На правах рукописи

ШИРЯЕВ ЕГОР МИХАЙЛОВИЧ

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ, МЕТОДЫ И АЛГОРИТМЫ ЭФФЕКТИВНОЙ РЕАЛИЗАЦИИ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ, СОХРАНЯЮЩИХ КОНФИДЕНЦИАЛЬНОСТЬ

Специальность: 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва — 2025

Работа выполнена на кафедре вычислительной математики и кибернетики факультета математики и компьютерных наук имени профессора Н. И. Червякова федерального государственного автономного образовательного учреждения высшего образования «Северо-Кавказский федеральный университет» (ФГАОУ ВО СКФУ).

Научный руководитель:	Доктор физико-математических наук, доцент, **Бабенко Михаил Григорьевич**
Официальные оппоненты:	**Фёдоров Роман Константинович,** доктор технических наук, заведующий лабораторией комплексных информационных систем, Федеральное государственное бюджетное учреждение науки Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук **Петровский Михаил Игоревич,** кандидат физико-математических наук, доцент, доцент кафедры интеллектуальных информационных технологий факультета вычислительной математики и кибернетики Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет им. М.В. Ломоносова»
Ведущая организация:	Федеральное государственное автономное образовательное учреждение высшего образования «Южный федеральный университет»

Защита состоится «27» ноября 2025 г. в 16:00 на заседании диссертационного совета 24.1.120.01 при Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В. П. Иванникова Российской Академии Наук по адресу: 115035, г. Москва, ул. Садовническая, д. 41 ст. 2.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного учреждения науки Института системного программирования им. В. П. Иванникова Российской академии наук.

Автореферат разослан «___» _____2025 г.

Ученый секретарь
диссертационного совета 24.1.120.01,
кандидат физико-математических наук Турдаков Д. Ю.

Общая характеристика работы

Актуальность работы. Развитие методов искусственного интеллекта, а также вычислительной аппаратуры привело к тому, что машинное обучение и искусственные нейронные сети используются уже не только для узконаправленных специализированных задач, но и для широкого круга задач различных сфер деятельности человека. Особое внимание стоит уделить методам анализа Больших данных, отметить их быстрое развитие и повсеместное внедрение в производство для аналитического сопровождения различных процессов. С развитием методов машинного обучения и искусственных нейронных сетей растет и вычислительная сложность решаемых с их помощью задач, что приводит к дефициту вычислительных мощностей, как у рядового пользователя, так и у малого и среднего бизнеса.

Наиболее рациональным способом снижения затрат на необходимые вычислительные ресурсы является применение облачных технологий, а именно услуг облачных провайдеров. Однако в данном случае возникают определенные риски, связанные с конфиденциальностью обрабатываемых данных. В целом риски можно разделить на две категории:

1. Внешние риски.
2. Внутренние риски.

К внешним рискам можно отнести любого рода кибератаки на безопасность системы облачного провайдера. В данном случае угрозы конфиденциальности информации заключаются в том, что злоумышленник может украсть информацию, которая в настоящий момент обрабатывается в облаке, либо повредить ее, нарушив целостность и/или внеся в нее изменения. К внутренним рискам обычно относятся случаи, когда злоумышленник либо внедрился в персонал облачного провайдера, либо имеет доступ к системе администрирования, либо существует предварительный сговор с персоналом облачного провайдера. Таким образом, в случае использования облачных сервисов, наряду со стандартными рисками нарушения конфиденциальности возникают дополнительные, связанные с удаленной обработкой конфиденциальных данных.

Вопрос конфиденциальности данных, обрабатываемых методами искусственного интеллекта в облачных вычислительных системах, в настоящий момент становится критическим. Такая ситуация складывается в связи с внедрением методов машинного обучения и искусственных нейронных сетей в чувствительные сферы деятельности человека, такие как, например, здравоохранение, финансовый сектор, государственный сектор. Здесь конфиденциальными данными могут выступать персональные данные пользователя, данные составляющие врачебную тайну, корпоративную тайну и в некоторых случаях государственную тайну. Кроме того, если рассматривать такие концепции, как «Умный город», то вопрос

конфиденциальности данных становится еще более острым, так как возникают риски для данных достаточного широкого спектра.

Сложившаяся ситуация осложняется тем, что в настоящий момент на территории Российской Федерации одним из приоритетов является переход к передовым цифровым интеллектуальным производственным технологиям, роботизированным системам, новым материалам и способам конструирования, созданию систем обработки больших объемов данных, машинного обучения и искусственного интеллекта. Что так же влечет за собой повышение внимания к вопросу конфиденциальности данных, обрабатываемых в облачных системах, в том числе с применением аппарата искусственных нейронных сетей.

Искусственные нейронные сети находят все большее применение в различных областях человеческой деятельности. В здравоохранении они применяются для распознавания и анализа результатов исследований пациентов. В финансовом секторе – для анализа курсов валют и прогнозирования роста/снижения стоимости отдельных активов и биржи в целом. На муниципальном уровне – для анализа трафика на дорогах, потребления коммунальных услуг и т.п. Корпорации применяют искусственные нейронные сети для анализа и оптимизации внутренних процессов. Рядовые пользователи используют искусственные нейронные сети для личных нужд. Если в процессе обучения нейронных сетей данные являются открытыми и общедоступными, то данные, которые подаются на вход обученной нейронной сети, зачастую являются конфиденциальными.

В целях обеспечения конфиденциальности обрабатываемых искусственными нейронными сетями данных требуется разработка методов и алгоритмов, реализующих функционал криптографических систем, но с возможностью корректно дешифровать результат после обработки зашифрованных данных. В данной области наиболее перспективным направлением является полностью гомоморфное шифрование (ПГШ), оно позволяет обрабатывать зашифрованные данные с помощью гомоморфного сложения и умножения. В контексте искусственных нейронных сетей операции сложения и умножения являются ключевыми, функции активации нейронов, хоть зачастую и построены на применении операции сравнения, могут быть приближенны с помощью численных методов. Такое мнение транслируют авторы многочисленных работ, посвященных исследованию применения ПГШ в машинном обучении и искусственных нейронных сетях, высоко оценивая его перспективность. Основными препятствиями для применения схем ПГШ являются высокая вычислительная сложность и ограниченность набора поддерживаемых операций, как правило сложением и умножением. Преодоление этих препятствий требует тщательного исследования, разработки и адаптации моделей, методов и алгоритмов, как гомоморфных операций ПГШ, так и самих искусственных нейронных сетей. Получение новых научных результатов в данной области

исследований позволит повысить эффективность искусственных нейронных сетей, сохраняющих конфиденциальность, и расширить область их применения в первую очередь за счет расширения класса задач, допускающих применение облачных технологий.

Значительный научный вклад в рассматриваемую область внесли следующие исследователи: C. Gentry, V. Vaikuntanathan, Z. Brakerski, Yu. Polyakov, V. Shoup, A. Alexandru, A. Kim, S. Halevi, V. Zucca, D. Micciancio, J. H. Cheon, A. Al Badawi, Л. К. Бабенко, М. Г. Бабенко и М. А. Дерябин.

Объектом диссертационного исследования являются искусственные нейронные сети, сохраняющие конфиденциальность.

Предмет исследования – методы и алгоритмы искусственных нейронных сетей, сохраняющих конфиденциальность.

Целью диссертационного исследования является разработка методов и алгоритмов базовых операций искусственных нейронных сетей, сохраняющих конфиденциальность, уменьшающих время их обработки.

Научная задача диссертационной работы состоит в исследовании и разработке математической модели, методов и алгоритмов для повышения эффективности реализаций искусственных нейронных сетей, сохраняющих конфиденциальность на основе схем полностью гомоморфного шифрования.

Для решения поставленной общей научной задачи была произведена ее декомпозиция на ряд **частных задач**:

1. Разработка методов и алгоритмов для проектирования слоев искусственной нейронной сети, сохраняющей конфиденциальность.
2. Адаптация существующих моделей искусственных нейронных сетей под ограничения гомоморфных шифров.
3. Разработка комплекса программ, реализующего функционал искусственных нейронных сетей на базе схем гомоморфного шифрования.

Методология и методы исследования включают использование математического аппарата линейной алгебры, математического анализа, теории алгоритмов, численных методов, математического моделирования, теории распознавания образов и математической статистики.

Соответствие паспорту научной специальности.

Область исследования соответствует паспорту специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» по следующим пунктам:

п. 3. Модели, методы, архитектуры, алгоритмы, языки и программные инструменты организации взаимодействия программ и программных систем.

п. 4. Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов.

п. 9. Модели, методы, алгоритмы, облачные технологии и программная инфраструктура организации глобально распределенной обработки данных.

