

Программные системы и вычислительные методы*Правильная ссылка на статью:*

Трофимова В.С., Каршиева П.К., Рахманенко И.А. Метод трансферного обучения для дообучения нейронных сетей под особенности набора данных в задаче верификации диктора // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.71630 EDN: XHZCTS URL: https://nbpublish.com/library_read_article.php?id=71630

Метод трансферного обучения для дообучения нейронных сетей под особенности набора данных в задаче верификации диктора**Трофимова Варвара Сергеевна**

ORCID: 0009-0008-5044-2321



студент; кафедра комплексной информационной безопасности электронно-вычислительных систем;
Томский государственный университет систем управления и радиоэлектроники

634045, Россия, Томская область, г. Томск, ул. Красноармейская, 146, ауд. 509

✉ varvara.trofimova.01@mail.ru

Каршиева Полина Константиновна

ORCID: 0009-0004-8390-2348



студент; кафедра комплексной информационной безопасности электронно-вычислительных систем;
Томский государственный университет систем управления и радиоэлектроники

634045, Россия, Томская область, г. Томск, ул. Красноармейская, 146, ауд. 509

✉ polinakarshieva1@gmail.com

Рахманенко Иван Андреевич

ORCID: 0000-0002-8799-601X

кандидат технических наук



доцент; кафедра Безопасности информационных систем; Томский государственный университет
систем управления и радиоэлектроники

634045, Россия, Томская область, г. Томск, ул. Красноармейская, 146, ауд. 509

✉ ria@fb.tusur.ru

Статья из рубрики "Методы, языки и модели человека-машинного взаимодействия"

DOI:

10.7256/2454-0714.2024.3.71630

EDN:

XHZCTS

Дата направления статьи в редакцию:

30-08-2024

Дата публикации:

06-09-2024

Аннотация: Предметом исследования являются нейронные сети, дообученные с использованием метода трансферного обучения под особенности набора данных. Объектом исследования являются методы машинного обучения, используемые для решения задач верификации по голосу. Цель исследования заключается в повышении эффективности нейронных сетей в задаче голосовой верификации. В данной работе для процесса дообучения были подготовлены три набора данных на разных языках: на английском, на русском и на китайском. Также с использованием современных и предварительно обученных моделей ResNetSE34L и ResNetSE34V2 было проведено экспериментальное исследование, направленное на повышение эффективности нейронных сетей в задаче верификации диктора по произвольной фразе. Особое внимание уделялось оптимизации параметров обучения нейронных сетей с помощью метода тонкой настройки. Методология проведения исследования включает в себя оценку эффективности дообучения нейронных сетей под особенности набора данных в задаче верификации диктора, опираясь на значение равной ошибки 1-ого и 2-ого рода. Также был проведен ряд экспериментов, в ходе которых варьировались параметры и применялась техника замораживания слоев. Максимальное понижение равной ошибки 1-ого и 2-ого рода при использовании английского набора данных было достигнуто при помощи подбора эпох и скорости обучения, ошибка уменьшилась на 50%. Аналогичный подбор параметров при использовании русского набора данных уменьшил ошибку на 63,64%. При дообучении с использованием китайского набора данных наименьшее значение было достигнуто в эксперименте с замораживанием полностью связанного слоя, изменении скорости обучения и оптимизатора – ошибка снизилась на 16,04%. Полученные результаты могут использоваться при проектировании и разработке систем голосовой верификации и в образовательных целях. Также был сделан вывод, что метод трансферного обучения является эффективным при дообучении нейронных сетей под особенности набора данных, так как в подавляющем количестве экспериментов было достигнуто понижение EER, что говорит о повышении точности распознавания диктора.

Ключевые слова:

трансферное обучение, тонкая настройка, набор данных, верификация диктора, распознавание диктора, извлечение признаков, обработка речи, нейронные сети, глубокое обучение, распознавание образов

1 Введение

В настоящее время широко используются технологии, позволяющие осуществлять верификацию пользователя по различным биометрическим параметрам. Одним из таких параметров является голос [1]. Верификация диктора — это форма распознавания диктора, при которой принимается решение о принадлежности голосового образца индивиду, чья личность была заявлена [2]. Данная процедура позволяет обеспечить высокий уровень безопасности и удобства при получении пользователям доступа к

информации. Использование голосовой биометрии становится всё более актуальным в различных сферах, включая банковское дело, учреждения здравоохранения, системы безопасности и даже повседневную жизнь [\[3\]](#).

