речеобразования, акустические свойства сигнала и современные вычислительные методы. В данной работе рассматриваются ключевые аспекты анализа голоса, начиная от базовых параметров (джиттер, шиммер, HNR) до сложных алгоритмов классификации (К-ближайших соседей, машины опорных векторов, алгоритм дерева решений, линейный дискриминантный анализ, наивный байесовский классификатор).

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Данная тема исследования актуальна по следующим причинам:

1. Ранняя диагностика заболеваний.

Для проведения эксперимента используется база голосов Minsk2020_ALS. База содержит записи голосов людей с боковым амиотрофическим склерозом (БАС), собранные в период 2019-2020 гг. в Минске, Беларусь. Боковой амиотрофический склероз (БАС) — это быстро прогрессирующее заболевание, при котором раннее выявление критически важно для замедления его развития. Акустический анализ голоса позволяет выявлять доклинические признаки БАС (например, изменения джиттера, шиммера, HNR) до появления выраженных двигательных нарушений. Это открывает возможность для своевременного начала терапии и улучшения прогноза пациентов.

2. Объективизация клинической оценки

Традиционные методы диагностики БАС (например, шкала ALSFRS-R) часто субъективны и зависят от опыта врача. Акустические алгоритмы представляют количественные метрики (вариация частоты, амплитуды, спектральные искажения), которые позволяют точно оценивать динамику заболевания и эффективность лечения.

3. Доступность

Для анализа голоса достаточно микрофона и программного обеспечения, что делает метод доступным для удаленных регионов и домашнего использования. Это особенно важно для пациентов с БАС, которым сложно регулярно посещать медицинские учреждения.

4. Мониторинг прогрессирования заболевания

Акустические параметры голоса (например, снижение индекса гармоник HNR) коррелируют с степенью поражения двигательных нейронов. Регулярный анализ речи позволяет отслеживать развитие болезни и адаптировать терапию в режиме реального времени.

5. Снижение нагрузки на систему здравоохранения

Автоматизированные системы способны обрабатывать тысячи голосовых записей, минимизируя ручной труд врачей. Это повышает эффективность и снижает затраты на диагностику.

6. Персонализированная медицина

Анализ голосовых паттернов помогает выявлять индивидуальные особенности течения БАС у разных пациентов. Это позволяет разрабатывать персонализированные схемы лечения, включая речевую терапию и нейрореабилитацию.

Цели и задачи исследования

Цель исследования – разработка и анализ методов обработки голосового сигнала для создания системы медицинской диагностики (на примере бокового амиотрофического склероза). В работе решаются следующие задачи:

1. Изучение акустических особенностей голосового сигнала (периодичность, субгармоники, модуляция, функции возмущения).
2. Выделение информативных параметров (джиттер, шиммер, HNR) и их связь с патологиями.
3. Сравнение алгоритмов классификации (k-NN, SVM, Decision Tree, LDA, наивный байесовский классификатор) для диагностики заболеваний.
4. Программная реализация системы с функциями обработки сигнала, вычисления акустических мер и классификации.

Степень разработанности проблемы

Исследования в области акустического анализа голоса для медицинской диагностики активно развиваются, однако остаются ключевые пробелы:

1. Неполное сравнение алгоритмов. Существующие работы часто фокусируются на отдельных методах, а не на выборе наилучшего метода в результате их сравнения.

2. Субъективность клинических методов. Традиционные подходы зависят от экспертной оценки, что снижает их воспроизводимость.

Данное исследование направлено на устранение этих недостатков:

– проведено комплексное сравнение классических и современных алгоритмов.

– разработаны методы, адаптированные для анализа голоса пациентов с БАС.

– предложена автоматизированная система, обеспечивающая объективную количественную оценку параметров.

Область исследования

Содержание диссертации соответствует образовательному стандарту высшего образования второй ступени (магистратуры) ОСВО 1-40 80 01-2019 специальности 1-40 80 01 Компьютерная инженерия (профилизация «Встраиваемые системы»).

Теоретическая и методологическая основа исследования

В основу диссертации легли работы белорусских и зарубежных учёных в следующих областях:

1. Цифровая обработка сигналов (работы И. Р. Титце).
2. Методы выделения акустических параметров голоса (джиттер, шиммер, HNR) на основе работ Р. Вендал, Э. Юмото, У. Дж. Гулд, Т. Бэр, А. Хаммуш.