Научная новизна диссертационной работы:

1. Предложен модифицированный метод умножения матриц для реализации слоев искусственной нейронной сети, сохраняющей конфиденциальность, позволяющий уменьшить вычислительную сложность алгоритма с $O(n^4)$ до $O(n^2)$.
2. Разработаны методы дистилляции и квантизации параметров для адаптации искусственных нейронных сетей к ограничениям гомоморфных шифров, позволяющие сократить потребление памяти примерно в 1500 раз, уменьшить время обработки данных в среднем в 30 раз, при потере доли верных классификаций в диапазоне от 0.5% до 1%.
3. Разработаны функции активации для проектирования искусственных нейронных сетей, сохраняющих конфиденциальность.
4. Разработан программный комплекс для реализации искусственных нейронной сети, сохраняющих конфиденциальность.

Теоретическая и практическая значимость.

Теоретическая значимость данной диссертационной работы заключается в разработанных математической модели, методах и алгоритмах, включая методы повышения эффективности операции умножения матриц для гомоморфных шифров, методы адаптации искусственных нейронных сетей, учитывающие ограничения гомоморфных шифров, такие как дистилляция знаний и квантизация параметров, алгоритмы полиномиального приближения функций активации и функции активации на основе полиномов с обучаемыми коэффициентами.

Практическая значимость разработанных математической модели, методов и алгоритмов заключается в возможности реализации на их основе моделей искусственных нейронных сетей, сохраняющих конфиденциальность, для обработки входных данных в зашифрованном виде. Полученные результаты могут быть использованы при построении моделей искусственных нейронных сетей в областях с повышенными требованиями к конфиденциальности данных, как, например, здравоохранение, финансовый сектор, банковская деятельность и т. п.

На защиту выносятся следующие положения:

1. Метод умножения зашифрованных матриц.
2. Алгоритмы проектирования сверточного и скрытого слоев нейронной сети на базе модифицированного метода умножения матриц, сохраняющего конфиденциальность.
3. Методы адаптации искусственных нейронных сетей учитывающие ограничения гомоморфных шифров.

4. Алгоритмы полиномиального приближения стандартных функций активации и функции активации на основе полиномов с обучаемыми коэффициентами.

Основные результаты диссертационного исследования были использованы в рамках следующих научно-технических работ:

1. РНФ № 19-71-10033 «Эффективная, безопасная и отказоустойчивая система распределенного хранения и обработки конфиденциальных данных с регулируемой избыточностью для проектирования мобильных облаков на маломощных вычислительных устройствах»;
2. РНФ № 22-71-10046 «Разработка новых методов и алгоритмов для повышения надежности и безопасности хранения, передачи и обработки данных в туманных вычислениях»;
3. Северо-Кавказский федеральный университет «Интеллектуальный блок управления распределенной системой хранения данных в гетерогенных средах с регулируемой избыточностью и безопасностью»;
4. РНФ № 25-71-30007 «Новые технологии для проектирования облачных сервисов машинного обучения, сохраняющих конфиденциальность» (2 и 3 глава).

Достоверность полученных результатов обеспечивается строгостью проведения математических доказательств, при получении которых был использован научно-методический аппарат математического анализа, теории чисел и численных методов, и подтверждается проведенным сравнительным анализом разработанных методов и алгоритмов с известными ранее с точки зрения скорости обработки данных, потребления оперативной памяти и точности распознавания образов.

Авторский вклад соискателя.

Все изложенные в диссертационной работе результаты получены при непосредственном участии автора. Из результатов работ, выполненных коллективно, в диссертацию включены только полученные непосредственно автором. В работах [1, 2, 8, 11-15, 24-26] автором рассмотрены модели распределенных вычислений, а именно облачных и туманных вычислений, проанализированы и обозначены их уязвимости с точки зрения безопасности. В работах [3, 16] автором проанализирована безопасность «Умных городов», уточнены требования к безопасности. В работах [17, 18] автором исследованы методы машинного обучения и искусственных нейронных сетей. В работах [4-5, 10, 19-22] исследованы вычислительные характеристики гомоморфных шифров, а также методы повышения их эффективности. В работах [6, 7, 9, 23] автором проведены исследования искусственных нейронных сетей, сохраняющих конфиденциальность, на базе гомоморфных шифров, разработан метод умножения матриц, сохраняющий конфиденциальность входных данных, характеризующийся меньшим потреблением памяти и меньшей вычислительной сложностью, достигаемыми путем сокращения количества гомоморфных операций.

Разработан комплекс программ для разработки и исследования сверточных нейронных сетей, сохраняющих конфиденциальность, с применением схем ПГШ [27-29].

Апробация работы. Основные результаты диссертационного исследования докладывались на международных конференциях, среди которых «International Workshop on Advanced in Information Security Management and Applications (AISMA-2023)» (г. Алигарх, Индия, г. Ставрополь, г. Красноярск, Россия), «Conference on Current Problems of Applied Mathematics and Computer Systems (CPAMCS 2023)» (г. Ставрополь, Россия), «V Ibero-American Congress of Smart Cities (ICSC-CITIES 2022)» (г. Куэнка, Эквадор), «International Conference on Mathematics and its Applications in new Computer Systems (MANCS 2021)» (г. Ставрополь, Россия), «2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus 2021)» (г. Москва и г. Санкт-Петербург, Россия), «Advances in Automation II: Proceedings of the International Russian Automation Conference (RusAutoConf 2020)» (г. Сочи, Россия), «The International Workshop on Information, Computation, and Control Systems for Distributed Environments» (г. Иркутск, Россия), «Spring/Summer Young Researchers' Colloquium on Software Engineering (SYRCoSE 2020, 2024)» (г. Ставрополь, Россия), «International Conference Engineering and Telecommunication (En&T 2021)» (г. Москва, Россия), «IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW 2021)» (г. Портленд, Орегон, США), «International Conference on High Performance Computing & Simulation (HPCS 2019)» (г. Дублин, Ирландия).

Публикации по теме диссертации. Основные результаты по теме диссертационного исследования изложены в 29 публикациях, 4 из которых изданы в журналах, рекомендованных ВАК [2, 4, 6, 7], 15 – в тезисах докладов конференций [11-18, 20-26], 8 – в журналах, входящих в международные базы цитирования Web of Science и Scopus [1-3, 5-9]. Получено 3 свидетельства о государственной регистрации программ для ЭВМ [27-29].

Моделирование и вычислительный эксперимент проведены на базе платформы Yandex Cloud в сервисе DataSphere на конфигурации c1.32, которая включает 32 виртуальных центральных процессора Intel Ice Lake и 256 ГБ оперативной памяти, с использованием языка программирования высокого уровня Python для разработки модулей, библиотеки PyTorch для проектирования открытых сверточных нейронных сетей и библиотек TenSEAL и Concrete-ML для проектирования сверточных нейронных сетей, сохраняющих конфиденциальность. В качестве критериев оценки эффективности разработанных моделей, методов и алгоритмов, использовались время одной итерации работы сверточной нейронной сети (миллисекунды, мс и секунды, с); память, занимаемая сверточной нейронной

сетью (гигабайты, ГБ); точность распознавания 10 000 изображений сверточной нейронной сетью (проценты, %).

Структура диссертации. Полный объем диссертации составляет 156 страниц, включая 39 рисунков и 7 таблиц. Список литературы содержит 186 наименований.

Основное содержание работы

Во введении обоснована актуальность темы диссертации, сформулированы цель и задачи работы, выбраны объект и предмет исследования, показаны научная новизна, практическая и теоретическая ценность полученных результатов, приведены основные положения, выносимые на защиту.

В первой главе рассмотрены распределенные вычислительные системы (РВС), а именно облачные вычисления (ОВ) и туманные вычисления (ТВ). Проведен анализ их безопасности, выявлены проблемы конфиденциальности таких систем. Рассмотрены методы искусственного интеллекта (ИИ), особое внимание уделено искусственным нейронным сетям, а именно классу сверточных нейронных сетей (СНС). Проанализировано применение методов искусственного интеллекта в публичных распределенных системах (табл. 1).

Таблица 1 – Историческая справка по рассмотренным технологиям РВС, ОВ и ИИ [6]

Год	**Распределенные вычисления**	**Облачные технологии**	**Искусственный интеллект**
1943	-	-	Концепция искусственной нейронной сети
1954	-	-	Зарождение генетических алгоритмов
1959	-	-	Машинное обучение
1962	Модель коллективных вычислений	-	-
1966	-	-	Появление языковых моделей
1978	Принципы распределения работы между процессорами	-	-
1980		-	Теоретическое описание глубокого обучения
1992	Зарождение GRID	-	-
1996	Проект GIMPS по поиску целых чисел	-	-
1999	Проект SETI на базе BOINC	-	-
2000	-	-	Начало практического применения глубокого обучения
			Компьютерное зрение

2006	-	Зарождение концепции облачных вычислений	
2008	-	Определение концепции облачных вычислений как услуги	
2009	-	Запуск Google Apps	
2011	-	Стандартизация SaaS, PaaS и IaaS как моделей обслуживания в облачных вычислениях	
2015	-	Развитие туманных вычислений, как основы для Интернета Вещей	
		Запуск OpenFog	
2018	-	-	GPT

Рассмотрен вопрос конфиденциальности данных, а также методы ее обеспечения в искусственных нейронных сетях, развернутых в публичных распределенных системах (табл. 2).

