ISSN 2658-5014 (Print) СИИТ ISSN 2686-7044 (Online)
2024. Т. 6, № 2 (17). С. 11–32 http://siit.ugatu.su
СИСТЕМНАЯ ИНЖЕНЕРИЯ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
Научный обзор

УДК 004.852 DOI 10.54708/2658-5014-SIIT-2024-no2-p11 EDN VNIEEV

ВЫСОКОНАДЕЖНАЯ БИОМЕТРИЧЕСКАЯ АУТЕНТИФИКАЦИЯ НА ОСНОВЕ ЗАЩИЩЕННОГО ИСПОЛНЕНИЯ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ И АЛГОРИТМОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

А. Е. Сулавко

Аннотация. В статье представлены результаты исследования, посвященного решению научно-технической проблемы повышения надежности многофакторной биометрической аутентификации и защищенности биометрических систем от компьютерных атак. Объект исследования – системы биометрической аутентификации на основе методов, моделей и алгоритмов доверенного ИИ. Предмет исследования – нейросетевые модели и алгоритмы машинного обучения на малых выборках для высоконадежной биометрической аутентификации и защиты биометрических данных от компрометации. Цель работы – повысить надежность многофакторной биометрической аутентификации на основе защищенного исполнения нейросетевых моделей доверенного ИИ и алгоритмов их автоматического синтеза и обучения на малых выборках биометрических данных. Для достижения цели были выполнены следующие задачи: 1. Разработка концепции защищенного исполнения нейросетевых алгоритмов ИИ. 2. Разработка моделей искусственных нейронов и нейросетевого преобразователя биометрия-код, потенциально устойчивых к деструктивным воздействиям, и алгоритмов их робастного автоматического обучения на малых выборках. 3. Разработка адаптивной модели ИИ и алгоритмов ее обучения, позволяющих предупредить или снизить влияние концептуального дрейфа данных в системах биометрической аутентификации. 4. Разработка методов многофакторной аутентификации на базе тайных биометрических образов с обеспечением конфиденциальности биометрических данных. 5. Разработка технологии автоматического синтеза и обучения нейросетевых моделей для высоконадежной многофакторной биометрической аутентификации.

Ключевые слова: защищенное исполнение искусственного интеллекта; нейросетевые преобразователи биометрия-код; биометрическая аутентификация; биометрические параметры голоса; особенности воспроизведения рукописных паролей; эхограммы ушного канала; корреляция между признаками; корреляционные нейроны; автоматическое машинное обучение; обучение с подкреплением.

Введение

Сегодня мировой рынок биометрии проходит фазу активного роста (по данным MarketsAndMarkets к 2025 г. его объем составит 68 млрд долл.). Биометрические системы внедряются повсеместно: на объектах критической информационной инфраструктуры [1], в банковской сфере, государственном секторе (более 80 стран используют биометрические паспорта), в сфере управления транспортом и городом. Рост рынка биометрических систем обусловлен новыми тенденциями и вызовами, с которыми столкнулись общество и государство: увеличение объемов данных о действиях пользователей в сети Интернет, которые могут быть использованы в злоумышленных целях (обострились проблемы приватности, анонимности пользователей и защищенности биометрических шаблонов от компрометации); применение технологий искусственного интеллекта (ИИ) для реализации хакерских атак, дезинформации, мошенничества, фальсификации биометрических образов человека (например, при помощи deepfake, голосовых синтезаторов); замена традиционных биометрических образов отпечатка пальца на более удобные образы голоса, лица и др., пригодные для бесконтактной аутентификации, но в большей степени подверженные дрейфу (изменчивости). В связи с этим современная высоконадежная биометрическая система должна строиться на основе доверен-

ного ИИ, устойчивого к деструктивным факторам (дрейф биометрических данных, компьютерные атаки) и обладающего поддержкой защищенного режима исполнения. Под «защищенным исполнением» понимается невозможность анализа логики работы ИИ, управления ИИ и извлечения знаний из памяти ИИ любым неавторизованным субъектом.

Настоящее исследование посвящено решению научно-технической проблемы, 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.

Степень научной разработанности темы и обсуждение решаемой задачи

На данный момент действуют ряд международных стандартов, связанных с вопросами защиты биометрических систем от компьютерных атак (ISO/IEC 19792:2009, ISO/IEC 24761:2019, ISO/IEC 24745:2022, ISO/IEC 30107). Однако эти стандарты не позволяют устранить ряд актуальных угроз (извлечение знаний моделей ИИ, компрометация открытых биометрических образов, состязательные атаки). В России действует серия национальных стандартов ГОСТ Р 52633, не имеющих международных аналогов. Стандарты ГОСТ Р 52633 регламентируют особенности разработки, обучения и тестирования систем высоконадежной биометрической аутентификации, которые должны строиться на базе нейросетевых преобразователей биометрия – код (НПБК), позволяющих связать криптографический ключ или пароль пользователя с его биометрическим образом. Тем не менее из-за наличия ряда недостатков применимость данных стандартов ограничена (высокая вероятность ошибок, малая длина ключа, подверженность атакам).

В мировой практике сложилось несколько подходов к повышению надежности биометрических систем аутентификации с обеспечением конфиденциальности биометрических данных, которые основаны на использовании нечетких экстракторов, искусственных нейронных сетей, искусственных иммунных систем, применении шифрования (в том числе гомоморфного). Развитию аппарата искусственных нейронных сетей и искусственных иммунных систем, а также вопросам создания доверенного ИИ посвящены работы многих ведущих российских и зарубежных ученых. Среди них Ю. А. Брюхомицкий, А. М. Вульфин, С. В. Гарбук, А. И. Галушкин, А. И. Иванов, И. В. Котенко, С. И. Николенко, B. Baker, Y. Bengio, L. N. De Castro, C. Fung, J. Greensmith, G. E. Hinton, V. Kurkova, Y. LeCun, P. K. Mishra, R. E. Schapire, K. O. Stanley, J. Timmis и др. Вопросам высоконадежной биометрической аутентификации, оценки изменчивости биометрических параметров, обеспечения конфиденциальности биометрических данных, а также защите биометрических систем от компьютерных атак посвятили множество своих работ Б. С. Ахметов, Л. К. Бабенко, А. В. Безяев, В. И. Васильев, В. И. Волчихин, Б. Н. Епифанцев, А. В. Еременко, А. И. Иванов, А. С. Катасёв, П. С. Ложников, Г. Б. Маршалко, T. H. Akkermans, F. O. Catak, Y. Dodis, F. Hao, L. G. Hafemann, A. K. Jain, A. Kumar, E. Maiorana, Y. Mulionoa, N. D. Roy, L. Wang, L. Yuan и др. Анализ этих работ позволил определиться с направлением исследования и выявить перспективные подходы к решению обозначенной научно-технической проблемы. Эти подходы связаны с разработкой концепции защищенного исполнения нейросетевых алгоритмов ИИ, моделей искусственных нейронов и НПБК на их основе, изначально устойчивых к деструктивным воздействиям и атакам, адаптивных моделей ИИ, способных подстраиваться под изменяющиеся данные, снижая влияние концептуального дрейфа в задачах высоконадежной биометрической аутентификации, а также алгоритмов их обучения. Из проведенного анализа следует, что на основе предложенных концепций, моделей и алгоритмов необходимо разработать методы, технологию и программный комплекс для создания систем высоконадежной многофакторной биометрической аутентификации с обеспечением защиты биометрических данных от компрометации.

