ОТЗЫВ

официального оппонента

доцента кафедры вычислительной техники и защиты информации
ФГБОУ ВО «Уфимский университет науки и технологий»,
доктора технических наук
Вульфина Алексея Михайловича
на диссертацию
Сулавко Алексея Евгеньевича
«Высоконадежная биометрическая аутентификация на основе защищенного исполнения нейросетевых моделей и алгоритмов искусственного интеллекта»,
представленную на соискание ученой степени доктора технических наук
по специальности 2.3.6. Методы и системы защиты информации,
информационная безопасность

Актуальность темы исследования

Представленная диссертационная работа посвящена проблеме повышения надежности биометрической аутентификации, защите от компрометации биометрических данных и защите биометрических систем от компьютерных атак с помощью методов доверенного искусственного интеллекта. На сегодняшний день тематика исследования пользуется большим вниманием научного сообщества и является актуальной.

В последние годы принято множество технологических инициатив и нормативных актов, направленных на регулирование стремительно развивающихся технологий искусственного интеллекта, обусловленных необходимостью цифровизации экономики и укрепления информационной безопасности страны (Указ Президента РФ от 10.10.2019 г. № 490, Федеральный закон № 572 (об осуществлении идентификации и аутентификации физических лиц с использованием биометрических персональных данных) и множество других руководящих и законодательных документов). С точки зрения содержания данных документов работа автора также является актуальной и своевременной. В исследовании детально проработаны вопросы, которые лежат в плоскости обеспечения информационной безопасности и с одной стороны касаются защиты данных от неавторизованного доступа с помощью биометрических систем аутентификации, а с другой стороны касаются методов построения доверенного искусственного интеллекта в целом, а также обучения и защиты подобных моделей и систем от компьютерных атак.

В работе также рассматриваются вопросы стандартизации технологий и систем искусственного интеллекта.

Таким образом, тема диссертационного исследования, посвященного решению научно-технической проблемы повышения надежности многофакторной

биометрической аутентификации и защищенности биометрических систем от компьютерных атак на основе технологии защищенного исполнения нейросетевых моделей и алгоритмов искусственного интеллекта, является актуальной и своевременной.

Оценка структуры и содержания работы

Диссертационная работа Сулавко А.Е. состоит из введения, 5 глав, заключения, списка сокращений, списка литературы и приложений. Диссертация содержит 391 страницу текста, 108 рисунков и 28 таблиц. Список литературы включает 362 источника.

Во введении обозначена актуальность исследования и дана общая характеристика работы.

В первой главе выполнено аналитическое исследование решаемой в работе проблемы, и проведен системный обзор научных публикаций, а также международных и национальных стандартов в области биометрических систем защиты, доверенного искусственного интеллекта, а также защиты подобных систем и технологий от состязательных атак. Описаны основные атаки на системы искусственного интеллекта и биометрические системы, рассматриваемые в рамках диссертации. Показано, что для решения обозначенных проблем требуется разработка моделей искусственных нейронов и сетей, позволяющих повысить длину ключа, а также снизить влияние концептуального дрейфа. Сформулированы цель и задачи исследования.

Во второй главе разработана и описана концепция защищенного исполнения нейросетевых моделей и алгоритмов искусственного интеллекта. Предложена модель корреляционных нейронов, которые анализируют силу связи между входными данными путем определения так называемой «точечной корреляции». Разработан алгоритм обучения корреляционного нейрона. Предложена модель нейросетевого преобразователя «биометрия-код» (НПБК) на основе корреляционных нейронов и алгоритм его обучения. НПБК создает устойчивую связь между биометрическим шаблоном пользователя и ключом шифрования, который может быть использован для аутентификации. При предъявлении биометрического образа в НПБК генерируется ключ. Модель обеспечивает конфиденциальность биометрического шаблона и криптографического ключа пользователя на всех этапах – от хранения и передачи по каналам связи до обработки данных. Представлена авторская теория об искривлении пространства признаков при наличии корреляции между ними. Это искривление признакового пространства позволяет получить дополнительную информацию о корреляционных связях, которую можно оценить, оперируя мета-признаками в пространстве Байеса-Минковского.