С учетом цели и задач исследования в качестве методов обеспечения конфиденциальности выбраны гомоморфные шифры, среди которых были выделены три основных схемы: TFHE, CKKS и BFV.

Таблица 2 – Результаты аналитического обзора методов обеспечения конфиденциальности ИИ [6]

№	**Метод**	**Вычислительная сложность**	**Конфиденциальность при передаче данных**	**Конфиденциальность при обработке данных**	**Обеспечение надежности**	**Рассмотренный метод ИИ**
1	Модифицированный градиентный спуск	$O(D^2N + (D+\varepsilon)^3)$, где D – количество участников, ε – максимальный сдвиг набора данных, который может наблюдаться при добавлении или удалении одного участника	низкая	средняя	Отсутствует	AC-GAN
2	Асинхронный градиентный спуск	$O(D^2(N \cdot \delta) + D^3)$, где θ – отношение общего числа параметров СНС к скорости	низкая	средняя	Отсутствует	Сверточные нейронные сети

		выбора параметров СНС				
3	СРС Шамира	$O(N^2)$	высокая	низкая	Отсутствует	Глубокое обучение
4	СРС Асмута-Блума	$O(\log_2^2(N) + \log_2(N^2))$	высокая	низкая	Корректирующие коды СОК	Федеративное обучение
5	СРС Миньотта	$O(N^2)$	низкая	низкая	Корректирующие коды СОК	Федеративное обучение
6	BFV	$O(N \log N)$	высокая	высокая	Корректирующие коды СОК (в теории)	Искусственные нейронные сети, Сверточные нейронные сети
7	BGV	$O(N \log N)$	высокая	высокая	Отсутствует	Искусственные нейронные сети, Сверточные нейронные сети
8	CKKS	$O(N \log N)$	высокая	высокая	Корректирующие коды СОК (в теории)	Искусственные нейронные сети, Сверточные нейронные сети, Глубокое обучение

Во второй главе представлено исследование библиотек, программно реализующих схемы гомоморфного шифрования. Определены две библиотеки, наиболее подходящие для проведения дальнейших исследований: TenSEAL и Concrete-ML. Данные библиотеки реализованы на языке высокого уровня Python, предоставляют инструментарий для работы с гомоморфными шифрами и их внедрения в искусственные нейронные сети.

Исследованы методы матричного умножения, сохраняющего конфиденциальность. Сформулированы и доказаны теоремы об умножении зашифрованной матрицы на открытую и об аппроксимации функции искусственной нейронной сетью.

Теорема 1

Если матрица A зашифрована криптографической полностью гомоморфной схемой, а матрица B представлена в открытом виде, тогда умножение этих матриц

$$A \cdot B = \sum_{k=0}^{n-1} \left(\phi^k \circ \sigma(A)\right) \odot \left(\psi^k \circ \tau(B)\right),$$

принимает вид умножения матрицы на скаляр

$$A \cdot B = \sum_{k=0}^{n-1} \left(\phi^k \circ \sigma(A)\right) \odot \left(B^{k^T}\right),$$

где B^{k^T} – компонентный вектор транспонированной матрицы B. В таком случае количество операций сложения составляет n^2, количество операций умножения n^2, количество операций вращения n, а сложность уменьшается с $O(n^4)$ до $O(n^2)$.

На основе полученной теоремы предложена модификация гомоморфного умножения матрицы зашифрованных конфиденциальных входных данных на открытую матрицу параметров. Уменьшение пространственной сложности с $O(n^4)$ до $O(n^2)$ позволило сократить объем памяти в среднем в 7.89 раза для произведения матриц размера $n \times n$ и повысить быстродействие в среднем в 1.49 раза.

Проведено исследование функций активации, в результате которого была сформулирована и доказана следующая теорема.

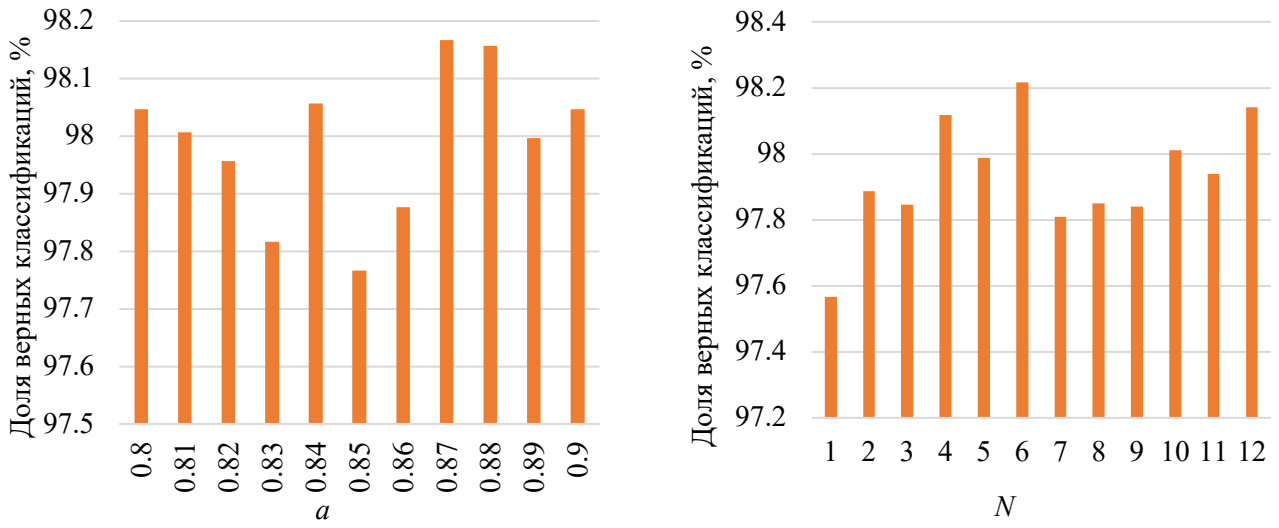
Теорема 2

Любую функцию $f(x)$, имеющую конечное число разрывов первого рода, можно аппроксимировать с помощью полиномов на ограниченной области с требуемой точностью, которая зависит от степени полинома аппроксимирующей функции $\hat{f}(x)$.

На основе полученной теоремы была выведена полиноминальная функция активации с обучаемыми коэффициентами (ФАОК).

$$p(x) = x^n \cdot \alpha_1 + x^{n-1} \cdot \alpha_2 + \cdots + x^1 \alpha_{n-1} + \alpha_n.$$

Проведено исследование методов дистилляции сверточных нейронных сетей, модифицирован метод дистилляции для задач классификации, который позволяет проводить дистилляцию для задач распознавания образов, определены оптимальные константа контроля ошибки, равная $a \approx 0.875$ (рис. 1а)), и размер коалиции сверточных нейронных сетей учителей, равный 6 (рис. 1б)), что позволило уменьшить размеры СНС более чем в 1609 раз при потере точности ±0.5-1%.



а) б)

Рисунок 1 – Зависимость качества обучения СНС с дистилляцией: а) от величины константы a; б) от количества СНС учителей в коалиции, где N – размер коалиции

Проведено исследование различных методов квантизации для построения сверточных нейронных сетей, сохраняющих конфиденциальность (СНССК), на базе схем BFV и CKKS. Разработан алгоритм конфиденциальной свертки (алг. 1).

Алгоритм 1. Конфиденциальный сверточный слой СНССК

Вход: $image_{enc}, kernel_{size}, stride, w_{filter}, b_{filter}, context$

Выход: $image_{enc_{conv}}$

$M, N \leftarrow size(image_{enc})$

Цикл от $i = 0$ **до** $M - 1$ **выполнять**

  Цикл от $j = 0$ **до** $N - 1$ **выполнять**

    $Temp_{mult_j} \leftarrow Mult_{Plain}\left(image_{enc_j}, w_{filter_i}, context\right)$

  Конец цикла

Конец цикла

Цикл от $i = 0$ **до** $M - 1$ **выполнять**

  Цикл от $j = 0$ **до** $\log_2 kernel_{size} - 1$ **выполнять**

    $Temp_{rot_i} \leftarrow Rot(Temp_{mult_i}, j, context)$

    $Temp_{add_i} \leftarrow Add\left(Temp_{rot_i}, Temp_{mult_i}, context\right)$

    $image_{enc_{conv},i} \leftarrow Add_{Plain}(Temp_{add_i}, kernel_{size}, context)$

  Конец цикла

Конец цикла

Цикл от $i = 0$ **до** $N - 1$ **выполнять**

  $image_{enc_{conv}} \leftarrow Add_{Plain}(image_{enc_{conv},i}, b_{filter_i}, context)$

Конец цикла

Возвратить $image_{enc_{conv}}$

Кроме того, разработан алгоритм для реализации слоя прямого прохода (алг. 2).