Объект исследования: системы биометрической аутентификации на основе методов, моделей и алгоритмов доверенного ИИ. Предмет исследования: нейросетевые модели и алгоритмы

машинного обучения на малых выборках для высоконадежной биометрической аутентификации и защиты биометрических данных от компрометации. Цель работы: повысить надежность многофакторной биометрической аутентификации на основе защищенного исполнения нейросетевых моделей доверенного ИИ и алгоритмов их автоматического синтеза и обучения на малых выборках биометрических данных. Для достижения цели были выполнены следующие задачи:

1. Разработка концепции защищенного исполнения нейросетевых алгоритмов ИИ.
2. Разработка моделей искусственных нейронов и нейросетевого преобразователя биометрия – код, потенциально устойчивых к деструктивным воздействиям, и алгоритмов их робастного автоматического обучения на малых выборках.
3. Разработка адаптивной модели ИИ и алгоритмов ее обучения, позволяющих предупредить или снизить влияние концептуального дрейфа данных в системах биометрической аутентификации.
4. Разработка методов многофакторной аутентификации на базе тайных биометрических образов с обеспечением конфиденциальности биометрических данных.
5. Разработка технологии автоматического синтеза и обучения нейросетевых моделей для высоконадежной многофакторной биометрической аутентификации.

Для решения указанных задач применялись методы распознавания образов, машинного обучения, кодирования информации и защиты данных от компрометации, аппарат искусственных нейронных сетей (ИНС), ансамблевые методы, биоинспирированные алгоритмы и модели классификации образов, методы теории вероятностей и математической статистики, спектрального и корреляционного анализа, обеспечения дифференциальной конфиденциальности данных и знаний, идентификации и аутентификации.

Определение ключевых понятий

Под *надежностью* понимается способность биометрической системы сохранять во времени требуемый уровень точности аутентификации и защищенности от компьютерных атак в изменяющихся условиях функционирования. Согласно ГОСТ Р 52633.0, система биометрической аутентификации является высоконадежной, если показатель вероятности ошибки «ложного допуска» составляет менее 10^{-12}. Приведены угрозы и атаки («на решающий бит», «извлечение знаний», «ключ под ковриком», «состязательные», «представления») (рис. 1), перед которыми уязвимы биометрические системы и другие приложения ИИ. Указаны недостатки методов защиты знаний ИИ и биометрических шаблонов на основе гомоморфного шифрования, а также ограничения концепции федеративного обучения. Проведен анализ существующих стандартов и научных публикаций в области защиты ИИ и биометрических систем от обозначенных угроз.

В рамках деятельности Международных технических комитетов (ТК) по стандартизации ISO/IEC JTC 1/SC 42 «Artificial intelligence», ISO/IEC JTC 1/SC 37 Biometrics и национальных ТК 164 «Искусственный интеллект» и ТК 098 Биометрия и биомониторинг пока не разработаны стандарты в области защиты решающих правил от обозначенных атак. Наиболее детально эти вопросы проработаны для приложений биометрии. Национальные стандарты по нейросетевой биометрии серии ГОСТ Р 52633, закрепленные за ТК 362 «Защита информации», основаны на концепции НПБК.

Концепцию защищенного исполнения нейросетевых алгоритмов ИИ в задачах биометрической аутентификации позволяют реализовать: нечёткие экстракторы; гибриды нечеткого экстрактора и многослойных ИНС; автоматически обучаемые НПБК на базе ИНС; НПБК на базе квадратичных нейронов. Выявлены принципиальные недостатки методов защиты знаний путем гомоморфного шифрования. Анализ работ показал, что для задач классификации в защищенном режиме архитектуру ИИ можно разделить на два блока: блок извлечения признаков и блок классификации образов.

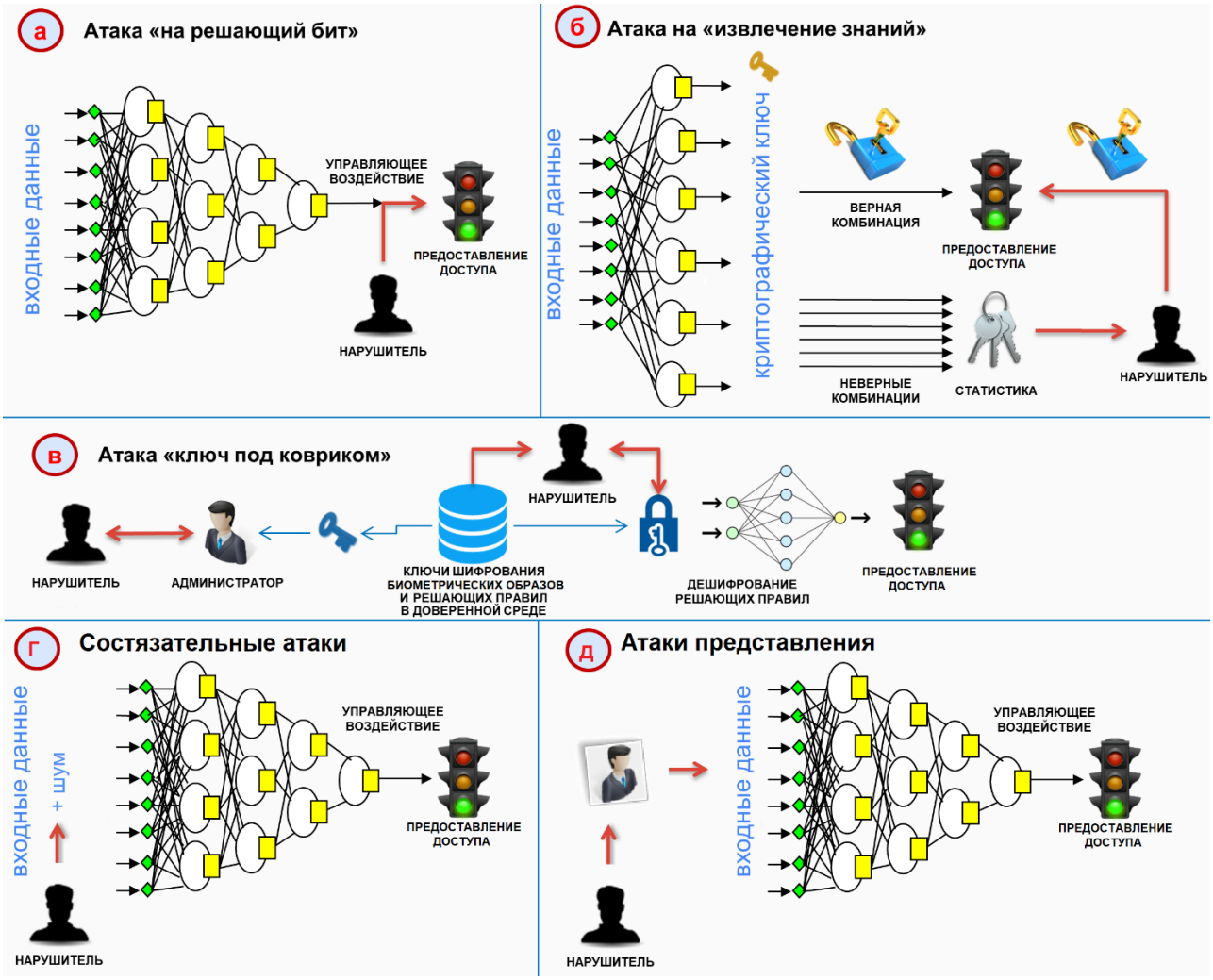


Рис. 1 Атаки на высоконадежные биометрические системы.

