

УДК 004.891

doi: 10.15622/rcai.2025.098

ПРОГРАММНАЯ ПЛАТФОРМА ОБНАРУЖЕНИЯ МОШЕННИЧЕСТВА В АУДИОЗАПИСЯХ С ПОМОЩЬЮ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ¹

Е.С. Мытарин (*e.mytarin@ulstu.ru*)

В.С. Мошкин (*v.moshkin@ulstu.ru*)

А.Ю. Журавлев (*zhuravlev.a@ulstu.ru*)

Ульяновский государственный технический университет,
Ульяновск

В статье представлено описание гибридной методики программная платформа для автоматизированного обнаружения мошенничества в аудиозаписях телефонных разговоров, основанная на нейросетевых моделях и онтологическом анализе. Реализующая предложенный подход программная платформа интегрирует методы глубокого обучения, такие как распознавание речи (Vosk, Whisper), лемматизацию текста, анализ интонации и сопоставление с онтологической базой данных для оценки вероятности мошеннических действий. Решение ориентировано на применение в сфере финансовой безопасности и противодействия социальной инженерии. Результаты проведенных экспериментов в сравнении с современными подходами работы демонстрируют эффективность платформы в обработке аудиоданных, визуализации результатов и интеграции с базами данных.

Ключевые слова: противодействие мошенничеству, нейронные сети, онтологии, распознавание речи, социальная инженерия.

Введение

Современные мошеннические схемы, включая телефонное мошенничество, характеризуются высокой адаптивностью и использованием психологических приемов, что затрудняет их своевременное выявление. Традиционные методы анализа аудиоданных зачастую не справляются с обработкой

¹ Работа выполнена в рамках государственного задания № 075-03-2023-143 по проекту «Исследование интеллектуального предиктивного мультимодального анализа больших данных, конструирование признаков гетерогенных динамических данных для машинного обучения».

нечетких речевых паттернов и требуют значительных вычислительных ресурсов. В этой связи актуальной задачей становится разработка интеллектуальных систем, способных автоматизировать процесс детекции мошенничества на основе машинного обучения и формализованных знаний [1].

Одним из перспективных направлений является применение нейросетевых моделей, позволяющих анализировать речевые данные с учетом контекста и семантики. В сочетании с онтологическим подходом, обеспечивающим структурированное представление экспертных знаний, такой метод повышает точность классификации подозрительных диалогов [2].

В данной работе рассматривается программная платформа, интегрирующая методы глубокого обучения и онтологический анализ для предиктивной оценки вероятности мошенничества в аудиозаписях. Решение ориентировано на практическое применение в сфере финансовой безопасности [3].

1. Сравнительный анализ методов и программного обеспечения детектирования мошенничества в аудиозаписях

а. Программные аналоги

В настоящее время для решения задачи выявления мошеннических звонков применяются различные специализированные программные решения. Среди существующих разработок можно выделить несколько ключевых продуктов.

IBM Watson Speech to Text представляет собой платформу для преобразования речевых данных в текстовый формат с возможностью последующего семантического анализа. Система обеспечивает высокую точность распознавания речи, однако не содержит специализированных функций для детекции мошеннических схем [4].

Google Cloud Speech-to-Text – облачное решение для обработки аудиопотоков, поддерживающее интеграцию с NLP-алгоритмами. Несмотря на широкие возможности обработки естественного языка, платформа не предлагает готовых механизмов для выявления мошеннических сценариев [5].

Pindrop – специализированная система для обнаружения телефонного мошенничества, использующая анализ голосовых биометрических параметров и поведенческих паттернов. Решение демонстрирует высокую эффективность, но требует значительных вычислительных ресурсов и ориентировано преимущественно на корпоративный сегмент [6].

Nice Perform – программный комплекс для мониторинга звонков в колл-центрах, включающий функции анализа подозрительных диалогов. Система обладает развитыми возможностями машинного обучения, однако не поддерживает работу с пользовательскими онтологиями [7].

OpenVoice OS – открытая платформа для обработки голосовых данных с возможностью подключения сторонних аналитических модулей. Гибкость архитектуры компенсируется отсутствием предустановленных алгоритмов детекции мошенничества [8].