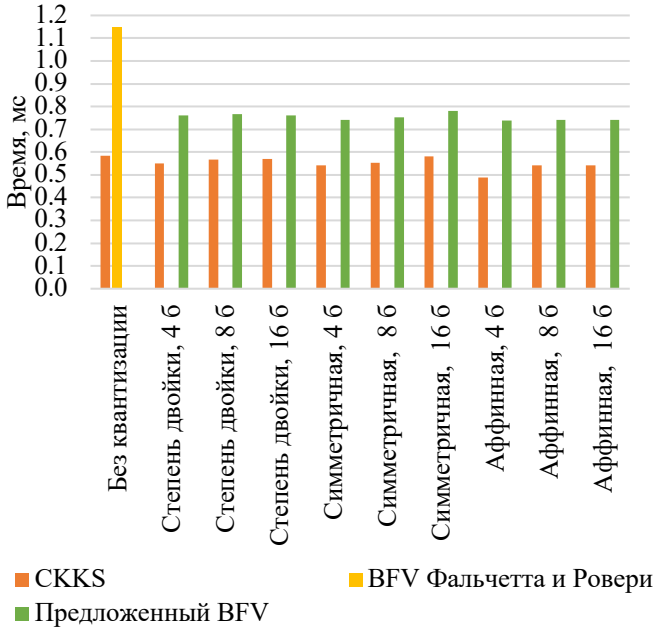
Алгоритм 2. Конфиденциальный прямой проход CHCCK

Вход: $image, w, b, contex$

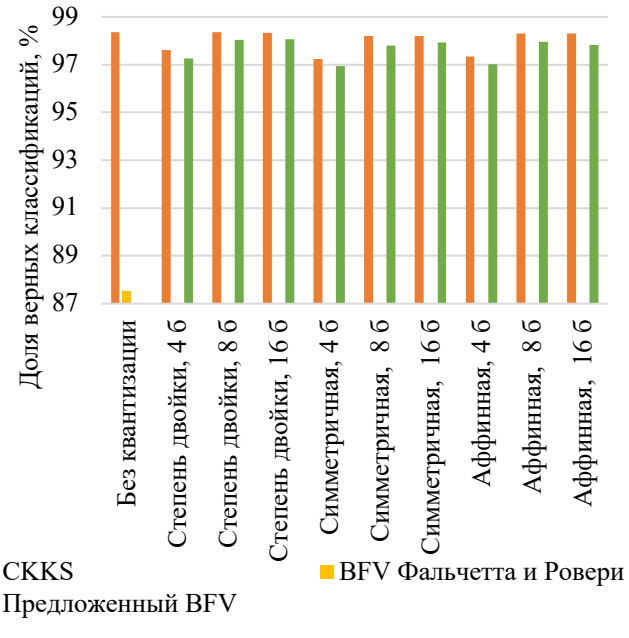
Выход: $image_{enc_{forward}}$

1: $P, Q \leftarrow w$
2: $Temp = 0$
3: $Temp_{rot} = 0$
4: **Цикл от** $i = 0$ **до** $P - 1$ **выполнять**
5: $Temp_{rot_i} \leftarrow Rot(image, \log_2 Q, context)$
6: **Цикл от** $j = 0$ **до** $Q - 1$ **выполнять**
7: $Temp_i \leftarrow Mult_{Plain}(Temp_{rot_i}, w_j, context)$
8: **Конец цикла**
9: **Конец цикла**
10: **Цикл от** $i = 0$ **до** $Q - 1$ **выполнять**
11: $image_{enc_{forward_i}} \leftarrow Add_{Plain}(Temp_i, b_i, context)$
12: **Конец цикла**
13: **Возвратить** $image_{enc_{forward}}$

Исследование показало, что метод квантизации 2^N дает наилучшее соотношение скорости обработки данных и точности распознавания изображений (рис. 2). Кроме того, исследование показало, что схема BFV проигрывает схеме CKKS и в точности распознавания изображений, и по скорости обработки данных.



а)



б)

Рисунок 2 – Экспериментальное сравнение квантованных моделей СНССК:
а) по времени работы; б) по точности распознавания

Такой результат можно объяснить высокой скоростью накопления внутренней ошибки, что мешает корректному срабатыванию функции активации. Таким образом, была установлена необходимость расширения исследования в сторону анализа схемы TFHE.

В третьей главе разработаны математические модели сверточных нейронных сетей, как открытых, так и обеспечивающих конфиденциальность, на базе нескольких приближенных функций активации. Это необходимо для более детального исследования моделей сверточных нейронных сетей, сохраняющих конфиденциальность, на базе гомоморфных схем CKKS и TFHE.

Построенные математические модели были исследованы в два этапа и на двух наборах данных. Первый этап – исследование моделей сверточных нейронных сетей, сохраняющих конфиденциальность, без применения дистилляции, второй этап – с применением дистилляции. В качестве обучающих наборов данных были выбраны MNIST с изображениями размером 28×28 и 8×8. На Рисунках 3-4 представлены модели СНС-учителей,

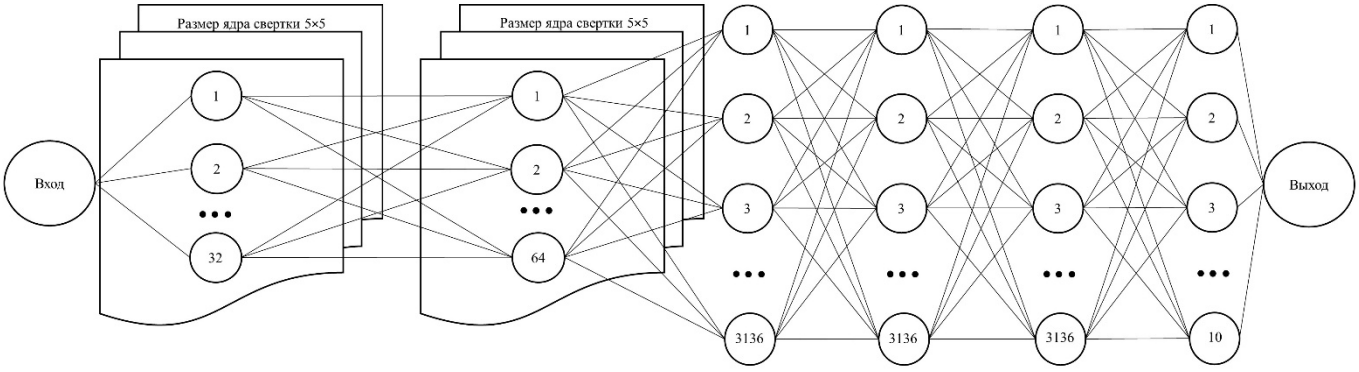


Рисунок 3 – Модель СНС-учителя для изображений размером 28×28

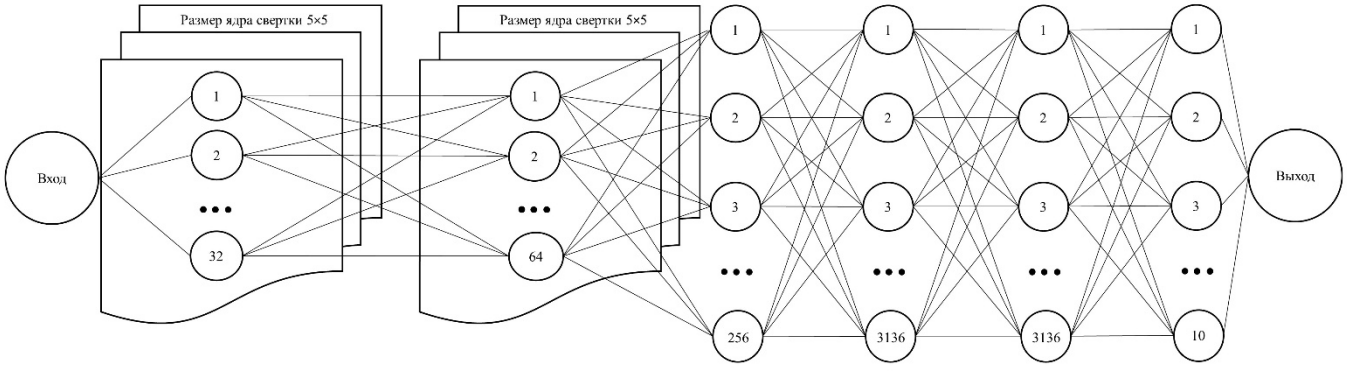


Рисунок 4 – Модель СНС-учителя для изображений размером 8×8

а на Рисунках 5-6 модели СНС-учеников.