Длина ключа для нечеткого экстрактора и НПБК ограничена из-за ряда принципиальных недостатков и подверженности атакам соответственно [2]. Для сетей квадратичных нейронов длина ключа может быть примерно в 4 раза выше, чем для классических НПБК [3]. Ограничения длины ключа не позволяют использовать такие нейроны для широкого спектра классификационных задач. Применение методов глубокого обучения и аппарата многослойных нейронных сетей, по всей видимости, ограничивается блоком извлечения признаков. Блок классификации должен иметь простую архитектуру, такую, чтобы его синтез и обучение могли выполняться в автоматическом режиме.

В биометрических системах крайне важно обеспечить актуальность знаний ИИ. С течением времени возрастает количество сбоев и ошибок из-за изменчивости биометрических данных, например, в зависимости от психофизиологического состояния (ПФС) [4]. Данная проблема является частным случаем дрейфа концепций и данных.

Выводы: для решения обозначенных проблем требуется разработать иные модели искусственных нейронов и сетей, позволяющих повысить длину ключа в соответствии с текущими требованиями (например, ГОСТ Р 34.10-2012) или выше (на перспективу), а также позволяющих предотвратить или снизить влияние концептуального дрейфа. Безопасность блока классификации может быть дополнительно усилена криптографическими методами.

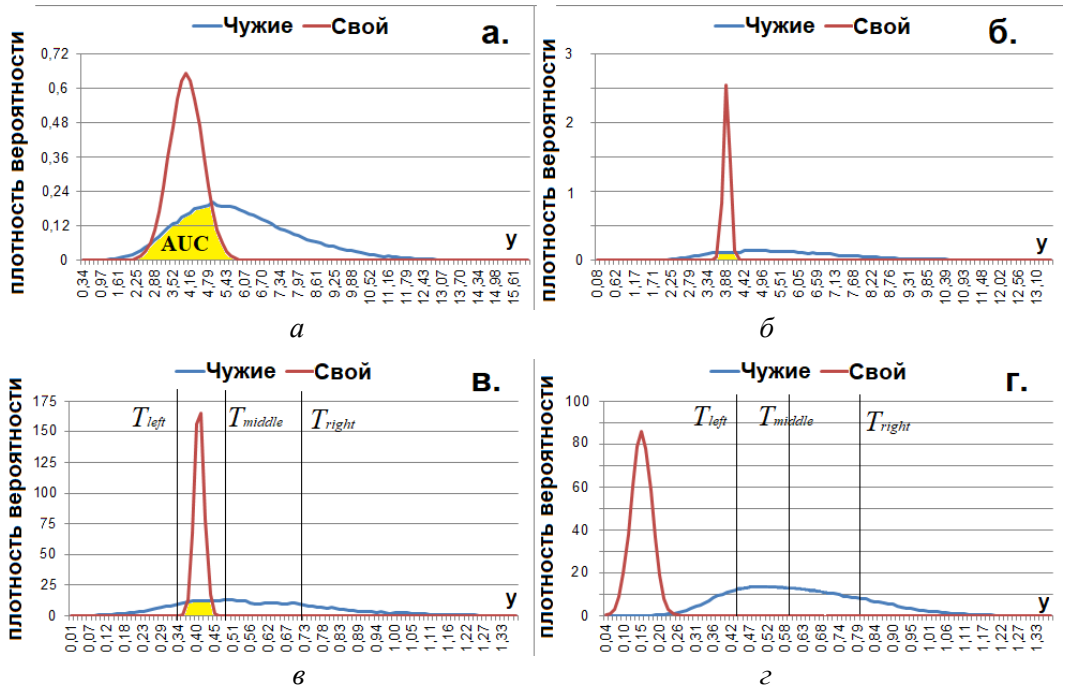
Концепция защищенного исполнения нейросетевых алгоритмов

Разработаны концепция защищенного исполнения нейросетевых алгоритмов ИИ и модель корреляционных нейронов для высоконадежной биометрической аутентификации – это новый класс нейронов, анализирующих корреляционные связи между признаками вместо признаков. Анализ внутренних корреляционных связей образов и их классификация происходят без необходимости хранения информации о корреляционных связях или признаках, характерных для биометрических образов пользователей (биометрические шаблоны не компрометируются при хранении). Корреляционные нейроны используются для формирования высоконадежного блока классификации, устойчивого к деструктивным воздействиям. Предложены модель НПБК на основе корреляционных нейронов и алгоритм ее автоматического обучения на малых выборках биометрических данных.

Показано, что пространство признаков искривляется из-за корреляционных связей между измерениями. Классы образов имеют уникальные матрицы коэффициентов корреляции $C_{j,t}$ между признаками. Поэтому относительно различных классов пространство признаков искривлено по-разному. Корреляция переносит часть информации об образах, касающейся уровней искривления, в «скрытые» измерения. Чтобы извлечь эту информацию, предложено множество метрик Байеса–Минковского [5, 6], в частности (1).

$$y = \sum_{j=1}^{n} \left| \left| \frac{a_t}{\delta_t} \right|^p - \left| \frac{a_j}{\delta_j} \right|^p \right|, j \neq t, \qquad y = \sqrt[p]{\sum_{j=1}^{n} \left| \left| \frac{a_t}{\delta_t} \right|^p - \left| \frac{a_j}{\delta_j} \right|^p \right|}, j \neq t, \tag{1}$$

где y – оценка близости образа к классу «Свой»; a_j – значение j-го признака из вектора $\bar{a}$, представляющего распознаваемый образ; n – количество признаков; δ_j – нормирующие коэффициенты, вычисляемые как среднеквадратичное отклонение признака для класса «Чужие», представляющего множество обезличенных образов, поэтому δ_j не компрометируют данные класса «Свой» (*обеспечивается дифференциальная конфиденциальность*).



а *б* *в* *г*

Рис. 2 Графики плотностей вероятности:
значений меры (1) при $p = 1$: $|C_{j,t}| < 0.3$, $n' = 5$ (*а*); $1 > C_{j,t} > 0.95$, $n' = 5$ (*б*);
значений меры (3): $1 > C_{j,t} > 0.95$, $n' = 10$ (*в*); $-1 < C_{j,t} < -0.95$, $n' = 10$ (*г*).

На рис. 2 *а*, *б* видно, что $AUC_{|C| > 0.95}(\Phi_G(y), \Phi_I(y)) < AUC_{|C| < 0.3}(\Phi_G(y), \Phi_I(y))$, где *AUC* – площадь, ограниченная функциями плотности вероятности (ФПВ) «Свой» $\Phi_G(a_j)$, «Чужие» $\Phi_I(a_j)$, и осью абсцисс. Чем выше корреляция между признаками, тем меньше неверных решений дает мера (1). Размерность пространства Байеса–Минковского составляет

$$n' = 0.5(n(n–1)) = 0.5n^2 – 0.5n.$$

Под мета-признаками подразумеваются разности вида (2):

$$a'_{j^*} = a'_{t,j} = f(a_t, a_j) = \left| \left|\frac{a_t}{\delta_t}\right|^p - \left|\frac{a_j}{\delta_j}\right|^p \right|, j > t, j^* = \sum_{j^*=1}^{t-1}(n - t^*) + j - t, \quad (2)$$

которые фактически являются грубыми оценками корреляционной зависимости между двумя исходными признаками под номерами *j* и *t*, сделанными всего по одному примеру, но при наличии некоторых априорных знаний, полученных в процессе обучения на выборке небольшого объема.