Дообучение нейронных сетей под особенности набора данных играет ключевую роль для повышения точности работы систем верификации диктора по голосу [\[4\]](#). Каждый набор данных имеет уникальные признаки. При дообучении используются предварительно обученные модели, которые в процессе адаптируются под новые признаки, что приводит к значительному улучшению эффективности работы системы.

Эффективность нейронных сетей, выполняющих верификацию диктора по голосу, напрямую влияет на безопасность и удобство использования различных сервисов. Точная система верификации диктора по голосу позволяет предотвращать несанкционированный доступ к конфиденциальной информации и защищать пользовательские аккаунты от злоумышленников.

2 Проведение эксперимента

2.1 Предварительные данные

Для дообучения нейронной сети были выбраны предварительно обученные модели на английском языке, такие как ResNetSE34L и ResNetSE34V2. Эти модели, основанные на архитектуре ResNet34, используются для идентификации диктора по произвольной фразе и последующей верификации. Основное отличие между ResNetSE34L и ResNetSE34V2 заключается в методах извлечения и адаптации признаков аудиоданных. Кроме того, ResNetSE34V2 предлагает более эффективную обработку данных, что обеспечивает более высокой уровень точности обучения.

В качестве обучающего набора данных для модели ResNetSE34L был использован речевой корпус VoxCeleb2 [\[5\]](#). С помощью данного корпуса были получены веса «baseline_lite_ap.model». Этот акустико-фонетический корпус речи представляет собой один из крупнейших наборов данных, применяемых для оценки систем автоматического распознавания речи.

Также для модели ResNetSE34L использовался речевой корпус TIMIT [\[6\]](#), на котором была обучена модель и получены веса «model000000100.model».

Далее для последующего дообучения моделей были подготовлены три речевых корпуса: на английском, русском и китайском языке.

В качестве первого был использован речевой корпус TIMIT [\[6\]](#), содержащий аудиозаписи на английском языке. Он был разделен на обучающий набор данных, содержащий 4620 аудиозаписей, и тестовый набор данных, состоящий из 1680 аудиозаписей. Этот набор данных был задействован в процессе дообучения с использованием обученных весов «baseline_lite_ap.model» и «baseline_v2_ap.model». В случае с «model000000100.model» этот подход не применялся, так как соответствующие веса были получены в ходе обучения модели ResNetSE34L на основе речевого корпуса TIMIT.

Речевой корпус на русском языке включает в себя аудиозаписи, сделанные 50 носителями русского языка. Каждый диктор имеет 50 аудиозаписей. Этот набор данных был разделен на две группы: обучающий набор, содержащий записи 30 дикторов, и тестовый набор, состоящий из записей 20 других дикторов [\[7\]](#).

В данном исследовании также использовался речевой корпус на китайском языке под названием «HI-MIA». Данные для корпуса собирались в домашних условиях с помощью микрофонных решеток и микрофона HI-FI. Датасет был разделен на два набора: обучающий, включающий записи 42 дикторов, и тестовый, состоящий из записей 40 дикторов [8].

Речевые корпусы на русском и китайском языке использовались в экспериментах со всеми представленными обученными весами, поскольку ни один из них ранее не использовался для обучения моделей.

Эти наборы данных были выбраны для дообучения нейронной сети с целью повышения ее эффективности в распознавании речи. Благодаря разнообразию диалектов, голосов и языков в аудиозаписях, можно получить более широкий охват вариативности в произношении слов и фраз.

Для самой верификации пользователя используется тестовый сценарий, включающий попытки аутентификации легального пользователя и атакующего (в случае несовпадения голоса). Легальная попытка в текстовом файле с тестовым сценарием помечена меткой 1, а атака на систему — меткой 0.

В контексте стремления улучшения эффективности нейронной сети в задачах голосовой верификации основное внимание уделялось понижению равной ошибки 1-го и 2-го рода [9]. Цель заключалась в достижении более точного распознавания диктора, что важно для обеспечения высокой надежности системы.

В рамках проведенного исследования для качественного уменьшения ошибок 1-ого и 2-ого рода был применен метод трансферного обучения с использованием предварительно обученных весов модели, выполняющей задачу верификации диктора по произвольной фразе.