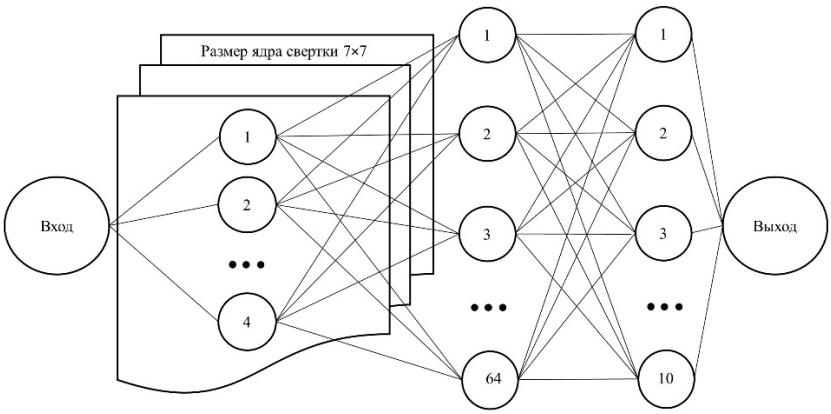


Рисунок 5 – Модель СНС-ученика для изображений размером 28×28

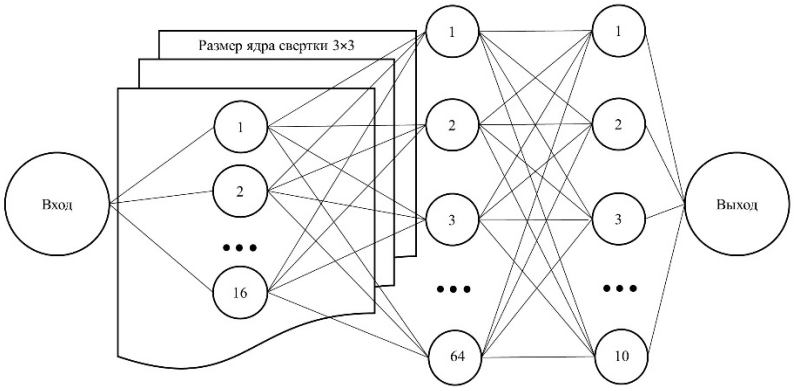


Рисунок 6 – Модель СНС-ученика для изображений размером 8×8

Получены полиноминальные аппроксимации различных функций активации. Для приближенной сигмоидной функции активации (ПС) известен полином наилучшего приближения третьей степени следующего вида:

$$\hat{\sigma}(x) = -0.004 \cdot x^3 + 0.197 \cdot x + 0.5.$$

Найдены коэффициенты для приближенных полиноминальных аппроксимаций функции активации ReLU (ПР) различных степеней

$$\hat{y}^2 = 0.0464108 \cdot x^2 + 0.5 \cdot x + 0.946969,$$

$$\hat{y}^4 = -3.983725 \cdot 10^{-4} \cdot x^4 + 0.0812353 \cdot x^2 + 0.5 \cdot x + 0.591856.$$

Найдены коэффициенты для приближенных полиноминальных аппроксимаций функции активации ELU (ПЕ) различных степеней

$$\hat{y}^2 = 0.0447007 \cdot x^2 + 0.5728 \cdot x + 0.554538,$$

$$\hat{y}^3 = -0.000949 \cdot x^3 + 0.0447007 \cdot x^2 + 0.630903 \cdot x + 0.554538,$$

$$\hat{y}^4 = -3.40065 \cdot 10^{-4} \cdot x^4 - 9.493416 \cdot 10^{-4} \cdot x^3 +$$

$$+ 0.074428 \cdot x^2 + 0.630903 \cdot x + 0.2514,$$

$$\hat{y}^5 = 1.99471 \cdot 10^{-5} \cdot x^5 - 3.40065 \cdot 10^{-4} \cdot x^4 -$$

$$- 3.209105 \cdot 10^{-3} \cdot x^3 + 0.074428 \cdot x^2 + 0.680269 \cdot x + 0.2514.$$

Для ФАОК в ходе обучения СНС были получены следующие полиномы второй степени:

для сверточного слоя

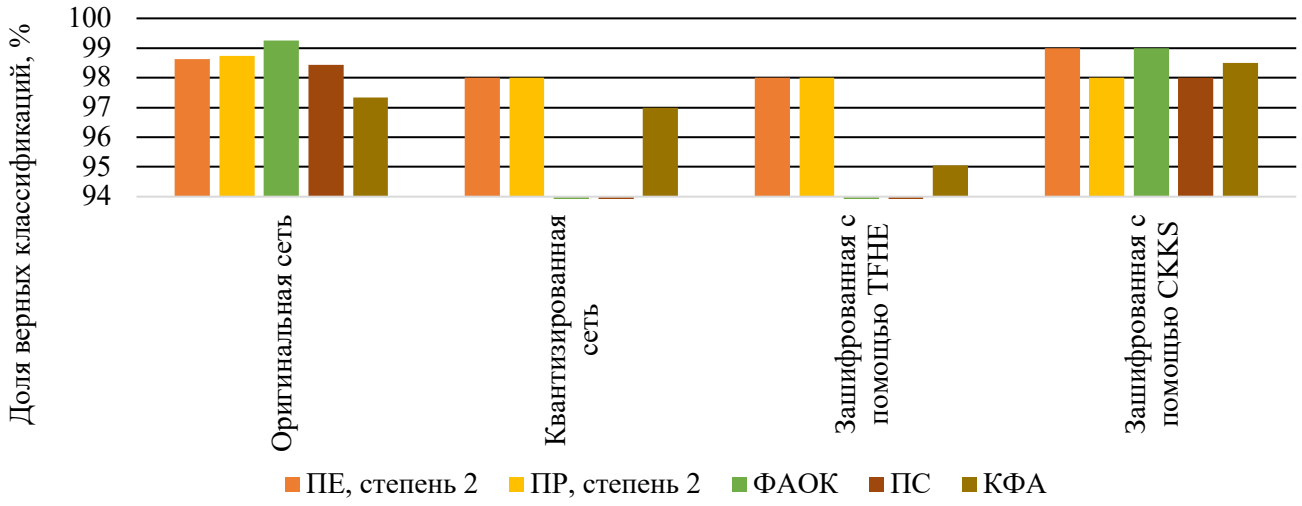
$$y_{\text{ФАОК}_{8\times8}}(x) = -0.0588202 \cdot x^2 + 0.0128461 \cdot x + 0.0204959,$$

$$y_{\text{ФАОК}_{28\times28}}(x) = -0.003377 \cdot x^2 - 0.020059 \cdot x - 0.145558,$$
$$y_{\text{ФАОК\_}dist_{8\times8}}(x) = -0.058766 \cdot x^2 + 0.013047 \cdot x + 0.020398,$$
$$y_{\text{ФАОК\_}dist_{28\times28}}(x) = -0.002223 \cdot x^2 + 0.094038 \cdot x + 1.648167,$$

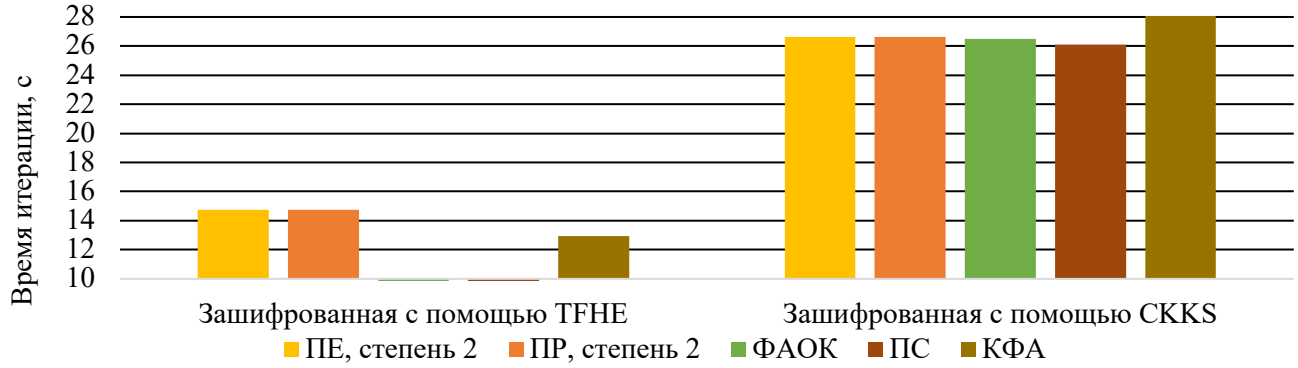
для скрытого слоя

$$y_{\text{ФАОК}_{8\times8}}(x) = -0.0531935 \cdot x^2 - 0.0466886 \cdot x + 0.0349889,$$
$$y_{\text{ФАОК}_{28\times28}}(x) = -0.004618 \cdot x^2 + 0.033658 \cdot x + 3.277034,$$
$$y_{\text{ФАОК\_}dist_{8\times8}}(x) = -0.053145 \cdot x^2 - 0.046645 \cdot x + 0.034919,$$
$$y_{\text{ФАОК\_}dist_{28\times28}}(x) = 0.029810 \cdot x^2 - 0.089146 \cdot x - 2.771453.$$