Экспериментально установлено (на больших объемах сгенерированных данных), что один метапризнак Байеса–Минковского может содержать в 2–3 раза больше информации, чем содержится суммарно в паре исходных признаков, от которых он порожден, если они сильно коррелированы.

Корреляционный нейрон предлагается строить на метрике взвешенного среднеквадратичного отклонения (3), которая позволяет отделить как положительно, так и отрицательно коррелированные данные:

$$y = \sqrt{\frac{1}{\eta}\sum_{j^*=1}^{n'} w_{j^*}\left(a'_{j^*} - m'\right)^2} = \sqrt{\frac{1}{\eta}\sum_{\iota=1}^{\eta} w_\iota\left(a'_\iota - m'\right)^2}, m' = \frac{1}{\eta}\sum_{\iota=1}^{\eta} a'_\iota \quad (3)$$

где η – количество входов нейрона; w_{j^*} – вес синапса под номером j^* ($w_{j^*} \geq 0$, если $w_{j^*} = 0$, то j^*-й метапризнак не влияет на сумму, то есть не соединяется с нейроном); ι – номер метапризнака без учета синапсов с нулевым весом (для сквозной нумерации). Вес синапса рассчитывается по формуле (4):

$$w_\iota = \frac{|m''_{(G),\iota} - m''_{(I),\iota}|}{\sigma''_{(G),\iota} \cdot \sigma''_{(I),\iota}}, \quad (4)$$

где $m_{(G),\iota}''$, $m_{(I),\iota}''$ – математические ожидания, а $\sigma_{(G),\iota}''$, $\sigma_{(I),\iota}''$ – среднеквадратичные отклонения значений ι-го метапризнака второго порядка ($a''_\iota = (a'_\iota - m')^2$) для образов «Свой» и «Чужие», соответственно, рассчитанные по данным обучающей выборки. После обучения нейрона параметры $m_{(G),\iota}''$, $m_{(I),\iota}''$, $\sigma_{(G),\iota}''$, $\sigma_{(I),\iota}''$ должны быть удалены. В качестве функции активации предлагается использовать многоуровневую пороговую функцию квантования (5):

$$\phi(y) = \begin{cases} 3, y < T_{left} \\ 2, T_{left} \leq y < T_{middle} \\ 1, T_{middle} \leq y < T_{right} \\ 0, y \geq T_{rightt} \end{cases}, \quad (5)$$

где T_{left}, T_{middle} и T_{right} – левый, средний и правый пороговые значения активации нейрона (рис. 2 *в*, *г*). В соответствии с предлагаемой моделью нейрон имеет четыре варианта активации {0, 1, 2, 3} и только одно из них соответствует гипотезе «Свой», остальные – гипотезе «Чужие». О том, какое состояние активации соответствует гипотезе «Свой» (далее ϕ_G), известно только на этапе синтеза и обучения НПБК, злоумышленник не обладает этой информацией, так как она не сохраняется после настройки нейрона.

При $0.1 < P(\phi(y)) < 0.4$ обеспечивается достаточно высокая энтропия выходов нейронов в ответ на образы «Чужие», где $P(\phi(y))$ – относительная частота появления $\phi(y)$ при поступлении на вход образа «Чужой». При вычислении порогов сначала рассчитываются граничные

значения откликов нейрона y на обучающие примеры «Свой» ($y_{G\min}$, $y_{G\max}$) и «Чужие» ($y_{I\min}$, $y_{I\max}$), а также значения их функций распределения $F_G(y)$ и $F_I(y)$ исходя из гипотезы нормального распределения y (подтверждено методом хи-квадрат).

Предложен алгоритм настройки порогов (рис. 3).

Введем коэффициент AUC_{MAX}, равный максимально допустимому показателю AUC ($\Phi_G(y)$, $\Phi_I(y)$) для нейрона, чтобы исключить «слабые» нейроны, которые дают близкие отклики на образы «Свой» и «Чужие». К значению функции активации применяется одна из возможных таблиц перевода состояний {0, 1, 2, 3} в двухбитный код. При обучении нейрона хеш-таблица выбирается случайно, но с учетом того, на какие два ключевых бита (далее b) настраивается нейрон.

Разработана модель НПБК на базе корреляционных нейронов для реализации блока классификации, устойчивого к деструктивным воздействиям (рис. 4). Перед поступлением образов на входы НПБК они должны быть обработаны блоком извлечения признаков. Эксперименты с использованием синтетических данных показали, что при параметре $p = 0.9$ отображения (2) в большинстве случаев удается достичь наименьшего количества ошибок классификации. Коэффициенты δ_j вычисляются на основании тренировочной выборки «Чужие» (один набор коэффициентов может использоваться для нормирования признаков сразу для множества НПБК, принимающих на входы аналогичные признаки).

Так, для обучения корреляционного нейрона достаточно определить связанные мета-признаки, вычислить веса, пороги и задать хеш-таблицу.

Количество входов η для всех корреляционных нейронов НПБК должно быть равным. При синтезе НПБК для конкретного пользователя необходимо убедиться, что имеется достаточное количество пар признаков с уровнями взаимной корреляции $C_{j,t} < C_-$ и $C_{j,t} > C_+$ (по результатам экспериментов оптимальными являются $C_+ = 0.5$ и $C_- = -0.5$). Следует рассчитать корреляционную матрицу по данным выборки «Свой». Любая пара коррелированных признаков потенциально порождает один мета-признак Байеса–Минковского. Пусть N_- и N_+ – количества нейронов, ориентированных на обработку данных со взаимной корреляцией $C_{j,t} < C_-$ и $C_{j,t} > C_+$. Должно соблюдаться условие $N_- \approx N_+$ (допускается расхождение на 1 – 3 нейрона). Каждый нейрон должен обрабатывать уникальную комбинацию мета-признаков и генерировать на выходе 2 бита. Нужное количество нейронов определяется исходя из длины ключа L. Например, при $L = 1024$ бит $N_- = N_+ = L/2/2 = 256$. Тогда, если $\eta = 4$, то для синтеза сети потребуется 2048 пар признаков. Алгоритм синтеза и обучения НПБК можно изложить как последовательность шагов:

1. Расчет корреляционной матрицы признаков.
2. Подсчет пар отрицательно коррелированных признаков ($C_{j,t} < C_-$). Если количество пар менее ηN_-, то C_- увеличивается на 0.05, и шаг 2 повторяется.
3. Синтез и обучение N_- нейронов для анализа отрицательно коррелированных данных в соответствии с алгоритмом на рис. 3. Если количество нейронов, удовлетворяющих условиям алгоритма на рис. 3, оказалось менее N_-, то C_- увеличивается на 0.05, и шаги 2 – 3 повторяются.
4. Подсчет пар положительно коррелированных признаков ($C_{j,t} > C_+$). Если количество пар менее ηN_+, то C_+ уменьшается на 0.05, и шаг 4 повторяется.
5. Синтез и обучение N_+ нейронов для анализа положительно коррелированных данных в соответствии с алгоритмом на рис. 3. Если количество нейронов, удовлетворяющих условиям алгоритма на рис. 3, оказалось менее N_+, то C_+ уменьшается на 0.05, и шаги 4 – 5 повторяются.