3. Методы классификации – применение К-ближайших соседей (k-NN), метод опорных векторов (SVM), алгоритм дерева решений (Decision Tree), линейный дискриминантный анализ (LDA), наивный байесовский классификатор для классификации патологий.

Информационная база исследования включает:

– Открытые датасеты голосовых записей пациентов с БАС (база голосов Minsk2020_ALS. Режим доступа: https://github.com/Mak-Sim/Minsk2020_ALS_database).

– Программные инструменты: Python и его библиотеки.

Научная новизна, теоретическая и практическая значимость

Научная новизна заключается в:

1. Комплексном анализе алгоритмов акустического анализа голоса с акцентом на диагностику БАС, включая:

- сравнение алгоритмов классификации в контексте медицинской диагностики;
- разработку гибридного подхода, сочетающего выделение параметров джиттер, шиммер, HNR.

2. Адаптации методов обработки сигналов для работы с данными пациентов, включая подавление шумов.

Теоретическая значимость:

1. систематизация методов акустического анализа для медицинской диагностики;

2. углубление понимания связи акустических параметров голоса при БАС.

Практическая значимость:

1. разработка автоматизированной диагностической системы с функциями: предобработки голосовых записей, расчёта диагностических параметров (джиттер, шиммер, HNR), классификации патологий на основе наилучшего классификатора, который устанавливается экспериментально.

2. возможность применения в телемедицине для удалённого мониторинга пациентов.

Основные положения, выносимые на защиту

1. Результаты анализа, как длина КИХ-фильтра, используемого в ZC-методе, влияет на точность определения ЧОТ голосовых сигналов.

2. Оценка эффективности пяти различных классификаторов в задаче бинарной классификации состояния здоровья на основе акустических признаков, извлеченных из аудиозаписей.

3. Программная реализация системы медицинской диагностики на основе акустического анализа голосового сигнала.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе проведено детальное исследование методов анализа голосовых сигналов, применяемых в медицинской диагностике. Рассмотрены ключевые термины, математические модели, методы оценки возмущений, классификация сигналов. Основное внимание уделено анализу циклических параметров голоса, их модуляциям и применению в клинической практике.

Анализ возмущений основан на предположении, что в речевом сигнале всегда присутствуют небольшие колебания частоты, амплитуды и формы волны, отражающие внутренние «шумы» человеческого тела. Любая попытка говорящего воспроизвести абсолютно стабильный звук приводит к появлению апериодической формы волны. Движения тканей и воздушных потоков модулируются неравномерностью внутренних процессов, таких как распространение электрических импульсов, движение жидкостей и клеточная активность в речевом аппарате. Таким образом, то, что может показаться устойчивым движением в макроскопическом масштабе, часто оказывается пульсирующим движением в микроскопическом масштабе.

Голосовые связки являются основной колебательной системой, которая производит то, что можно назвать несущим сигналом (гортанный поток воздуха). Тогда все остальные генераторы можно рассматривать как модуляторы несущего сигнала. Некоторые модуляции близки к синусоидальным (дыхание, сердцебиение), другие же имеют более сложную структуру (потенциалы действия мышц, воздушные завихрения, движение слизи). Третьи являются пассивными осцилляторами (трахеальный резонатор, надгортанный голосовой тракт, различные синусы), которые могут влиять на первичную колебательную систему.

Анализ голоса позволяет разрабатывать технологии для записи, передачи и воспроизведения речи. Также он помогает изучать процессы, связанные с производством и восприятием речи. Например, анализ голосового сигнала позволяет понять, как работают наши голосовые связки и диафрагма при произношении звуков. Кроме того, голосовые сигналы могут использоваться для передачи информации по телефону или через интернет. И наконец, анализ голоса может помочь в диагностике заболеваний, например, если мы обнаруживаем изменения в звуке голоса, то это может указывать на проблемы со здоровьем легких или горла.

В науке о голосе существуют термины, описывающие изменения в голосе: возмущение — незначительное отклонение частоты и амплитуды; джиттер — небольшое отклонение основной частоты голоса от ожидаемого поведения; шиммер — случайное кратковременное возмущение амплитуды, которое может восприниматься как потрескивание или шум. Эти термины

помогают описывать речевые явления, хотя не имеют точных математических определений.