Таким образом, на основе представленных моделей (рис. 5-6) и найденных полиномиальных функций активации было проведено моделирование полученных моделей СНС, на базе набора данных MNIST с изображениями размером 28×28 и 8×8. В результате исследования было установлено, что в большинстве случаев схема TFHE обрабатывает данные быстрее в среднем на 50%, однако CKKS имеет более высокую точность распознавания, в среднем на 5% (рис. 7-10).



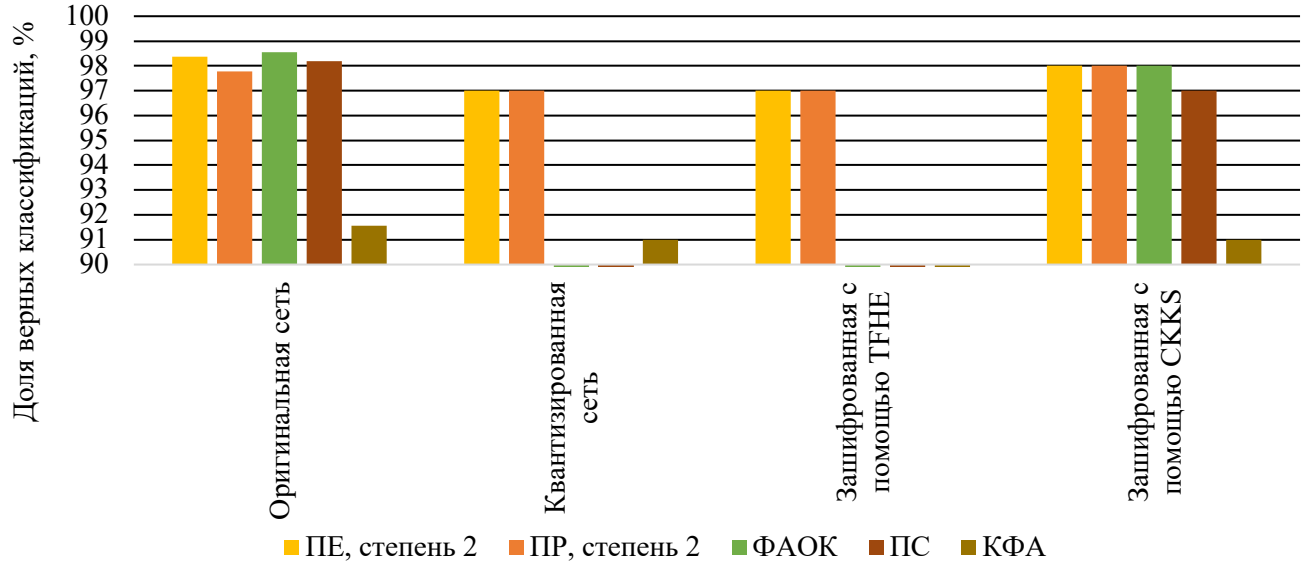
а)



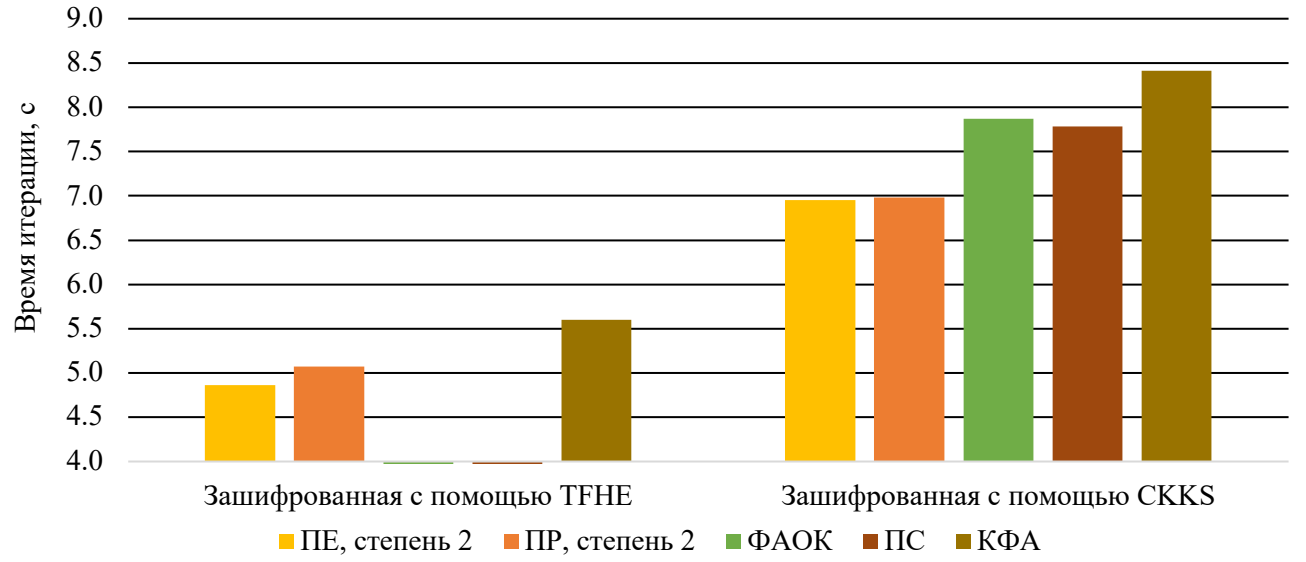
б)

Рисунок 7 – Результаты моделирования СНС(СК) с дистилляцией на базе полученных функций активации для изображений размера 8×8: а) время работы; б) точность распознавания

Если рассматривать соотношение скорости обработки данных и точности распознавания изображений, то модели на базе схемы TFHE имеют наилучший результат при применении дистилляции и работе с наборами данных размера 8×8, модели на базе CKKS напротив дают наилучший результат при распознавании изображений размера 28×28. Получены результаты применения приближенных функций активации: модели на базе CKKS показывают наилучшие результаты, в то время как модели на базе TFHE зачастую не могут обеспечить точность выше 90% на заданных параметрах. Лучший результат распознавания был получен на основе функции активации с обучаемыми коэффициентами в моделях на базе CKKS.



а)



б)

Рисунок 8 – Результаты моделирования СНС(СК) с дистилляцией на базе полученных функций активации для изображений размера 28×28: а) время работы; б) точность распознавания

Таким образом, в ходе проведения исследования были получены оптимальные модели сверточных нейронных сетей, сохраняющих конфиденциальность, для различных наборов данных с применением группы функций активации, обеспечивающих требуемое соотношение точности и скорости обработки данных. Полученные в ходе исследования результаты были использованы при разработке программного комплекса для проектирования сверточных нейронных сетей, сохраняющих конфиденциальность (ПКП СНССК) (рис. 9).