По мере сужения интервала (C_-; C_+) уже созданные нейроны допустимо не удалять. Алгоритм исполняется, пока не выполнится условие $N_- = N_+ = L/4$ либо пока не будет нарушено условие $|C_{-+}| \geq 0.3$.

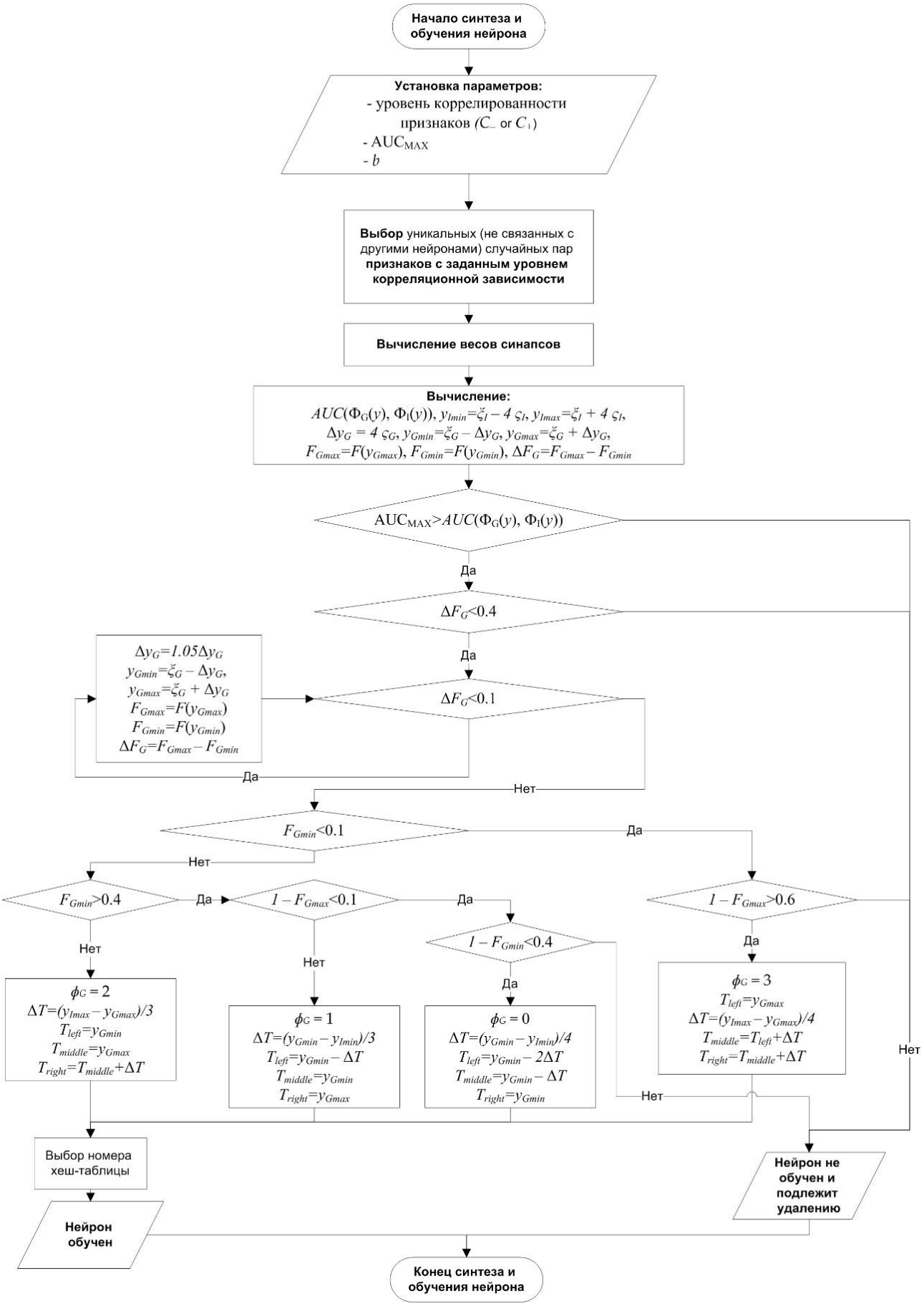


Рис. 3 Схема алгоритма синтеза и обучения корреляционного нейрона.

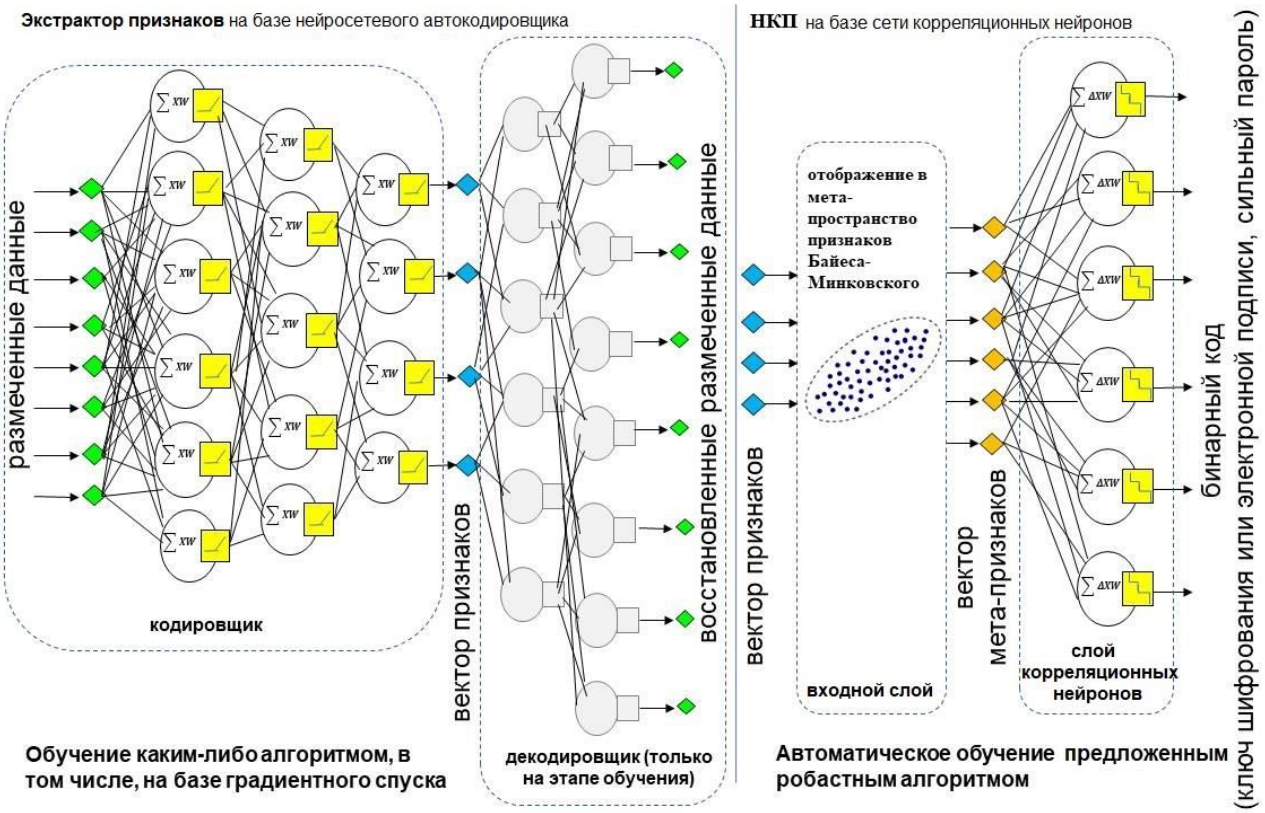


Рис. 4 Структурная схема связывания ключа и биометрического образа: слева – блок извлечения признаков, справа – блок классификации.