Третья глава целиком посвящена оценке влияния дрейфа биометрических данных на результаты классификации образов в биометрических системах,

и способах его устранению путем использования адаптивных моделей искусственного интеллекта, основанных на нейро-иммунном подходе к машинному обучению. Предложена нейро-иммунная модель и алгоритм ее обучения на малых выборках. Разработан алгоритм дообучения предложенной модели в процессе ее работы в реальной практике, т.е. когда пользователи используют модель для аутентификации и распознавания своих образов. Эффективность модели оценена в задаче классификации пользователей по их клавиатурному почерку на трех наборах данных, два из которых находятся в открытом доступе, один – соискатель сформировал самостоятельно. Тщательная экспериментальная проверка предложенных моделей и алгоритмов обуславливает высокую оценку полученных результатов.

Четвертая глава носит более прикладной характер. В ней рассматривается многофакторная и однофакторная аутентификация по голосу, подписи акустическому образу уха. Для анализа биометрических данных уха используются нейронные сети – автоэнкодеры, а для извлечения признаков голоса и подписи – классические алгоритмы анализа сигналов. Для классификации образов в пространстве извлеченных признаков используются разработанные в предыдущих главах модели НПБК и нейро-иммунной сети. Для обучения моделей применяются алгоритмы, разработанные в предыдущих главах. Предлагается алгоритм, комплексирующий признаки, и позволяющий обеспечить высокую общую эффективность и надежность системы аутентификации. Для проверки эффективности предложенных решений автор использовал достаточно объемные наборы данных, которые собрал самостоятельно.

В пятой главе приводится описание практической части работы, а именно разработанных программных моделей и технологии автоматического синтеза и обучения нейросетевых моделей доверенного ИИ, национального стандарта на ее основе и внедрения основных результатов, которых насчитывается восемь.

В заключении сформулированы основные результаты и выводы, подведены итоги работы, обозначены направления дальнейших исследований.

В приложениях представлены результаты дополнительных экспериментов, исходный код разработанного программного модуля, ряд документов, акты внедрения и использования разработанных программах средств, свидетельства о регистрации программ и патент.

Структура и содержание диссертации соответствует цели и поставленным задачам исследования. Диссертации отражает суть проделанной работы, а автореферат раскрывает содержание диссертации.

Тема диссертации *соответствует* специальности 2.3.6. Методы и системы защиты информации, информационная безопасность.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

Обоснованность научных положений и выводов, сформулированных в диссертационной работе, не вызывает сомнений, т.к. основана на проведенном автором всестороннем анализе большого количества трудов отечественных и зарубежных авторов, нормативных и правовых документов, корректной постановке задач, использовании общепризнанных научных методов и лучших практик, строгости применения используемого математического аппарата. Результаты диссертации опубликованы в 80 работах, из них 38 статей в рецензируемых журналах из перечня ВАК либо в изданиях, индексируемых в RSCI, 21 статья в изданиях, включенных в международные базы Web of Science или Scopus, 1 монография, 11 статей в других изданиях, получен патент и 8 свидетельств о регистрации программ. Автор принял участие во множестве международных и всероссийских конференциях. В целом количество публикаций полностью соответствует уровню докторской диссертации.

Автор системно подходил к решению поставленных задач, логично и последовательно описал полученные результаты. Выводы автора не противоречат известным фактам из практики, а также, не противоречат доказанным научным фактам и подкрепляются результатами экспериментов и обоснованиями. Автор использовал признанные методологии работы с данными. Все защищаемые положения работы корректны и обоснованы.

Диссертация соответствует пунктам 9, 12 и 15 паспорта специальности 2.3.6. Методы и системы защиты информации, информационная безопасность.

Новизна полученных результатов

Сулавко А.Е. предложил и использовал в своей работе ряд оригинальных подходов, а именно: автоматический синтез и обучение корреляционных нейронов и их сетей, нейро-иммунные модели, базирующиеся на предложенной архитектуре искусственных детекторов. Полученные автором результаты обладают высокой степенью оригинальности и научной новизной.

Научная новизна результатов диссертации заключается в том, что в ней получены следующие новые научные результаты:

1. Концепция защищенного исполнения нейросетевых алгоритмов ИИ, *позволяющая* обеспечить устойчивость моделей и алгоритмов ИИ к извлечению знаний в задачах классификации образов. В *отличие* от существовавших ранее концепций, предложенная автором концепция основана на преобразовании корреляционных связей между признаками в высокоинформативные мета-признаки Байеса-Минковского. В рамках работы установлено, что корреляция между признаками увеличивает количество информации об образе (один мета-признак Байеса-Минковского может содержать в 2-3 раза больше информации, чем со-

держится суммарно в паре исходных признаков, от которых он порожден), что повышает надежность распознавания образов.