Ряд событий называется периодическим, если он повторяется с точным интервалом T_0 по формуле $f(t \pm nT_0) = f(t)$, при этом T_0 — фундаментальный период; однако в речевых сигналах строгая периодичность невозможна. Серия событий считается циклической, если повторяется определённый шаблон, не обязательно регулярный, а циклические параметры, например, период, амплитуда, частота описывают эти повторения и могут варьироваться во времени, образуя контуры, такие как контур основной частоты F_0 , который отражает реальные особенности голосового сигнала и его изменения.

Для удаления предсказуемых шаблонов в контуре циклических параметров применяется линейное предсказание, оставляющее в остатке лишь случайные вариации. Автокорреляционная функция, наоборот, выявляет долгосрочные закономерности, снижая влияние краткосрочного шума, помогая обнаружить невидимые вибраторы и особенности голоса, а гистограмма и спектр мощности показывают распределение и модуляции частот в параметрах голоса, демонстрируя, что возмущения обычно случайны и близки к нормальному распределению.

Меры возмущения, такие как стандартное отклонение (σ) и среднее абсолютное отклонение (δ) функции возмущения, количественно оценивают уровень вариаций циклических параметров голоса, однако они не отражают закономерности в отклонениях и чувствительны лишь к среднему значению. Для оценки общих акустической аperiодичности используют отношение гармоник к шуму (HNR), которое показывает соотношение энергии гармонического сигнала и шума в голосе, но не позволяет определить конкретные нарушения в периоде, амплитуде или форме волны. Новые методы корректируют влияние джиттера для более точного разделения источников шума, улучшая понимание природы голосовых возмущений.

Голосовые сигналы классифицируют на три типа: почти периодические с низкоэнергетическими субгармониками, сигналы с бифуркациями и субгармониками, энергию которых близка к основной частоте, и аperiодические сигналы без явной структуры. Спектrogramма и анализ основной частоты помогают выявлять бифуркации и классифицировать голосовые расстройства, например, при дисфонии, где амплитудные и частотные изменения отражают нестабильность фонации.

Клинические голосовые тесты призваны оценить влияние голоса на понимание речи и эффективность общения, а также состояние гортани и организма в целом; для этого используются задания разной сложности, от отдельных звуков до сложных фраз, позволяющие проверить контроль высоты,

громкости и качества голоса, включая дыхательные и мышечные функции, причём тесты адаптируются под индивидуальный голосовой диапазон.

Для точной регистрации акустических речевых сигналов необходимы профессиональный конденсаторный микрофон (всенаправленный или кардиоидный, чувствительность не хуже - 60 дБ), расстояние до микрофона менее 10 см (оптимально 3-4 см) при протяжном произнесении гласных для минимизации низкочастотных модуляций и высокого отношения сигнал/шум, размещение микрофона вне оси рта ($45\text{--}90^\circ$) для уменьшения аэродинамического шума, контроль реверберации и шумов помещения, 16-битный АЦП (отношение сигнал/шум 85-95 дБ), частота дискретизации 20-100 кГц, прямая связь по постоянному току, входы линейного уровня для подключения преобразователей без дополнительных предусилителей и звукоизолированное помещение (уровень шума < 50 дБ) с подавлением шумов на частоте 60 Гц и её гармониках.

Вторая глава посвящена углубленному анализу параметров голосового сигнала, необходимых для диагностики патологий. Начав с фундаментального параметра – частоты основного тона (F_0), глава детально рассматривает методы ее точного определения. Для этого исследуются алгоритмы сегментации сигнала на периоды основного тона (ОТ): метод подгонки формы сигнала (WM) и метод отбора локальных максимумов (РР), оба использующие предварительную грубую разметку, получаемую методом пересечения нуля (ZC). Эффективность ZC-метода, в свою очередь, анализируется через длину КИХ-фильтра, демонстрируя преимущества адаптивной длины фильтра, оптимизирующей подавление гармоник и повышающей точность. Далее, глава расширяет анализ, включая оценку нестабильности голоса с помощью параметров джиттера и шиммера, а также оценку чистоты сигнала через Harmonic-to-Noise Ratio (HNR). Завершается глава экспериментальным исследованием. Эксперимент проведен на базе 66 записей здоровых голосов (Minsk2020_ALS) исследовал влияние длины КИХ-фильтра в ZC-методе на точность определения частоты основного тона (ЧОТ), измеряемую джиттером ($<5\%$ — допустимая погрешность). Результаты показали минимальное количество ошибок (6-7) при длине фильтра 301 отсчета и при использовании адаптивного фильтра, характеристики которого оказались близки к фильтру длиной 301. Это подтверждает эффективность адаптивного фильтра для определения ЧОТ. Результаты эксперимента приведены на рисунке 1.