Каждый из рассмотренных инструментов обладает уникальным набором функций и может быть эффективен в определенных сценариях работы с аудиоданными. Однако ключевым ограничением существующих решений является отсутствие комплексного подхода, сочетающего прогностический анализ на основе нейросетевых моделей с гибкой системой настраиваемых онтологий для выявления мошеннических схем. Большинство представленных систем функционируют исключительно в режиме реактивного анализа, не обеспечивая предиктивной оценки потенциальных угроз.

b. Методы анализа

Для решения задачи выявления мошеннических звонков применяются различные программные решения, каждое из которых обладает уникальными характеристиками. Ниже представлен обзор ключевых методов и их параметров, основанный на анализе современных исследований. Основные критерии оценки эффективности применяемых методов: точность, скорость, интерпретируемость, устойчивость к шуму, потребление памяти [9].

Методы на основе NLP

- Точность – очень высокая (98.53%).
- Скорость – высокая (1-5 секунд на обработку записи).
- Интерпретируемость – средняя (правила анализа могут быть понятны, но сложность моделей затрудняет полное объяснение).
- Устойчивость к шуму – высокая (способность фильтровать шумы, но зависимость от качества распознавания речи).
- Потребление памяти – низкое (локальная обработка данных, 50-150 МБ) [10].

RAG-модели на основе LLM

- Точность – очень высокая (97.98%).
- Скорость – средняя (3-15 секунд, включая транскрипцию, поиск в базе знаний и генерацию ответа).
- Интерпретируемость – высокая (предоставляет обоснования решений).
- Устойчивость к шуму - высокая (использование cosine similarity для учета ошибок транскрипции).
- Потребление памяти – высокое (из-за использования больших языковых моделей, 4-8 ГБ) [11].

Гибридные системы (Pindrop)

- Точность – высокая (до 90%).
- Скорость – средне-высокая (зависит от сложности анализа, 1.5-15 секунд).

- Интерпретируемость – средняя (комбинация методов может усложнить объяснение).
- Устойчивость к шуму – высокая (использование акустических свойств голоса).
- Потребление памяти – средняя (из-за комплексного анализа, 2.5-4 ГБ) [12].

Методы на основе рекуррентных нейронных сетей (RNN)

- Точность – высокая (93.79%).
- Скорость – низкая (60–600 секунд).
- Интерпретируемость – низкая (затруднено понимание временных зависимостей).
- Устойчивость к шуму – высокая (эффективность для временных рядов).
- Потребление памяти – высокое (1–8 ГБ) [13].

Генеративно-сопоставительные сети (GAN)

- Точность – высокая (88-92%).
- Скорость – низкая (требуется время для обучения и анализа, 30-180 секунд).
- Интерпретируемость – низкая (сложность архитектуры GAN).
- Устойчивость к шуму – средняя (чувствительность к качеству данных).
- Потребление памяти – очень высокое (3–15 ГБ) [14]

В связи с тем, что все рассмотренные подходы не показывают максимальную точность при решении подобной задачи, актуальной задачей является разработка эффективного подхода для предиктивного выявления мошенничества в аудиозаписях, интегрирующей современные методы нейросетевого анализа и обработки естественного языка с учетом требований к скорости обработки и точности детекции, а также его реализация в виде программной платформы

2. Гибридная методика определения мошенничества посредством анализа аудиофайлов телефонных разговоров

Для решения поставленной задачи был разработан подход к детектированию мошенничества в аудиозаписях телефонных разговоров.

Блок-схема предложенной методики анализа содержания голосовых аудиофайлов включает представлена на рис. 1.

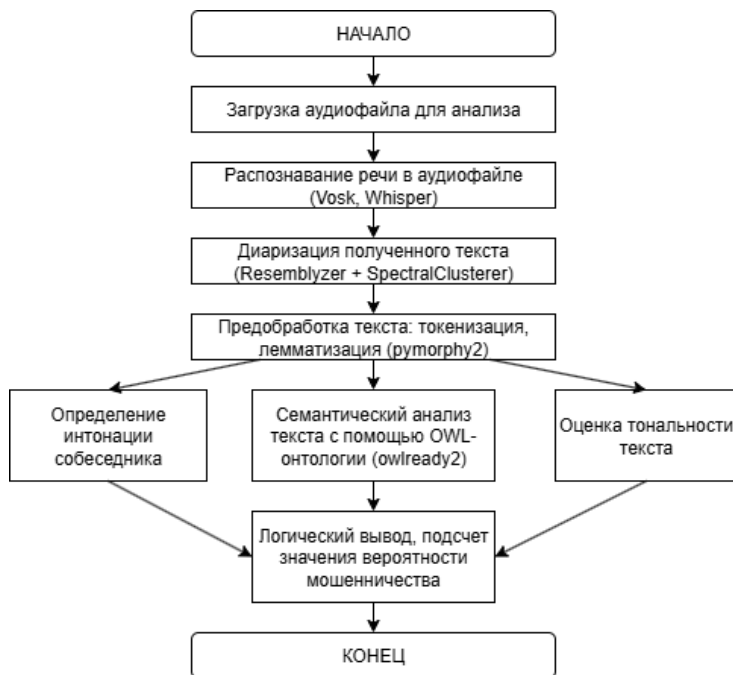


Рис. 1. Блок-схема предложенной методики анализа содержания голосовых аудиофайлов телефонных разговоров