Таблицы весовых коэффициентов w_i и номера хеш-таблиц обученного НПБК представляют собой защищенный эталон пользователя.

Дрейф биометрических признаков и онлайн-обучение нейро-иммунных моделей

ПФС отражается на динамических биометрических признаках (клавиатурного [7] и рукописного почерка [8, 9], голоса [10]) спонтанными изменениями. Если ПФС пользователя не совпадает на этапах обучения системы и аутентификации, то вероятность ошибок «ложного отказа» (FRR) и «ложного допуска» (FAR) многократно повышается. Эту проблему можно решить с использованием алгоритмов онлайн-обучения моделей ИИ.

Проведен анализ подходов к построению адаптивных моделей ИИ (моделей, способных к онлайн-обучению), включая методы глубокого обучения с подкреплением, эволюционный и иммунный подходы. Методы глубокого обучения с подкреплением и эволюционные нейронные сети работоспособны только при больших объемах обучающей выборки. Искусственные иммунные системы и сети (ИИС) обладают двойной пластичностью, позволяющей легко изменять в процессе функционирования не только собственные параметры, но и структуру (в отличие от нейронных сетей). Проанализированы существующие подходы к построению моделей ИИС (на базе дендритных клеток, негативного отбора, клональной селекции и сетевых алгоритмов) [11], а также методы и алгоритмы биометрической аутентификации на их основе [7, 8, 10].

Модели, в которых совместно применяются элементы аппаратов ИИС и ИНС, принято называть нейро-иммунными (частный случай нейросетевых моделей). Синтез и обучение нейро-иммунных моделей выполняются с использованием принципов ИИС, однако процесс классификации образов при помощи нейро-иммунной модели после ее обучения схож с работой ИНС.

Детектор (аналог искусственного нейрона) – искусственная иммунная клетка, которая обладает способностью обнаруживать чужеродные антигены, анализируя распознаваемый образ и реагируя на него пропорционально тому, насколько этот образ соответствует антигену (шкала реакций [0;1]) [12]. Каждый детектор следует рассматривать как бинарный классификатор, являющийся «обверткой» над корреляционным (классическим, квадратичным) нейроном, состоящий из нескольких функций, последовательно применяющихся к $\bar{a}$ (6):

$$u_i = \phi_\chi\left(y' = \varphi\left(y = f_x(\bar{\alpha} = R(\bar{a}, \Psi_i), \breve{g}, \Theta_i), T_i\right)\right), \quad (6)$$

1) $\bar{\alpha} = R(\bar{a}, \Psi_i)$ – функция-рецептор извлекает η из n признаков; Ψ_i – множество номеров признаков, которые должен анализировать i-й детектор;

2) $y = f_x(\bar{\alpha}, \breve{g}, \Theta_i)$ – функция-ядро вычисляет близость образа к эталону класса «Свой»; x – тип ядра (например, на базе корреляционного нейрона (3)); $\breve{g}$ – вектор параметров функционала, которые влияют на характер вычислений (например, степенной коэффициент p для перехода в пространство мета-признаков Байеса–Минковского); $\Theta_i = \{w_1, w_2, \ldots, w_\eta, \delta_1, \delta_2, \ldots, \delta_\eta\}$ – множество параметров обученного нейрона. Элементы из множеств $\breve{g}$ и Θ_i зависят от типа нейрона или меры близости, на базе которых строится детектор. Помимо корреляционных нейронов в основе детектора может лежать квадратичный, классический нейрон, байесовский классификатор и др. Однако большинство существующих метрик компрометируют знания ИИ, в отличие от корреляционного нейрона. Разные ядра образуют различные виды детекторов, которые дают слабо коррелированные решения относительно друг друга, что позволяет объединять такие нейроны (детекторы) в сеть для получения синергетического эффекта. Из любого ядра можно получить разные меры близости за счет изменения параметров $\breve{g}$;

3) $y' = \varphi(y, T_i)$ – функция нормирования откликов y относительно порога T_i, который вычисляется в процессе настройки i-го детектора. Для корреляционных нейронов функция имеет вид: $\varphi(y, T_i) = y / T_i$, где T_i – это максимальное значение функции-ядра i-го детектора, при поступлении на его вход обучающих образов «Свой»;

4) $u_i = \phi_\chi(y'_i)$ – функция активации, дополнительный нелинейный элемент детектора, который определяет особенности реагирования на антиген. Функция активации также необходима, чтобы привести отклик детектора к области значений [0;1]. В отличие от функции активации в защищенном режиме (5) адаптивная модель ИИ использует сигмоиды.

Одной из теоретических проблем аппарата ИИС является слабая обоснованность используемых мер близости. Согласно теореме «об отсутствии бесплатных завтраков» (No Free Lunch), ни одна мера близости не может быть оптимальной для всего множества задач распознавания образов. Поэтому в настоящей работе каждый детектор определяет близость уникальным способом, а состав детекторов «подстраивается» под задачу в процессе обучения.

Разработанная адаптивная модель ИИ (рис. 5) так же, как и модель защищенного исполнения, настраивается на верификацию образа конкретного пользователя и обучается на малых выборках образов «Свой» и «Чужие» (можно использовать одну выборку «Чужих» для обучения всех моделей).

Детекторы разделены на две группы: *врожденный (ВИ) и приобретенный иммунитет (ПИ)*, и рассматриваются как *два комитета (ансамбля)* слабых классификаторов, обучаемых при помощи разных алгоритмов. Коллективное решение N детекторов вычисляется как среднее частных решений:

$$\ddot{u} = \Phi(\overline{D}^* = \{D_1^*, \ldots D_N^*\}, \bar{a}) = \frac{1}{N} \cdot \sum_{i=1}^{N} \phi(D_i^*, \bar{a}) = \frac{1}{N} \cdot \sum_{i=1}^{N} u_i \quad . \quad (7)$$

ВИ передается посредством генов. Костный мозг (см. рис. 5) является местом пребывания детекторов, параметры и состав которых определяются в процессе итерационного обучения ИИС с учителем с использованием *тренировочной* и *валидационной* выборок, которые являются *непересекающимися* подмножествами *обучающей* выборки. Для этого предложен специальный алгоритм (рис. 6, *а*), отличающийся от алгоритма на рис. 3.