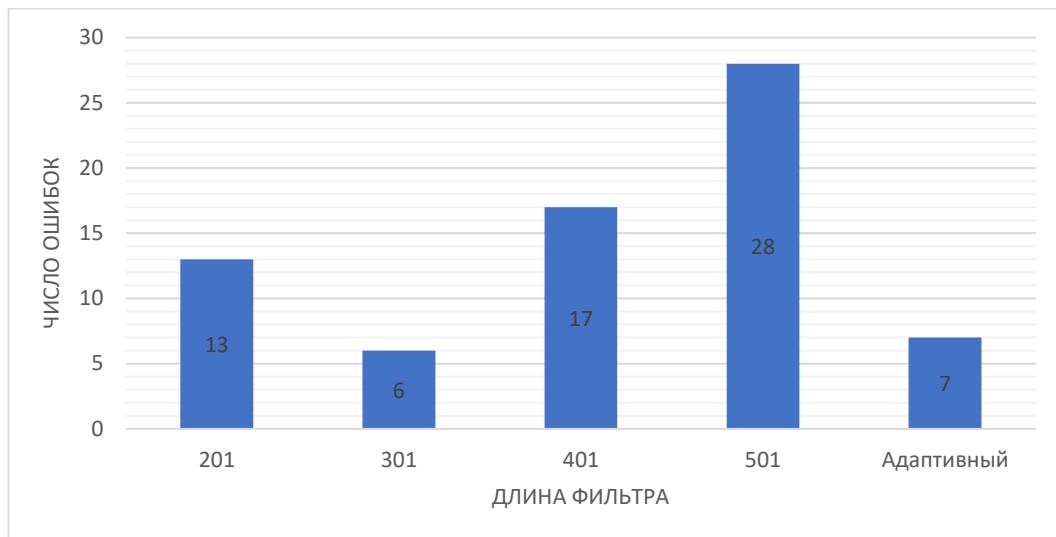


Рисунок 1 - График количества ошибок в зависимости от длины фильтра

Третья глава посвящена углубленному анализу классификаторов и эксперименту для поиска наилучшего классификатора для задачи бинарной классификации состояния здоровья человека (здоров или болен) на основе акустических признаков голоса.

Суть алгоритма К-ближайших соседей (KNN) -классификация/регрессия на основе "близости" объектов в пространстве признаков. Особенности алгоритма: нет фазы обучения, высокая вычислительная сложность предсказания, чувствителен к масштабу данных. Применение: простые задачи с небольшими данными.

Суть метода опорных векторов (SVM) - поиск оптимальной разделяющей гиперплоскости, максимизирующей зазор между классами. Особенности метода: эффективен в высокоразмерных пространствах, поддерживает нелинейные границы через ядра, требует масштабирования данных.

Суть дерева решений — это иерархическое разбиение данных на подмножества через последовательные условия. Особенности: интерпретируемость, работа с числовыми данными без масштабирования, склонность к переобучению. Применение: задачи, требующие прозрачности решений.

Суть линейного дискриминантного анализа (LDA) - проекция данных в пространство меньшей размерности для максимизации межклассового расстояния. Особенности анализа: предполагает нормальность распределения данных, ограничен по числу признаков (не более (число классов - 1)). Применение: понижение размерности и классификация линейно разделимых данных.

Суть наивного байесовского классификатора - предсказание класса на основе условных вероятностей признаков (с допущением их независимости). Особенности классификатора: высокая скорость, устойчивость к шуму,

проблемы с "нулевой частотой" признаков. Применение: текстовая классификация (спам-фильтры), высокоразмерные данные.

Сравнительный анализ представлен в таблице 1.