Рассмотрим наиболее важные элементы предложенного подхода:

1. Распознавание речи реализовано с использованием моделей Vosk (на базе Kaldi) и Whisper (от OpenAI):

- *Vosk* – это инструмент для офлайн-распознавания речи, построенный поверх фреймворка Kaldi, одного из самых известных проектов в области ASR (Automatic Speech Recognition). Он использует гибридную архитектуру HMM + DNN, сочетающую статистическое моделирование с нейросетевыми признаками.
- *Whisper* – это открытая нейросетевая модель, разработанная OpenAI, основанная на архитектуре трансформеров. Она относится к классу end-to-end моделей, обученных на огромном объеме мульти-язычных данных, включая шумные и реальные аудиозаписи. Модель, по которой будет произведено распознавание речи, может быть выбрана пользователем в интерфейсе платформы.

2. **Диаризация аудио** – это сегментация аудио по говорящим с целью анализа текста только одного говорящего (предположительного мошенника).
3. **Предобработка текста** включает токенизацию и лемматизацию.
4. **Семантический анализ** предобработанного текста речи собеседника с использованием разработанной OWL-онтологии.

OWL-онтология – это одна из форм графового-семантического представления знаний, реализующая дескрипционную логику 2-ого порядка. Используемая в рамках данного исследования лингвистическая OWL-онтология имеющую следующую структуру:

$$O_F = (T_F, P_F),$$

где T_F – множество объектов классов онтологии, представляющих собой ключевые термины, которые используются при реализации схем телефонного мошенничества с целью кражи денежных средств.

Например, к подобным терминам относятся словосочетания «*безопасный счет*», «*служба безопасности*», «*подозрительная активность*» и пр.

P_F – множество свойств объектов:

где – множество бинарных лингвистических свойств терминов (ObjectProperties). Например, синонимия, антонимия, паронимия и пр.

– свойство типа данных (DatatypeProperty), отражающее вероятность принадлежности термина к множеству терминов, употребляемых в телефонных разговорах мошенников,

В рамках данного исследования набор терминов и отношения между ними формировались специалистом самостоятельно на основе опыта анализа мошеннических телефонных звонков.

В дальнейшем планируется автоматизация процесса формирования этой лингвистической OWL-онтологии путем обработки текстов записей разговоров мошенников и извлечения наиболее аутентичных терминов.

5. **Определение интонации** собеседника реализовано посредством анализа аудиосигнала (среднее значение амплитуды).
6. **Оценка тональности текста** – это классификация текста с применением модели BERT (используемые классы тональности: положительный, отрицательный и нейтральный).
7. **Логический вывод** результата анализа включает:
 - 1) процент вероятности мошенничества;
 - 2) число найденных подозрительных слов в тексте;
 - 3) результат вычисленной интонации.

Таким образом, научной новизной предложенного решения является гибридизация современных нейросетевых моделей, моделей онтологического анализа и NLP в задачах обнаружения мошенничества в аудиозаписях телефонных разговоров.

3. Архитектура программной платформы

Разработанная программная платформа, реализующая предложенную методику, состоит из следующих программных модулей:

- Модуль обработки аудио преобразует записи в формат WAV и выполняет диаризацию речи.
- Модуль распознавания речи преобразует аудио в текст путем использования моделей Vosk и Whisper.
- Модуль обработки текста выполняет лемматизацию и анализ интонационных особенностей речи с применением библиотеки rumorphy2.
- Модуль работы с онтологией обеспечивает хранение и обработку данных с использованием библиотеки owlready2.
- Модуль хранения данных организует работу с базой данных на основе SQLite.
- Модуль визуализации отображает аналитические данные с помощью библиотеки matplotlib.

На рис. 2 представлена диаграмма компонентов разработанной программной платформы.