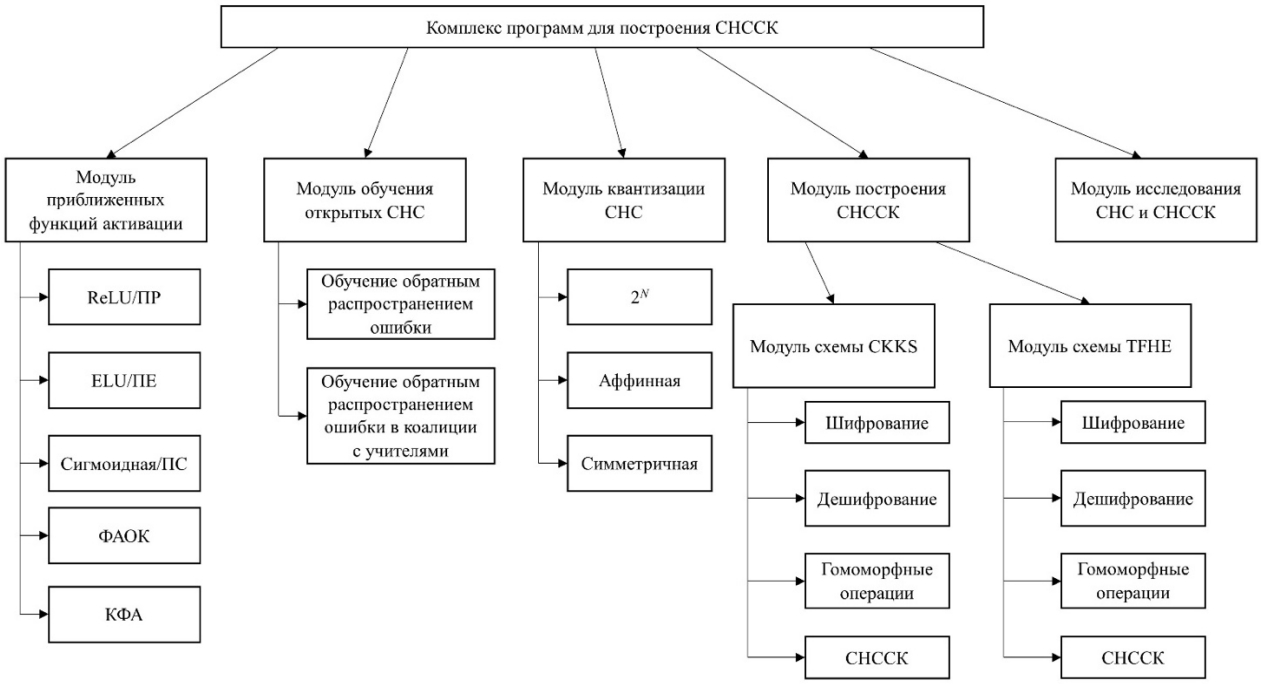


Рисунок 9 – Структурная схема работы ПКП

Разработанный программный комплекс потенциально позволит расширить применение моделей СНС за счет возможности обработки конфиденциальных данных в общедоступных облаках и туманных вычислениях. Основной функционал модуля вызова функций активации заключается в реализации приближенных функций активации для их последующего использования в моделях СНС и СНССК (рис. 10).

СНС или СНССК вызывает необходимую функцию активации, модуль вычисляет и возвращает либо ее значение, либо значение соответствующего приближающего полинома требуемой степени (если функция активации поддерживает более чем одну степень). Таким образом, на выходе получаем значения вызываемой функции активации $y(x)$, где x -- значения на выходах слоя, после которого функция вызывается. Модуль обучения открытых СНС реализует обучение открытых СНС алгоритмом обратного распространения ошибки как для случая с дистилляцией, так и без нее.

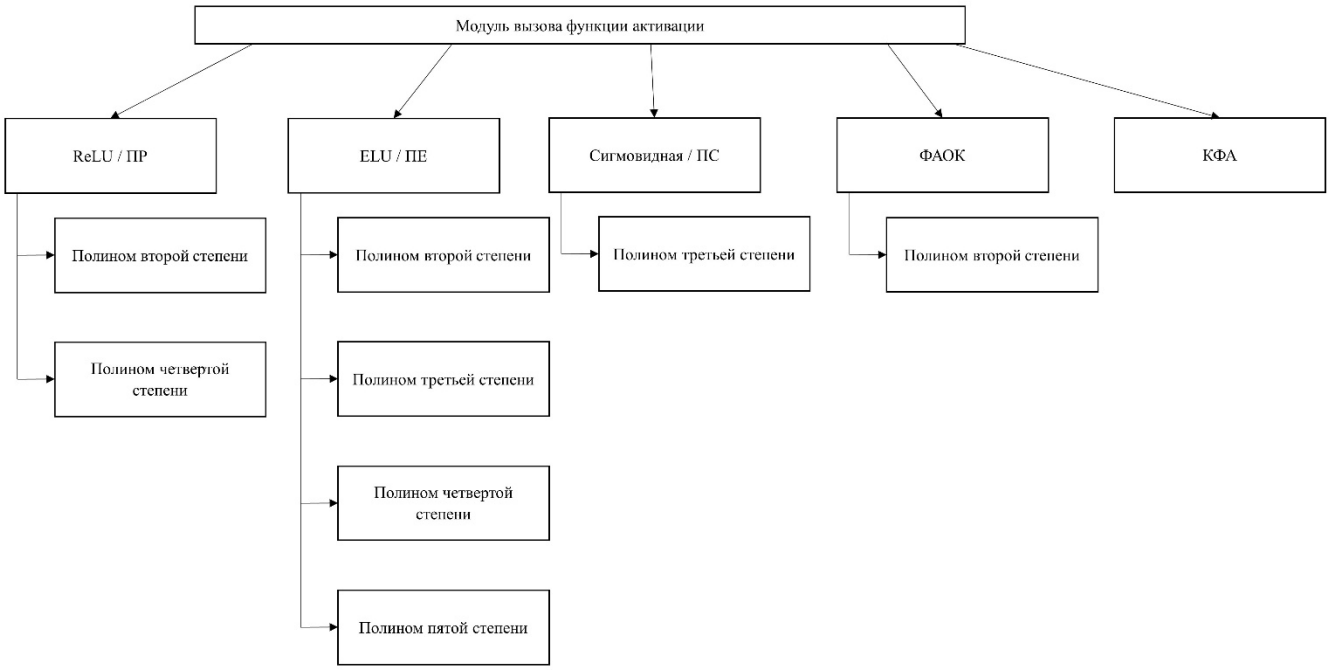


Рисунок 10 – Структурная модуля вызова функций активации

В нем так же инициализируются модели открытых сетей, обучаются СНС учителей. Модуль квантизации СНС применяется для квантизации СНС. Квантизация необходима для работы схемы ПГШ TFHE, так как она не поддерживает обработку вещественных значений. Модуль построения СНССК принимает открытую модель СНС, такие параметры слоев СНС как веса и смещения остаются открытыми. При работе СНССК зашифрованные входные данные проходят через слои СНС с открытыми параметрами и обрабатываются. Такой подход позволяет уменьшить избыточность и повысить скорость обработки, сохраняя конфиденциальность входных данных. Модуль поддерживает гомоморфные операции для двух схем ПГШ и позволяет проектировать модели СНССК на базе CKKS и TFHE. Модуль исследования СНС и СНССК используется для тестирования СНС и СНССК на точность распознавания и производительность, которая определяется как время, затраченное на одну итерацию нейронной сети. Для получения статистически достоверного результата, как точность, так и производительность рассчитываются для всей тестовой подвыборки набора, т.е. выполняется 10000 итераций. Модуль работает как с открытыми СНС с вещественными и целочисленными (квантованными) параметрами, так и с СНССК, использующими как схему ПГШ TFHE, так и схему CKKS.

Основные результаты и выводы по работе

Работа посвящена исследованию методов и алгоритмов, необходимых для построения нейронных сетей, сохраняющих конфиденциальность с использованием ПГШ. Основным сдерживающим фактором для практического использования ПГШ для построения нейронных сетей, сохраняющих конфиденциальность является его

высокая вычислительная сложность, складывающаяся из сложностей проблемных операций, таких как матричное умножение и определение знака закодированного числа, ограничений, накладываемых параметрами шифрования. Основные полученные и представленные в работе результаты исследования можно сформулировать следующим образом:

1. Разработан метод матричного умножения, который повышает быстродействие за счет уменьшения количества гомоморфных операций при сохранении конфиденциальности данных пользователя, уменьшая вычислительную сложность алгоритма с $O(n^4)$ до $O(n^2)$.

2. Модифицирован метод дистилляции, который использует коалицию учителей для уменьшения размеров целевой сети, что влечет за собой сокращение потребления памяти примерно в 1500 раз, уменьшение времени обработки данных в среднем в 30 раз, при уменьшении доли верных классификаций от 0.5% до 1%.

3. Выполнена адаптация методов квантизации, которая позволяет применять целочисленные и логические схемы ПГШ для построения нейронных сетей, сохраняющих конфиденциальность с большей эффективностью с точки зрения потребления памяти более чем в 3.5 раза и уменьшение времени обработки данных на 20%.

4. Разработана ФАОК, позволяющая повысить долю верных классификаций в среднем на 1.5% по сравнению с функциями, реализованными в библиотеках TenSEAL и Concete-ML.

5. Разработан комплекс программ и алгоритмов для проектирования и исследования приближенных функций активации позволяющий построить модель под конкретную задачу, повысить ее эффективность, а также расширить область применения для решения прикладных задач, требующих сохранения конфиденциальности.