Таблица 1 - Сравнительный анализ классификаторов

| Критерий | K-NN | SVM | Дерево решений | LDA | Наивный Байес |
|-----------------------|-----------------------------|---------------------------------------|-----------------------|----------------------------|-------------------------|
| 1 | 2 | 3 | 4 | 5 | 6 |
| Скорость обучения | Нет обучения | Медленная (большие данные) | Быстрая | Быстрая | Очень быстрая |
| Скорость предсказания | Медленная | Быстрая | Быстрая | Быстрая | Очень быстрая |
| Интерпретируемость | Низкая | Средняя (линейное ядро) | Высокая | Средняя | Средняя |
| Устойчивость к шуму | Низкая (малое (k)) | Высокая (зависит от опорных векторов) | Низкая (переобучение) | Низкая | Высокая |
| Требования к данным | Масштабирование обязательно | Масштабирование обязательно | Без масштабирования | Гауссово распределение | Независимость признаков |
| Проклятие размерности | Сильно страдает | Устойчив | Устойчив | Ограничен числом признаков | Устойчив |

Также в третьей главе проведен эксперимент для определения наилучшего классификатора состояния здоровья человека (здоров или болен) на основе акустических признаков голоса. Результаты эксперимента приведены на рисунке 2.

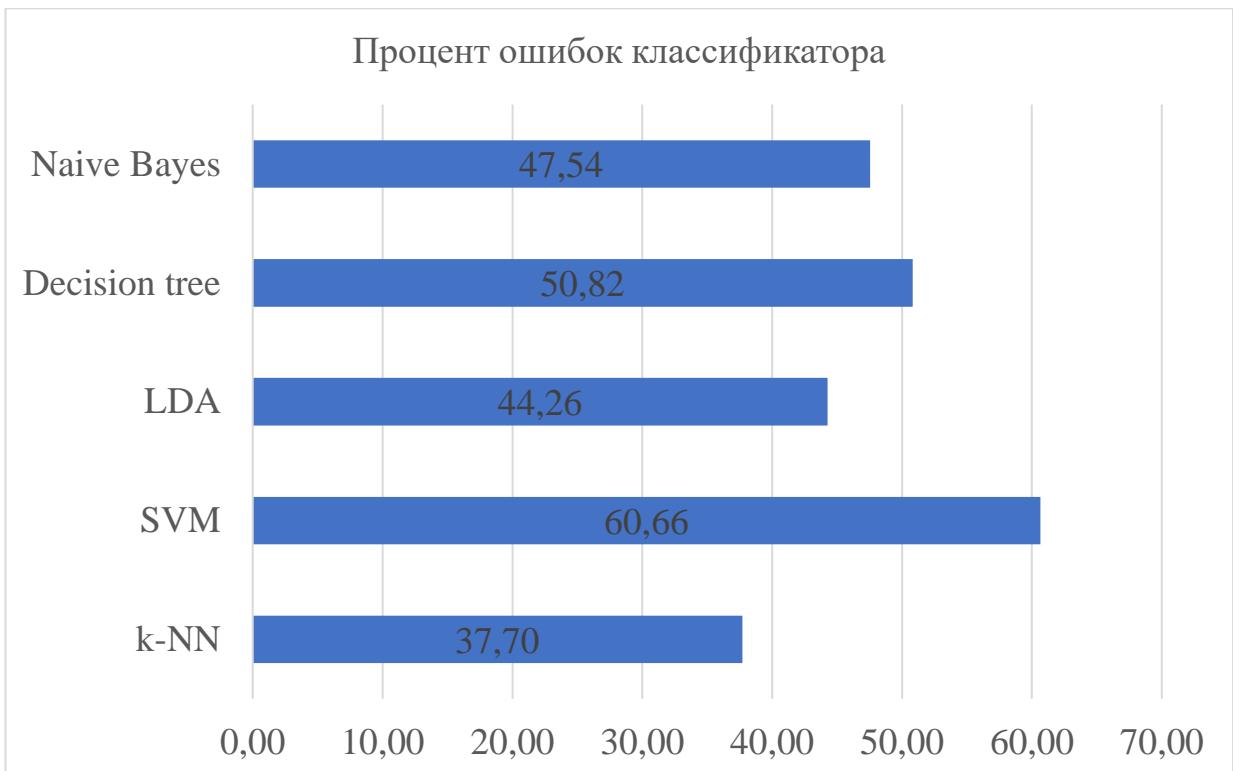


Рисунок 2 – Процент ошибок каждого классификатора

В четвертой главе описана подробная реализация программы и интерфейса для программы определения состояния здоровья. Интерфейс приведен на рисунке 3.