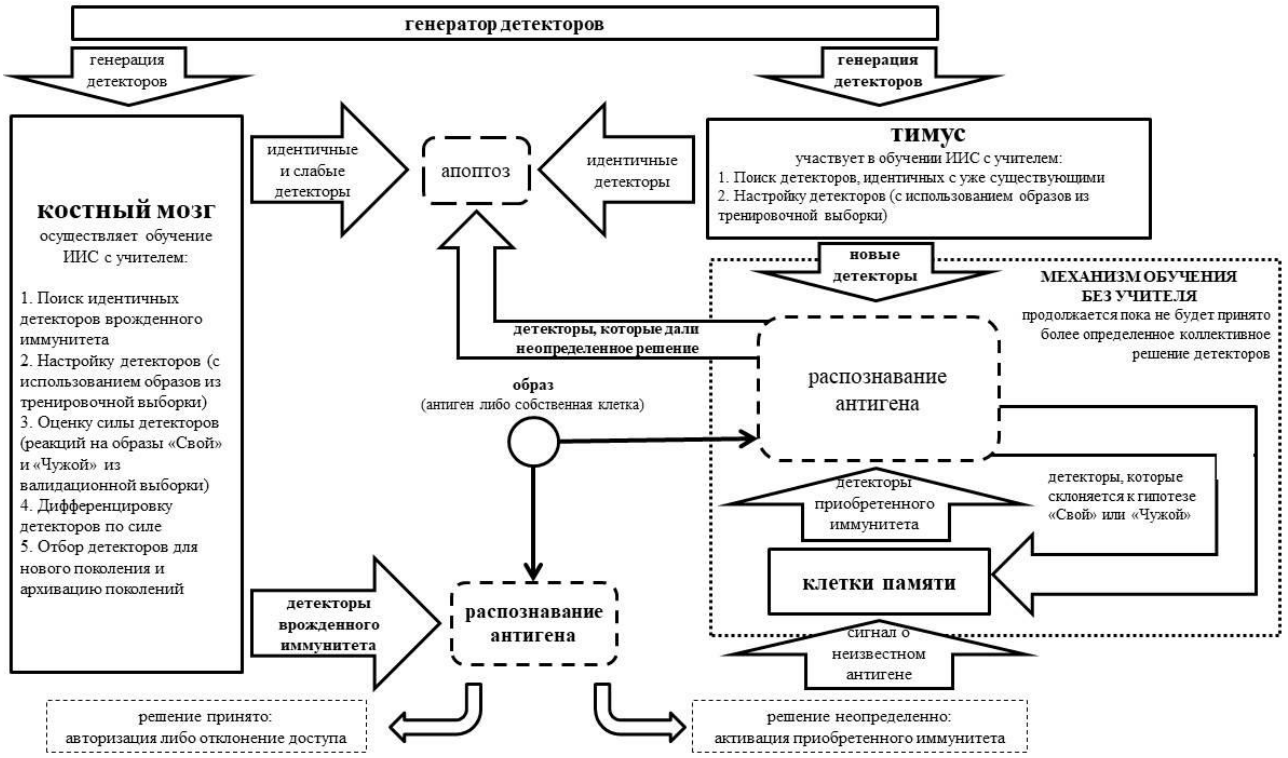


Рис. 5 Функциональная схема ИИС.

ПИ развивается с течением жизни и определяет способность организма обезвреживать специфические антигены, которые попадали в организм ранее. Адаптивный иммунный ответ приводит к появлению клеток памяти (также представленных детекторами), которые долгое время пребывают в «спящем состоянии» до повторной встречи с антигеном. В разработанной модели ПИ формируется в процессе функционирования ИИС. Если решение об отнесении образа к классам «Свой» или «Чужой» является неоднозначным, могут генерироваться новые иммунокомпетентные детекторы.

Идея объединения классификаторов в комитет основана на теореме Кондорсе, которая утверждает: если мнения экспертов независимы, и вероятность правильного решения каждого из них больше 0.5, то с увеличением количества экспертов вероятность правильного коллективного решения возрастает и стремится к единице. Однако на практике решения классификаторов, играющих роль экспертов, в той или иной мере коррелированы; чем ниже коррелированность решающих правил, тем более ощутим положительный эффект при их комплексировании. Таким образом, имеются следующие гиперпараметры, которые влияют на эффективность комитета детекторов:

RD – матрица коэффициентов корреляции $r(\bar{u}_i, \bar{u}_l)$ между реакциями всех возможных пар детекторов, где $\bar{u}_i$ – вектор реакций i-го детектора на примеры образов «Чужих» из тренировочной или валидационной выборки; N – количество детекторов; Δu – сила детекторов, их способность давать как можно более высокие показатели разницы средних уровней реакции на образы «Свой» μ_u (С) и «Чужой» μ_u (Ч) (8):

$$\Delta u = \mu_u \text{ (Ч)} - \mu_u \text{ (С)}, \ \mu_u \text{ (Ч)} > \mu_u \text{ (С)} . \qquad (8)$$

Апробирована стратегия снижения уровня коррелированности решений детекторов, которая оказалась недостаточно продуктивной. Не наблюдалось сходимости алгоритма: процесс обучения был длительным, и не всегда удавалось найти N детекторов с заданным минимальным уровнем взаимной коррелированности решений. По этой причине в настоящей работе выбрана стратегия повышения силы детекторов при условии, что они не должны быть идентичными. При появлении в ИИС идентичных или слабых детекторов происходит их уничтожение (апоптоз, см. рис. 5).

Детектор можно описать множеством параметров $D_i = \{\Psi_i, \check{g}, x, \chi\}$. Сгенерировать детектор означает сгенерировать D_i. Необходимо, чтобы решения всех детекторов ВИ и ПИ не являлись полностью коррелированными. Поэтому после генерации детектора осуществляется проверка идентичности его параметров и параметров уже существующих детекторов. При обнаружении «двойника» его следует удалить и сгенерировать детектор снова. При этом значения параметров $\check{g}$ можно считать равными, когда они отличаются менее чем на 10^{-1}. Чем сильней различия между D_i и D_l, тем менее коррелированы решения i-го и l-го детекторов.

В разработанной ИИС реализуется идея случайных подпространств признаков, но в отличие от алгоритма «случайный лес» Ψ_i задается с учетом корреляции между признаками. Этот прием называется симметризацией корреляционных связей. Другая идея заключается в объединении разнородных случайных классификаторов. Примером подобной техники является нейросетевое обобщение множества различных критериев. Настройка детектора связана с вычислением порога T_i и эталонных описаний признаков Θ_i (w_j, δ_j). Настроенный детектор обозначим $D^*_i = \{\Psi_i, \check{g}, x, \chi, \Theta_i, T_i\}$.

При разработке алгоритмов обучения учтены следующие методы:

– бэггинг (bootstrap aggregating) – обучение базовых классификаторов на разных подмножествах обучающей выборки. Бэггинг уменьшает дисперсию голосов базовых классификаторов и помогает избежать переобучения;

– бустинг (boosting) – семейство алгоритмов машинного обучения, преобразующих слабые модели к сильным. Бустинг строит ансамбль путём тренировки каждого нового классификатора, уделяя больше внимания обучению на примерах, которые предыдущие модели классифицировали ошибочно.