Предложенная математическая модель, методы и алгоритмы позволяют не только расширить спектр задач, решаемых нейронными сетями на удаленных облачных сервисах, но и повысить эффективность решений с точки зрения вычислительной сложности и потребления памяти за счет оптимизации нейронных сетей и математического аппарата обработки конфиденциальных данных.

Публикации по теме диссертации

Статьи автора в журналах, рекомендованных ВАК РФ, Scopus, Web of Science

1. Shiriaev E. Reliablity and Security for Fog Computing Systems / E. Shiriaev, T. Ermakova, E. Bezuglova, [et al.] // Information. – 2024. – Vol. 15, no. 6. – P. 317.
2. Shiriaev E. A survey on multi-cloud storage security: threats and countermeasures / E. S. Bezuglova, E. M. Shiriaev, M. G. Babenko, [et al.] // Computational Technologies. – 2023. – Vol. 28, no. 1. – P. 72-80.

3. Shiriaev E. DT-RRNS: Routing protocol design for secure and reliable distributed smart sensors communication systems / A. Gladkov, E. Shiriaev, A. Tchernykh, [et al.] // Sensors. – 2023. – Vol. 23, no. 7. – P. 3738.
4. Shiriaev E. Comparative analysis of homomorphic encryption algorithms based on learning with errors / M. G. Babenko, E. I. Golimblevskaia, E. M. Shiriaev // Труды института системного программирования РАН. – 2020. – Т. 32, № 2. – С. 37-51.
5. Shiriaev E. A Comparative Study of Secure Outsourced Matrix Multiplication Based on Homomorphic Encryption / M. Babenko, E. Golimblevskaia, A. Tchernykh, E. Shiriaev, [et al.] // Big Data and Cognitive Computing. – 2023. – Vol. 7, no. 2. – P. 84.
6. Shiriaev E. Analytical Review of Confidential Artificial Intelligence: Methods and Algorithms for Deployment in Cloud Computing / E. M. Shiriaev, A. S. Nazarov, N. N. Kucherov, & M. G. Babenko// Programming and Computer Software. – 2024. – Vol. 50, no. 4. – P. 304-314.
7. Shiriaev E. High-Speed Convolution Core Architecture for Privacy-Preserving Neural Networks / M. A. Lapina, E. M. Shiriaev, M. G. Babenko, & I. Istamov // Programming and Computer Software. – 2024. – Vol. 50, no. 6. – P. 417-424.
8. Shiriaev E. Data Storage with Increased Survivability and Reliability Based on the Residue Number System / N. Kucherov, M. Babenko, E. Shiriaev, N. V. Hung //Advances in Systems Science and Applications. – 2024. – Vol. 24. – no. 02. – P. 166-186.
9. Shiriaev E. An efficient method for comparing numbers and determining the sign of a number in RNS for even ranges / A. Tchernykh, M. Babenko, E. Shiriaev, [et al.]. // Computation. – 2022. – Vol. 10, no. 2. – P. 17.

Другие публикации автора по теме диссертации

10. Ширяев Е.М. Разработка алгоритма конфиденциального поиска в распределенных системах / Е.М. Ширяев, Н.Н. Кучеров, О.В. Криволапова // Современная наука и инновации. – 2023. – № 2. – С. 10-19.
11. Shiriaev E. An Approach to Reducing Device Uncertainty in Fog-Cloud Computing / N. Kucherov, E. Shiriaev, D. Zolotariov, & S. Neelakandan // International Workshop on Advanced Information Security Management and Applications. – Springer, 2024. – P. 161-171.
12. Shiriaev E. Modification and Adaptation of Methods and Algorithms of the Active Security Concept for Fog Systems / E. Shiriaev, N. Kucherov, V. Movzalevskaya, & M. Khamidov // International Workshop on Advanced Information Security Management and Applications. – Springer, 2024. – P. 277-285.
13. Shiriaev E. Analytical Review of Cryptographic Primitives to Use in Fog Computing / E. Shiriaev, N. Kucherov, E. Bezuglova // International Workshop on Advanced Information Security Management and Applications. – Springer, 2024. – P. 267-276.

14. Shiriaev E. On the Way to Building Reliable and Secure Cloud-Based Data Processing Systems / N. Kucherov, E. Shiriaev, E. Bezuglova // International Workshop on Advanced Information Security Management and Applications. – Springer, 2024. – P. 152-160.
15. Shiriaev E. Load Balancing Methods for Distributed Data Storage: Challenges and Opportunities / E. Shiriaev // Current Problems of Applied Mathematics and Computer Systems. – Springer, 2024. – P. 95-104.
16. Shiriaev E. SNS-Based Secret Sharing Scheme for Security of Smart City Communication Systems / A. Gladkov, E. Shiriaev, A. Tchernykh, M. Deryabin, [et al.] // Ibero-American Congress of Smart Cities. – Springer, 2022. – P. 248-263.
17. Shiriaev E. Modeling Hyperchaotic Datasets for Neural Networks / E. Shiriaev, E. Bezuglova, N. Kucherov, & G. Value // International Conference on Mathematics and its Applications in new Computer Systems. – Springer, 2021. – P. 441-453.
18. Shiriaev E. Neural network method for base extension in residue number system / M. G. Babenko, E. Shiriaev, A. Tchernykh, [et al.] // International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2020). – CEUR, 2020. – P. 9-22.
19. Ширяев Е.М. Обзор программных реализаций методов гомоморфного шифрования / Е.И. Голимблевская, Е.М. Ширяев // Фундаментальные проблемы управления производственными процессами в условиях перехода к индустрии 4.0. – 2020. – С. 284.
20. Shiriaev E. Survey software implementations of homomorphic encryption methods / E. Golimblevskaia, E. Shiriaev, N. Kucherov // Advances in Automation II: Proceedings of the International Russian Automation Conference (RusAutoConf-2020). – Springer, 2021. – P. 601-613.
21. Shiriaev E. Efficient implementation of the CKKS scheme using a quadratic residue number system / E. M. Shiriaev, A. S. Nazarov, N. N. Kycherov, & N. A. Sotikova // 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). – IEEE. 2021. – P. 665-669.
22. Shiriaev E. One Plaintext Attack on the BFV Scheme / E. M. Shiriaev, A. S. Nazarov, N.A. Sotikova // 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). – IEEE. 2021. – P. 670-673.
23. Shiriaev E. Development of an Approach to Confidential Learning with Errors in the Design of Neural Networks / E. Bezuglova, & E. Shiriaev // AISMA-2023: International Workshop on Advanced Information Security Management and Applications. – Springer, 2024. – P. 24-30.
24. Shiryaev E. Performance impact of error correction codes in RNS with returning methods and base extension / E. Shiryaev, E. Bezuglova, M. Babenko, [et al.] //2021

International Conference Engineering and Telecommunication (En&T). – IEEE, 2021. – С. 1-5.

25. Shiryaev E. RRNS base extension error-correcting code for performance optimization of scalable reliable distributed cloud data storage / M. Babenko, A. Tchernykh, B. Pulido-Gaytan, J. M. Cortés-Mendoza, E. Shiryaev [et al.] //2021 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW). – IEEE, 2021. – С. 548-553.
26. Shiryaev E. Weighted two-levels secret sharing scheme for multi-clouds data storage with increased reliability / V. Miranda-Lopez, A. Tchernykh, M. Babenko, V. Kuchukov, M. Deryabin, E. Golimblevskaia, E. Shiryaev [et al.] //2019 International Conference on High Performance Computing & Simulation (HPCS). – IEEE, 2019. – С. 915-922.

Свидетельства о государственной регистрации программ

27. *Свидетельство о гос. регистрации программы для ЭВМ*. Программа моделирования сверточных нейронных сетей, сохраняющих конфиденциальность на туманных устройствах [Текст] / Е.М. Ширяев [и др.]; Ф. государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». – № 2025617506; заявл. 05.03.2025; опубл. 19.03.2025, 2025615824 (Рос. Федерация).
28. *Свидетельство о гос. регистрации программы для ЭВМ*. Программа построения сверточных нейронных сетей, сохраняющих конфиденциальность на туманных устройствах [Текст] / Е.М. Ширяев [и др.]; Ф. государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». – № 2025617046; заявл. 05.03.2025; опубл. 21.03.2025, 2025614617 (Рос. Федерация).
29. *Свидетельство о гос. регистрации программы для ЭВМ*. Программа, оптимизирующая использование нейронных сетей на туманных устройствах [Текст] / Е. М. Ширяев [и др.]; Ф. государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». – № 2025616724; заявл. 05.03.2025; опубл. 19.03.2025, 2025614357 (Рос. Федерация).