Используемый метод представляет собой технику дообучения нейронной сети, когда предварительно обученная модель на большом наборе данных дополнительно обучается на более узком наборе данных для выполнения задачи верификации диктора

2.2 Параметры дообучения

В процессе дообучения нейронной сети использовались различные стратегии для адаптации модели к новым данным. Особое внимание уделялось выбору параметров обучения, включая скорость обучения, изменение оптимизатора, количество эпох обучения и структура нейронной сети.

Помимо описанных выше параметров, также учитывался размер батча. При дообучении модели на речевом корпусе TIMIT размер батча был 50. При работе с русским корпусом он составлял размер 30, а с китайским размер был равен 40. Описание параметров, используемых для дообучения нейронной сети:

- 1) Скорость обучения влияет на сходимость модели и предотвращает переобучение. Изменение весов должно быть умеренным, чтобы избежать нарушения выученных признаков.
- 2) Количество эпох в процессе дообучения влияет на переобучение модели. Также недостаточное количество эпох в процессе дообучения может привести к недостаточной адаптации.

- 3) Маленькие размеры батча в процессе дообучения позволяют найти баланс между вычислительной эффективностью, устойчивостью к изменениям и способностью модели к обобщению на новые данные.
- 4) Оптимизатор отвечает за настройку весов модели в процессе обучения, чтобы минимизировать функцию потерь.
- 5) Структура нейронной сети описывает архитектуру модели, включая количество слоев, а также их взаимосвязь. Процесс дообучения определяет, как новые данные интегрируются в существующую архитектуру. Изменение структуры сети может включать в себя замораживание слоев. Это позволяет модели изучить новые особенности данных и сохранять ее первоначальное обучение.

В процессе исследования аудиозаписи были разделены на два текстовых файла: «train» и «test». Это позволило систематизировать данные и обеспечить необходимую структуру для последующего анализа.

3 Результаты

3.1 Результат дообучения с использованием речевого корпуса TIMIT

В целях определения наилучшей стратегии дообучения была проведена серия экспериментов с использованием набора данных TIMIT для обученных весов «baseline_lite_ap.model» и «baseline_v2_ap.model». Чтобы определить, насколько эффективнее нейронная сеть распознает дикторов, было рассмотрено значение EER (Equal Error Rate), полученное в ходе тестирования и последующего дообучения модели [\[9\]](#).

Чтобы аналитически сравнить значения, полученные после дообучения, с изначальными данными, модель была оценена без применения обучающего процесса. Было получено значение EER равное 0.012 и 0.013 соответственно для «baseline_lite_ap.model» и «baseline_v2_ap.model».

Далее модель была подвергнута дообучению с внесением изменений в её параметры. По завершению двадцати экспериментов была составлена сводная таблица, в которой представлены результаты равной ошибки 1-ого и 2-ого рода для сравнения эффективности дообучения модели. В данной таблице указаны используемые модели, предварительно обученные веса и проделанные эксперименты. Каждый эксперимент включал в себя попытку дообучения моделей на речевом корпусе TIMIT [\[6\]](#).

Таблица 1 — Результаты экспериментов для набора TIMIT

№	Эксперимент	Модели и веса		
		model000000100.model	baseline_lite_ap	baseline_v2_ap
0	До дообучения	0,066	0,012	0,013
1	Без изменений параметров	*	0,014	0,017
2	Замораживание 1 и 2 сверточного слоя	*	0,015	0,011
3	Замораживание 1 сверточного слоя	*	0,012	0,011
4	Замораживание 2 сверточного слоя	*	0,011	0,008
	-			

5	Замораживание полносвязного слоя, скорость обучения и изменение оптимизатора	*	0,011	0,012
6	Замораживание полносвязного слоя	*	0,009	0,015
7	Изменение оптимизатора	*	0,008	0,007
8	Подбор эпох и скорости обучения (max)	*	0,012	0,007
9	Подбор эпох и скорости обучения (min)	*	0,006	0,010
10	Понижение lr_decay и lr	*	0,008	0,008

При обучении с применением весов модели «baseline_lite_ap.model» наилучший результат был получен при незначительном изменении скорости обучения. Исходное значение EER было равно 0,012, после дообучения оно снизилось до 0,006. Данное изменение свидетельствует об уменьшении EER на 50%.

При использовании предварительно обученных весов «baseline_v2_ap.model» лучший результат был достигнут при изменении оптимизатора и существенного снижения скорости обучения. Перед дообучением значение равной ошибки 1-ого и 2-ого рода составляло 0,013, после дообучения показатель сократился до 0,007, что свидетельствует об уменьшении на 46,15%.