2. Модель корреляционных нейронов и модель НПБК на их основе, *отличающиеся* тем, что они анализируют корреляционные связи между признаками вместо признаков, а также робастной алгоритм автоматического синтеза и обучения этих моделей на малых выборках. Это *позволяет* повысить защищенность биометрических данных от компрометации, длину ключа, связываемого с биометрическими образами субъектов, и устойчивость систем биометрической аутентификации к состязательным атакам.

3. Адаптивная нейро-иммунная модель ИИ, *отличающаяся* от существовавших ранее моделей предложенной гибкой архитектурой искусственных иммунных детекторов (антител и клеток памяти). Также отличительной особенностью является использование в основе детекторов ядерных функций, сочетание ансамблевых методов машинного обучения и метода обучения с подкреплением, что *позволяет* ей устойчиво обучаться на малых выборках и адаптироваться к изменению биометрических данных в процессе функционирования. Предложенные нейро-иммунная модель и алгоритмы ее обучения *в отличие* от существовавших ранее *позволяют* снизить влияние концептуального дрейфа и вероятность ошибок биометрической аутентификации, даже если исходная обучающая выборка недостаточно репрезентативна или незначительна в объеме.

4. Методы и алгоритм высоконадежной многофакторной биометрической аутентификации на базе НПБК и нейро-иммунной модели, *отличающиеся* использованием новых акустических биометрических параметров, характеризующих внутреннее строение ушного канала, комплексированием динамических и статических признаков с учетом их приоритезации, информативности и стабильности. Это *позволяет* обеспечить более высокую надежность аутентификации, робастность дрейфующих характеристик, защиту биометрических образов от компрометации, снизить вероятность ошибок «ложного допуска» и «ложного отказа» по сравнению с известными аналогами.

5. Технология синтеза нейросетевых моделей доверенного ИИ, которая *позволяет* снизить объем тренировочной выборки, повысить надежность и защищенность биометрических систем аутентификации и других приложений ИИ. Предложенная технология отличается от существовавших ранее наличием режимов автоматического обучения нейросетевых моделей ИИ, защищенного исполнения нейросетевых алгоритмов классификации образов и применением процедур автоматической оценки информативности признаков.

Достоверность полученных результатов

Достоверность результатов работы подтверждается корректным использованием положений и основных понятий российских ГОСТов в области ИБ, методов исследования и проведения экспериментов, признанных методик

статистической обработки данных. Автор применял методы машинного обучения, распознавания образов, анализа сигналов, теории информации, теории вероятностей и математической статистики, использовал искусственные нейронные сети и иммунные систем. Натурные эксперименты проводились с учетом необходимых требований, вычислительные эксперименты выполнялись в соответствии с принятыми практиками, при этом использованы наборы данных достаточно большого объема, что говорит о высокой достоверности результатов исследования. Также достоверность результатов подтверждается актами внедрения, которых насчитывается восемь. В совокупности можно сказать, что все представленные в работе результаты являются достоверными.

Теоретическая и практическая значимость полученных автором результатов

В соответствии с доктриной информационной безопасности РФ к одной из ключевых задач сейчас относится наращивание технологического задела, позволяющего нашей стране отражать все угрозы, используя отечественный стек технологий информационной безопасности. В связи с этим можно отметить значительную научную и практическую значимость, а также оригинальность полученных автором результатов, так как предлагаемые им решения (концепция, модели, методы, алгоритмы и технология) базируются на математических основах, которые разрабатывались только Омской научной школой.

Стандарт, разработанный научной группой Омского государственного технического университета под руководством соискателя и базирующийся на результатах его диссертационного исследования, позволяет сохранить суверенитет страны в отношении защиты искусственного интеллекта от широкого класса угроз информационной безопасности.

Представленный метод аутентификации дает высокую надежность: $FRR = 0{,}12$ при $FAR < 10^{-14}$. Это представляет ценность для практики.

На базе результатов диссертационной работы Сулавко А.Е. или с их непосредственным применением разработано множество программных продуктов, о чем свидетельствуют акты использования результатов диссертации, что также говорит о высокой практической значимости работы.

Теоретическая значимость результатов исследования, прежде всего, заключается в том, что проработаны математические основы для создания эффективных алгоритмов автоматического синтеза и обучения нейросетевых моделей искусственного интеллекта с использованием малых выборок данных, при этом много внимания уделено именно архитектурным особенностям таких моделей. Кроме того, в работе заложены основы научного направления, связанного с защищенным исполнением нейросетевых моделей и алгоритмов искусственного интеллекта.