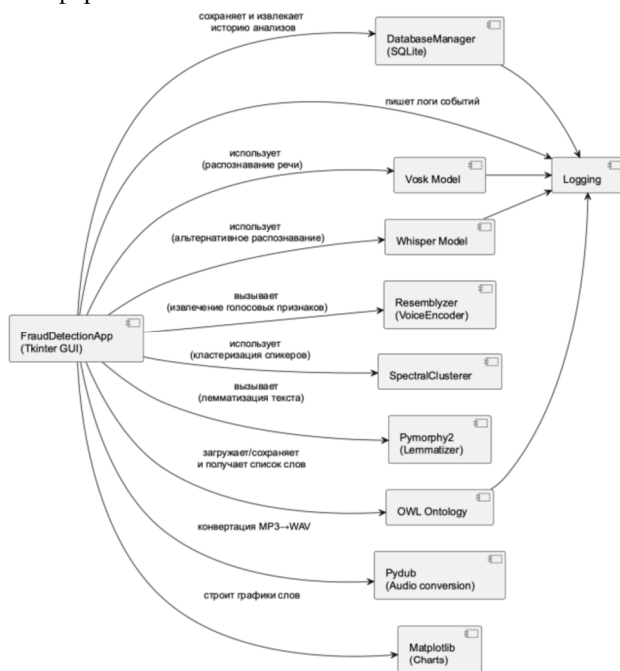


Рис. 2. Диаграмма компонентов

4. Эксперименты по определению мошенничества посредством анализа аудиофайлов телефонных разговоров

Для оценки эффективности разработанной программной платформы был проведён эксперимент на наборе аудиозаписей телефонных разговоров. Аудиозаписи были предоставлены Отделом по борьбе с киберпреступностью УМВД по Ульяновской области. Всего в выборку вошли 10 подтверждённых мошеннических звонков и 5 контрольных «чистых» записей, не содержащих признаков противоправного содержания. Такой набор позволил проверить работоспособность методики не только на единичном примере, но и при сравнительном анализе.

В ходе эксперимента система работала в двух режимах распознавания речи: Vosk и Whisper-small. Дополнительно учитывались результаты онтологического анализа, лемматизации текста и сегментации аудио по говорящим (Resemblyzer + SpectralClusterer). Для каждого разговора фиксировались следующие показатели:

- количество обнаруженных подозрительных терминов из онтологии (всего в базе 120 слов и словосочетаний);
- рассчитанная вероятность мошенничества (на основе весов терминов);
- оценка интонации собеседников (повышенный/спокойный/обычный тон);
- распределение подозрительных слов во времени.

4.1. Результаты на наборе записей

В среднем для мошеннических звонков система выявила от 8 до 15 ключевых терминов (например: «банк», «безопасный счёт», «служба безопасности», «подтверждение», «перевод»). Для контрольных звонков количество совпадений не превышало 2–3 слов, при этом итоговая вероятность мошенничества не превышала 5–7%, что указывает на низкий уровень ложных срабатываний. Для реальных мошеннических разговоров вероятность колебалась в диапазоне 15–42%, в зависимости от насыщенности диалога терминологией из онтологии.

Whisper показал более высокую точность распознавания речи на шумных записях ($\approx 92\%$) по сравнению с Vosk ($\approx 84\%$). При этом время обработки Whisper было выше (7–12 секунд против 4–8 секунд для Vosk). Использование онтологического анализа позволило повысить качество классификации: например, при распознавании только по ключевым словам точность составила 75%, а при учёте лемматизации и весов терминов – 83%.

4.2. Сравнительная оценка с другими подходами

Для сопоставления эффективности предложенного решения были рассмотрены результаты, опубликованные для других методов (см. раздел 1b). Итоговые показатели сведены в табл. 1.

Таблица 1

**Сравнение точности и скорости различных методов обнаружения
мошенничества**

Подход/Система	Точность, %	Время обработки	Особенности
Whisper + Онтология	83,2	7-12с	Учитывает семантику, интонацию, редактируема онтология
Vosk + Онтология	74,3	4-8с	Быстрое онлайн распознавание, хуже при шуме
IBM Watson STT + NLP	86-90	2-4с	Нет поддержки онтологии
Google Speech-to-Next + NLP	85-91	2-4с	Облачное решение
Pindrop (гибридная система)	85-90	1,5-15с	Голосовая биометрия, корпоративный сегмент
RNN-модели	93	60-600с	Высокая точность, но низкая скорость
GAN	88-92	30-180с	Высокие вычислительные затраты

Из табл. 1 видно, что предложенная методика занимает промежуточное положение: по точности она сопоставима с облачными решениями (Google, Watson), но выигрывает за счёт встроенной онтологической базы и анализа интонации. По времени обработки система уступает RNN и GAN, но работает существенно быстрее, что делает её применимой для прикладного использования.