Помимо проведения процесса дообучения все модели также были подвергнуты обучению на наборе данных TIMIT^[6]. Полученные результаты были сопоставлены с EER^[9], полученными в контексте дообучения, с целью демонстрации уменьшения EER при процессе дообучения по сравнению с исключительно базовым обучением.

3.2 Результат дообучения с использованием Русского речевого корпуса

Для оптимизации работы модели и повышения эффективности распознавания дикторов была проведена серия экспериментов на основе набора данных на русском языке с использованием предварительно обученных весов «model000000100.model», «baseline_lite_ap.model» и «baseline_v2_ap.model».

План постановки экспериментов, проводимых с русским речевым корпусом^[8], не отличался от экспериментального исследования с речевым корпусом TIMIT^[7]. Чтобы оценить эффективность распознавания диктора до процесса дообучения, модель была подвергнута оцениванию без обучения. Это было предпринято для оценки эффективности модели перед внесением корректив в её параметры и реализации процесса дообучения.

В связи с тем, что изначально модель обучалась на английском языке, а последующее предварительное обучение проводилось на русском, процесс смены языковой среды оказал влияние на точность модели. Точность модели в течение десяти эпох не поднималась выше шестидесяти. Этот переход требовал адаптации модели к новому языку и его особенностям. Английский и русский язык имеют множество различий в грамматике и синтаксисе. По мимо вышеперечисленного повлиять на дообучение может также морфология каждого языка. Подобная процедура проводится для того, чтобы расширить функциональность модели для работы с различными языками и культурными контекстами.

В результате проведения описанных выше экспериментов с использованием набора данных на русском языке была сформирована таблица 2. В данной таблице представлены результаты равных ошибок 1-ого и 2-ого рода, полученные в ходе всех экспериментов, в которых использовались предварительно обученные веса моделей «baseline_v2_ap.model», «baseline_lite_ap.model» и «model000000100.model».

Таблица 2 — Результаты экспериментов для Русского речевого корпуса

№	Эксперимент	Модели и веса		
		model000000100.model	baseline_lite_ap	baseline_v2_ap
0	До дообучения	0,066	0,012	0,078
1	Без изменений параметров	0,055	0,014	0,114
2	Замораживание 1 и 2 сверточного слоя	0,035	0,013	0,066
3	Замораживание 1 сверточного слоя	0,053	0,014	0,064
4	Замораживание 2 сверточного слоя	0,053	0,010	0,067
5	Замораживание полносвязного слоя, скорость обучения и изменение оптимизатора	0,032	0,007	0,078
6	Замораживание полносвязного слоя	0,048	0,010	0,062
7	Изменение оптимизатора	0,049	0,011	0,067
8	Подбор эпох и скорости обучения (max)	0,033	0,008	0,073
9	Подбор эпох и скорости обучения (min)	0,024	0,008	0,064
10	Понижение lr_decay и lr	0,028	0,009	0,068

В процессе обучения с применением весов «model000000100.model» значительное улучшение было получено при незначительной модификации скорости обучения. Начальное значение EER составляло 0,066, после дообучения показатель сократился до 0,024, что свидетельствует об уменьшении на 63,64%.

Путем использования предварительно обученных весов «baseline_lite_ap.model» наилучший результат был достигнут при замораживании полносвязного слоя, изменении скорости обучения и оптимизатора. Исходное значение EER было равно 0,012, после дообучения оно снизилось до 0,007, что означает сокращение EER на 41,67%.

При использовании предварительно обученных весов «baseline_v2_ap.model» самый качественный процесс обучения был реализован с помощью замораживания полносвязного слоя. Перед дообучением значение EER составляло 0,078, после дообучения показатель сократился до 0,062, что свидетельствует об уменьшении на 20,51%.

3.3 Результат дообучения с использованием речевого корпуса НI-MIA

В данном исследовании предварительно обученные веса были также дообучены на речевом корпусе, содержащим аудиозаписи на китайском языке [8].

Работа с этим набором данных не привела к успешному дообучению модели в большинстве случаев. Были проведены эксперименты, ранее проведенные с другими наборами данных. Также были добавлены эксперименты с другими значениями скорости обучения и замораживанием других слоев. В результате только при работе с предварительно обученными весами «baseline_lite_ap.model» в эксперименте с замораживанием полносвязного слоя, изменением скорости обучения и оптимизатора было достигнуто успешное дообучение.