Замечания по диссертационной работе

1. В работе не приведено обоснование использования нейросетевых автоэнкодеров с целью извлечения признаков из акустических образов уха. Возможно, схожие или даже лучшие результаты можно было получить, используя обычные методы спектрального анализа сигналов или иные модели. Причем, при анализе голосовых сигналов и подписей извлечение признаков выполнено с помощью классических алгоритмов анализа без использования автоэнкодеров, что также недостаточно подробно раскрыто описании предлагаемого решения.

2. В работе уделяется внимание оценке влияние психофизиологического состояния человека на возникновение дрейфа биометрических признаков (параграф 3.1). Однако помимо психофизиологического состояния человека на изменчивость динамических биометрических образов существенное влияние оказывает время. Данному аспекту в работе не уделено достаточного внимания. В связи с этим возникает вопрос: как и насколько учтен фактор времени?

3. В работе активное термин «дрейф» употребляется повсеместно. Однако понятие «дрейф» весьма многогранно, и границы между различными значениями данного понятия в работе выглядят размыто. Имеется рад вопросов, связанных с употреблением термина «дрейф»:

- какой именно дрейф будет наблюдаться в случае изменения психофизиологического состояния субъекта – дрейф концепции или данных?
- какой тип дрейфа связан с фактором изменения биометрических образов с течением времени?

4. В работе делается акцент на серию стандартов ГОСТ Р 52633. Однако для оценки качества работы биометрических систем существуют международные стандарты, в частности, стандарты ISO/IEC серии 19795. Часть этих стандартов гармонизирована для использования в России. Почему основной акцент делается именно на серию ГОСТ Р 52633?

5. В главе 5 речь идет об архитектуре и функциональных особенностях системы AIC ModelOps Platform. На рис. 5.6 указаны существующие бизнес-процессы, которые могут быть реализованы с использованием данной платформы. Однако далее представлена декомпозиция далеко не всех бизнес-процессов, которые фигурируют на рис.5.6. На рис. 5.7 и 5.8 есть BPMN диаграммы только для анализа результатов экспериментов, проведения НИР и управления командой, это не полный перечень бизнес-процессов. Также один из бизнес-процессов имеет разные названия на разных рисунках: «Планирование НИР» (рис. 5.6) и «Проведение НИР» (рис. 5.8).

Заключение

Выявленные недостатки работы, тем не менее, не снижают общей высокой оценки научной ценности и практической значимости и положительного впечатления о ней.

В работе решена актуальная научная проблема, которая заключается в повышении надежности многофакторной биометрической аутентификации и защищенности биометрических систем от компьютерных атак на основе технологии защищенного исполнения нейросетевых моделей и алгоритмов искусственного интеллекта. Внедрение результатов, полученных автором работы, вносит существенный вклад в развитие методологий биометрической аутентификации, создания систем доверенного искусственного интеллекта, а также значимый вклад в развитие теоретических основ методов машинного обучения в целом.

Диссертация Сулавко А.Е. является завершенной научно-квалификационной работой, которая выполнена на актуальную тему и содержит результаты, значимые для практики и теории, а также обладающие научной новизной. Диссертационная работа соответствует требованиям п. 9 «Положения о присуждении ученых степеней» ВАК, предъявляемым к докторским диссертациям, а ее автор – Сулавко Алексей Евгеньевич – заслуживает присуждения учёной степени доктора технических наук по специальности 2.3.6. Методы и системы защиты информации, информационная безопасность.

Официальный оппонент:
доктор технических наук,
доцент кафедры вычислительной
техники и защиты информации
ФГБОУ ВО «Уфимский университет
науки и технологий»

Вульфин Алексей Михайлович

Докторская диссертация защищена по специальности:
2.3.6. Методы и системы защиты информации, информационная безопасность

Даю согласие на обработку персональных данных.

Адрес места основной работы:
ФГБОУ ВО «Уфимский университет науки и технологий»
450076, Российская Федерация, Республика Башкортостан, г. Уфа, ул. Заки Валиди, д. 32
Телефон: 8 (347) 273-06-72
E-mail: vulfin.alexey@gmail.com

Подпись Вульфина А.М.
Удостоверяю «28» 03 2023 г.
Начальник общего отдела УУНиТ Рахимова Л.Ф.