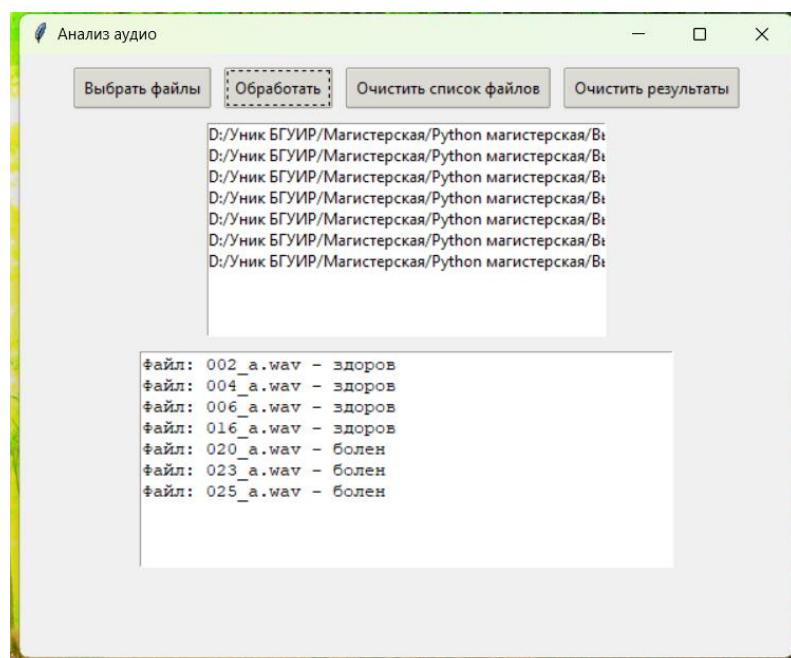


Рисунок 3 – Интерфейс приложения

ЗАКЛЮЧЕНИЕ

В ходе выполнения магистерской работы решена актуальная задача разработки системы диагностики заболеваний на основе акустического анализа голосового сигнала. Ключевые результаты и выводы исследования:

Установлено, что параметры джиттер (вариативность частоты), шиммер (вариативность амплитуды) и HNR (отношение гармоник к шуму) являются наиболее информативными для выявления патологий (например, БАС, неврологических расстройств).

Предложен адаптивный КИХ-фильтр для выделения периодов основного тона, где длина фильтра динамически определяется через период (T_0). Эксперимент на базе данных Minsk2020_ALS подтвердил эффективность метода: минимальное количество ошибок наблюдается при длине фильтра, равной 301, а также в условиях использования адаптивного фильтра. В этих случаях было зафиксировано всего 6 и 7 ошибок соответственно. На основании полученных результатов эксперимента можно также сделать вывод о том, что характеристики адаптивного фильтра в значительной степени соответствуют параметрам фильтра с длиной 301 отсчета. Это свидетельствует о том, что исследуемая база данных голосовых сигналов преимущественно включает записи голосов людей, чьи частотные характеристики наиболее эффективно обрабатываются с использованием фильтра длиной 301 отсчета.

Проведен сравнительный анализ пяти алгоритмов (k-NN, SVM, Decision Tree, LDA, Naive Bayes) и экспериментально было доказано: k-NN обеспечивает наименьший процент ошибок 37.7%, а SVM показал наихудший результат 60.66% ошибок.

Разработано приложение на Python с графическим интерфейсом, включающее: модули для вычисления джиттера, шиммера и HNR, обученную модель k-NN, функции загрузки аудио (.wav), обработки и сохранения результатов в Excel. И интуитивно простой и понятный интерфейс для пользователей приложения.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Жолуд, Е. И.; Вашкевич, М. И. // Метод сегментации голосового сигнала на периоды основного тона для систем медицинской диагностики / Белорусский государственный университет информатики и радиоэлектроники – Минск, 2024. – С. 144–148.
2. Жолуд, Е. И. // Методы предварительного анализа голосового сигнала для задач медицинской диагностики / Белорусский государственный университет информатики и радиоэлектроники – Минск, 2024. – С. 721–728.