При дообучении, где использовались веса «model000000100.model», исходная EER была равна 0,066 [9]. Новый набор данных содержал разнообразные фон-шумы, что представляет собой типичную среду взаимодействия в реальном мире. Параметры модели были адаптированы для качественного дообучения. После дообучения на наборе данных, содержащим аудиозаписи на китайском языке, было замечено увеличение EER до 0.214.

Все результаты, проведённых экспериментов с набором данных на китайском языке, представлены в таблице 3.

Таблица 3 — Результаты экспериментов для HI-MIA

№	Эксперимент	Модели и веса		
		model000000100.model	baseline_lit_ap	baseline_v2_ap
0	До дообучения	0,180	0,106	0,089
1	Без изменений параметров	0,214	0,108	0,091
2	Замораживание 1 и 2 сверточного слоя	0,197	0,125	0,094
3	Замораживание 1 сверточного слоя	0,201	0,138	0,096
4	Замораживание 2 сверточного слоя	0,207	0,141	0,097
5	Замораживание полносвязного слоя, скорость обучения и изменение оптимизатора	0,259	0,089	0,121
6	Замораживание полносвязного слоя	0,210	0,130	0,120
7	Изменение оптимизатора	0,231	0,147	0,092
8	Подбор эпох и скорости обучения (max)	0,302	0,222	0,148
9	Подбор эпох и скорости обучения (min)	0,229	0,134	0,123
10	Понижение lr_decay и lr	0,223	0,135	0,123

Анализ показал, что причиной повышения EER могут быть как лингвистические и фонетические различия между английским и китайским языком, так и присутствие шума, маскирующего ключевые акустические особенности речи. Дообучение на зашумленном китайском наборе данных привело к переобучению на специфических особенностях этого набора данных, что снизило обобщаемость модели. Также возможно, что стандартные методы дообучения не справляются со своей задачей в условиях сильной языковой вариативности и присутствия шума.

4 Область применения результатов и новизна

Данное исследование позволяет избежать дополнительных затрат на обучение моделей с нуля и, при использовании данных моделей, повышает уровень эффективности работы систем верификации диктора по голосу. Технология верификации диктора по голосу широко распространена в банковской сфере, так как при обращении клиента в контакт-центр единственным доступным биометрическим параметром является голос.

В работе были использованы известные предварительно обученные модели и метод трансферного дообучения в рамках заданной предметной области. Основным научным результатом являются результаты оценки эффективности и применимости конкретных методов дообучения нейронных сетей под особенности наборов данных в задаче текстонезависимой верификации диктора.

5 Выводы

В работе было проведено исследование процесса трансферного обучения нейронных сетей под особенности набора данных, что позволило получить более точные модели после проведенного дообучения.

Полученные наблюдения подчеркивают важность корректной настройки параметров при дообучении моделей. Тонкая настройка, проводимая в ходе дообучения моделей, благоприятно сказывается на оптимизации их эффективности.

При использовании предварительно обученных весов «baseline_lite_ap.model» и речевого корпуса TIMIT было получено максимальное понижение равной ошибки 1-ого и 2-ого рода, она уменьшилась на 50%.

С набором данных на русском языке заметно снизился показатель EER при использовании предварительно обученных весов «model000000100.model». Снижение составило 63.64%.

В работе с набором данных HI-MIA успешное дообучение модели было достигнуто только в эксперименте с замораживанием полносвязанного слоя, изменением скорости обучения и оптимизатора. По итогу было зафиксировано, что равная ошибка 1-ого и 2-ого рода (EER) снизилась на 16,04%.

Библиография

1. Гассиев Д. О., Сахаров В. А., Ермолаева В. В. Голосовая аутентификация // Тенденции развития науки и образования. 2019. № 56(2). С. 22-24.
2. ГОСТ Р 58668.11-2019 (ИСО/МЭК 19794-13:2018). Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 11. Данные голоса. Москва: Стандартинформ, 2019. 28 с.
3. Девятков В.В. Системы искусственного интеллекта. Гл. ред. И.Б. Федоров. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001. 352 с.
4. Галушкин А.И. Нейронные сети. Основы теории. – М.: Горячая линия – Телеком, 2012. 496 с.
5. Suzuki K. Artificial Neural Networks: Architectures and Applications. – Publisher: InTech, 2013. 256 p.
6. Евсюков М.В., Путято М.М., Макарян А.С., Немчинова В.О. Методы защиты в современных системах голосовой аутентификации // Прикаспийский журнал: управление и высокие технологии. 2020. № 3(59). С 84-92.
7. Nagrani A., Chung J.S., Zisserman A. VoxCeleb: A large-scale speaker identification