4.3. Выводы по серии экспериментов

Проведённая серия экспериментов показала, что разработанная программная платформа обеспечивает комплексный анализ аудиозаписей, включающий:

- высокоточное распознавание речи на основе современных нейросетевых моделей;
- онтологический анализ, позволяющий учитывать семантические связи терминов;
- оценку интонации и сегментацию по говорящим, что повышает интерпретируемость результата.

Таким образом, система способна не только фиксировать факт употребления подозрительных слов, но и выделять наиболее значимые фрагменты диалога, что делает её полезным инструментом для служб безопасности и противодействия телефонному мошенничеству.

Заключение

В рамках выполнения данного проекта была спроектирована и реализована программная платформа, предназначенная для автоматизированного анализа телефонных разговоров с целью выявления признаков мошенничества. Основу интеллектуальной составляющей системы составляют нейросетевые модели распознавания речи, алгоритмы обработки естественного языка, а также разработанная онтологическая модель, содержащая набор ключевых слов и их весов, отражающих степень вероятности мошеннического характера обращения.

Особое внимание в разработке было уделено построению структурированной и редактируемой онтологии, которая обеспечивает гибкость и расширяемость системы. Благодаря встроенному редактору словаря пользователь может самостоятельно изменять содержание онтологии – добавлять новые термины или удалять устаревшие, что позволяет оперативно адаптировать платформу к меняющимся сценариям мошенничества.

Дальнейшее развитие проекта может быть связано с внедрением онлайн-анализа телефонных звонков в реальном времени, расширением онтологической базы за счёт подключения к внешним источникам знаний, применением более глубоких моделей анализа эмоций, а также интеграцией с системами безопасности организаций для оперативного реагирования на инциденты.

Список литературы

1. Кулешов В.В., Карасев П.И., Х.А.Х. Шамсулдин, Али А.А.Х.Н.А. Обзор методов использования нейросетевого фишинга // Информационная безопасность и защита персональных данных. Проблемы и пути их решения: Сборник материалов и докладов XV межрегиональной научно-практической конференции, Брянск, 28 апреля 2023 года / под общей редакцией О.М. Голембиовской. – Брянск: Брянский государственный технический университет, 2023. – С. 151-153.
2. Трушин И С., Алексеев А.А. Применение современных нейронных сетей в речевых технологиях для задачи разведки // Вестник науки. – 2024. – Т. 4, № 6(75). – С. 1209-1217.
3. Ермаков С.Р., Зубов Н.А. Развитие системы предотвращения подделки голосовой биометрии // Научно-технический вестник Поволжья. – 2024. – № 3. – С. 175-178.
4. IBM Watson Speech to Text. – <https://www.ibm.com/products/speech-to-text> (дата обращения: 27.05.2025).
5. Google Cloud Speech-to-Text. – <https://cloud.google.com/speech-to-text> (дата обращения: 27.05.2025).
6. Pindrop. – <https://www.pindrop.com> (дата обращения: 27.05.2025).
7. Nice Perform. – <https://bslgroup.com/support-for-discontinued-products/nice-perform> (дата обращения: 27.05.2025).
8. OpenVoice OS. – <https://www.openvoiceos.org> (дата обращения: 27.05.2025).

9. **Lependin A.A., Filin Y.A., Malinin P.v.** Speech Replay Spoofing Attack Detection System Based on Fusion of Classification Algorithms // Известия Алтайского государственного университета. – 2018. – No. 1(99). – P. 107-112.
10. **Zhao Q., Chen K., Li T., Yang Y., Wang X.** Detecting telecommunication fraud by understanding the contents of a call // Cybersecur – 2018. – Vol. 1, No. 8.
11. **Singh G., Singh P., Singh M.** Advanced real-time fraud detection using RAG-based LLMs.
12. **Пономарев К.Г.** Способы генерации голосовых дипфейков и методы их выявления // Молодежь. Наука. Инновации. – 2024. – Т. 1. – С. 172-176.
13. **Pustovoirov P.S., Seilova N.A.** Анализ методов подделки голоса: риски, случаи и стратегии защиты // International Journal of Information and Communication Technologies. – 2024. – Vol. 5, No. 1(17). – P. 120-132.
14. **Пономарев К.Г., Верещагина Е.А.** Математический аппарат и технологическая инфраструктура системы прогнозирования синтетических голосовых дипфейков // Инженерный вестник Дона. – 2024. – № 6(114). – С. 338-353.