- dataset // arXiv:1706.08612. 2018. URL: <https://arxiv.org/pdf/1706.08612.pdf> (дата обращения: 15.07.2024).
8. Hinton G. E., Srivastava X., Krizhevsky A., Sutskever I., Salakhutdinov R. R. Improving neural networks by preventing co-adaptation of feature detectors // arXiv:1207.0580. 2012. URL: <https://arxiv.org/pdf/1207.0580.pdf> (дата обращения: 22.07.2024).
9. Конев А.А. Модель и алгоритмы анализа и сегментации речевого сигнала: автореф. дис. на соиск. ученой степ. канд. техн. наук. Томск, 2007. 150 с.
10. Qin X., Bu H., Li M. HI-MIA: A Far-field Text-Dependent Speaker Verification Database and the Baselines // IEEE International Conference on Acoustics, Speech, and Signal Processing, 2020. Рр. 7609-7613.
11. Рахманенко И.А., Шелупанов А.А., Костюченко Е.Ю. Автоматическая верификация диктора по произвольной фразе с применением свёрточных глубоких сетей доверия // Компьютерная оптика. 2020. № 44(4). С. 596-605.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Статья посвящена актуальной и востребованной теме – разработке методов улучшения верификации диктора на основе биометрии голоса с использованием нейронных сетей. В частности, исследуется трансферное обучение нейронных сетей, где предварительно обученные модели адаптируются под специфические особенности различных речевых данных. Предложенный метод решает задачу повышения точности и надежности систем голосовой верификации, что имеет большое значение для таких сфер, как банковское дело, системы безопасности, медицинские учреждения и повседневные цифровые сервисы.

Методология статьи демонстрирует высокий научный уровень. Авторы используют несколько известных нейронных сетей, таких как ResNetSE34L и ResNetSE34V2, которые дообучаются на различных наборах данных (английский, русский и китайский речевые корпусы). Приводятся конкретные параметры обучения, такие как скорость обучения, размер батча и количество эпох, что позволяет в полной мере оценить проведенные эксперименты и их результаты. Использование трансферного обучения является одним из ключевых аспектов в современных исследованиях искусственного интеллекта, что свидетельствует о высокой актуальности выбранного подхода.

Тема статьи находится на острие современных технологий. Применение биометрии голоса для идентификации личности набирает популярность, и повышение точности таких систем становится особенно важным в условиях роста числа кибератак и увеличения требований к безопасности. Кроме того, использование трансферного обучения для дообучения моделей на новых наборах данных позволяет существенно сократить затраты на вычислительные ресурсы и повысить производительность системы. Статья предлагает оригинальный подход к адаптации предварительно обученных моделей под особенности конкретных речевых данных, что позволяет повысить точность систем верификации диктора. Авторы исследуют влияние различных параметров обучения на эффективность моделей и предоставляют детализированные результаты, которые показывают значительное снижение ошибки (EER). Особенно интересным является применение метода на разных языковых данных, что расширяет область применения предложенной технологии.

Стиль изложения материала является логичным и последовательно выстроенным. Статья структурирована классическим образом, начиная с введения в тему, обзора

используемых моделей и заканчивая подробным анализом результатов. Особенno стоит отметить ясность и четкость представления экспериментов. Каждый этап исследования подробно описан, что позволяет легко следовать ходу мысли авторов и оценить значимость полученных данных.

Выводы статьи подкреплены результатами проведенных экспериментов и представляют собой логическое завершение исследования. Авторы подчеркивают, что использование трансферного обучения с предварительно обученными весами позволяет значительно снизить процент ошибок 1-го и 2-го рода (EER), что улучшает общее качество работы системы. Выводы полностью соответствуют заявленным целям исследования и подтверждают эффективность предложенного подхода.

Статья будет интересна как специалистам в области искусственного интеллекта и машинного обучения, так и исследователям, работающим над улучшением биометрических систем безопасности. Кроме того, результаты работы могут быть полезны практикам, занимающимся разработкой коммерческих решений в области верификации пользователей на основе голоса.

Статья представляет собой значимый вклад в исследование методов улучшения голосовой верификации с использованием нейронных сетей. Методология, представленная авторами, демонстрирует научную новизну и оригинальность подхода, а результаты экспериментов подтверждают эффективность предложенного метода. Рекомендую статью к публикации без существенных замечаний.