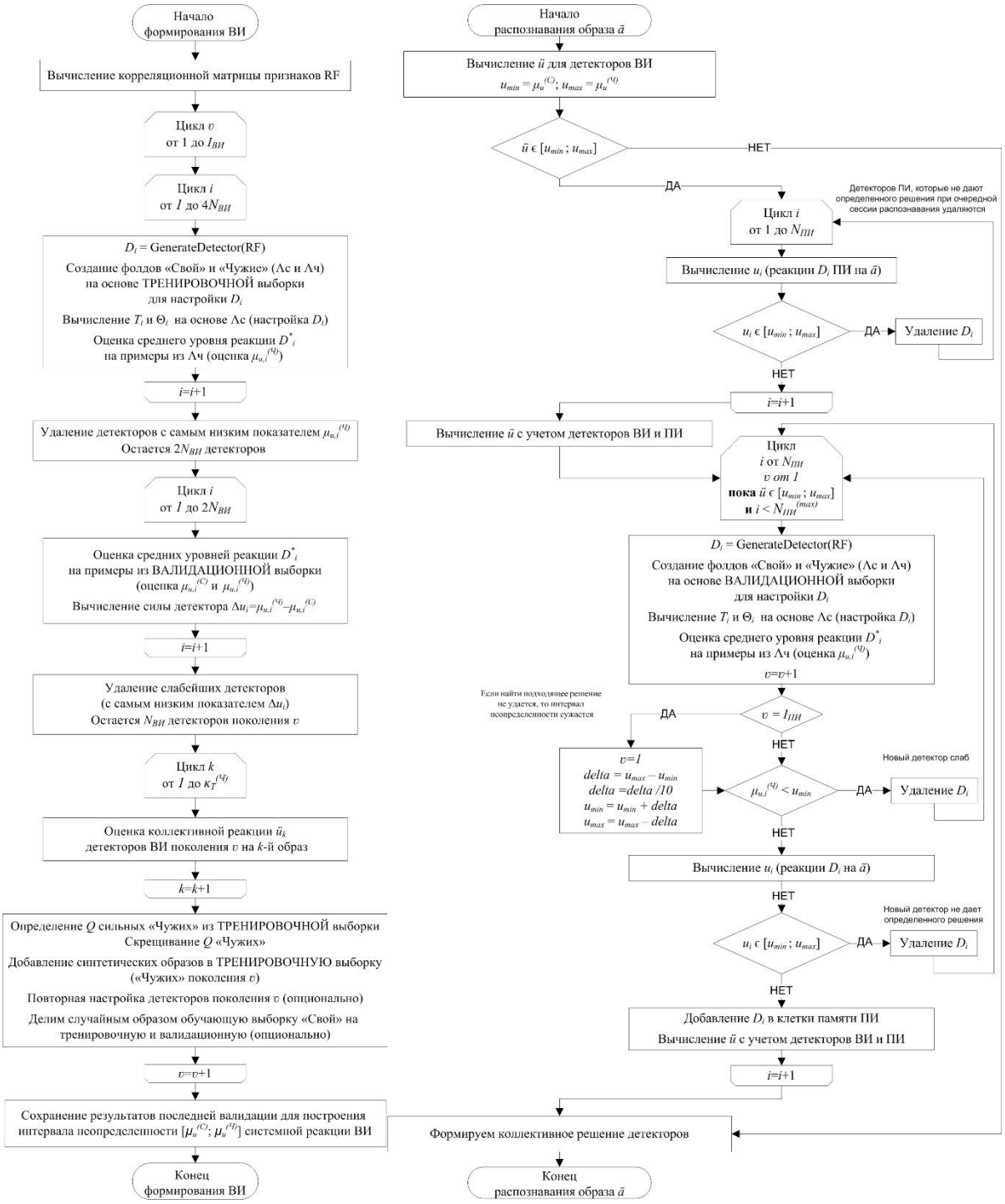
В разработанном алгоритме обучения с учителем на каждой итерации происходит генерация новой популяции детекторов, которые настраиваются с учетом нескольких случайных тренировочных примеров (бэггинг), и выполняется промежуточная оценка их эффективности (рис. 6, *а*). Слабые детекторы уничтожаются, и появляется новое поколение более эффективных детекторов. Мерой эффективности детекторов можно считать Δu (8). По результатам последней валидации вычисляются оценки $\mu_{u(С)}$ и $\mu_{u(Ч)}$ для коллективного решения детекторов ВИ. Эти параметры используются для построения интервала неопределенности решения (ИНР) $[\mu_{u(С)}; \mu_{u(Ч)}]$. ИНР является частью механизма подкрепления при онлайн-обучении модели в процессе ее функционирования. Этот механизм активируется при формировании ПИ.

На каждой итерации обучения с учителем синтезируются новые образы «Чужих» (рис. 6, *а*) путем скрещивания тренировочных примеров, которые хуже всего классифицируются детекторами ВИ. Скрещивание образов $\bar{a}_k$ и $\bar{a}_m$ происходит по формуле (9) (в соответствии с ГОСТ Р 52633.2-2010):

$$a_{c,j} = \frac{C_{syn}+1-c}{C_{syn}+1} \cdot a_{k,j} + \frac{c}{C_{syn}+1} \cdot a_{m,j}\ , \qquad (9)$$

где C_{syn} – количество синтетических примеров, порождаемых парой «сильных Чужих» предыдущего поколения (в настоящей работе $C_{syn} = 1$); c – номер синтетического примера; j – номер признака. Синтетические образы добавляются в тренировочную выборку, что позволяет следующему поколению детекторов эффективнее обучаться классифицировать образы «Чужих», наиболее близких к образам «Свой», с большей эффективностью (один из вариантов бустинга). Таким образом, ИИС одновременно «учится» создавать образы более сильных «Чужих» и распознавать их.

На скорость и эффективность алгоритма обучения с учителем влияют следующие основные параметры: $I_{ВИ}$ – количество итераций обучения; $N_{ВИ}$ – количество детекторов ВИ; Q – количество сильных «Чужих» (на каждой итерации генерируется $C_{syn} \cdot Q \cdot (Q - 1)/2$ примеров). Тренировочная выборка «Чужие» увеличивается с каждой итерацией обучения при добавлении синтетических примеров, валидационная выборка остается неизменной.



а *б*

Рис. 6 Алгоритмы обучения:
а – с учителем (или формирование ВИ);
б – онлайн-обучение с подкреплением (или формирование ПИ).

Разработан алгоритм онлайн-обучения с подкреплением (рис. 6, *б*) для дообучения модели в процессе исполнения [12]. Введем следующее правило, основанное на ИНР: если $u_i > \mu_u$ (Ч) или $ui < \mu_{u,i}$ (С), то решение детектора D^*_i считается определенным, а если μ_u (С) $< u_i < \mu_u$ (Ч), то его решение не определено. Если при распознавании образа коллективное (7) решение детекторов ВИ считается неопределенным, то активируется механизм ПИ (рис. 6, *б*). Тогда генерируются новые детекторы, которые настраиваются на других данных – примерах из валидационной выборки. Для новых детекторов вычисляются реакции u_i, но при формировании

коллективного решения учитываются голоса только тех детекторов, которые дают определенный ответ (эти детекторы становятся клетками памяти), детекторы ПИ с неопределенным ответом уничтожаются.

На скорость и эффективность алгоритма онлайн-обучения влияют следующие параметры: $I_{ПИ}$ – количество итераций обучения; $N_{ПИ(max)}$ – максимальное количество детекторов ПИ ($N_{ПИ}$ – их фактическое количество). Введение $I_{ПИ}$ позволяет избежать бесконечного цикла дообучения. Механизм ПИ компенсирует недостаток априорных знаний о классах «Свой» и «Чужой» и снижает вероятность возникновения концептуального дрейфа.

Проведен эксперимент с двумя общедоступными и одной собственной базами клавиатурного почерка. Опыты проводились при различном объеме обучающей выборки «Свой»: от $K_G = 20$ до $K_G = 40$. Тренировочная $K_{G(T)}$ и валидационная $K_{G(V)}$ выборки «Свой» делились в соотношении: $K_{G(T)} = 2 \cdot K_{G(V)}$. Тренировочная $K_{I(T)}$ и валидационная $K_{I(V)}$ выборки «Чужих» включали по одному примеру от каждого испытуемого. Остальные примеры использовались в качестве тестовой выборки. Тестирование проводилось методом перекрестного сравнения. Если обучать и тестировать систему на образах испытуемого, которые были записаны в разные дни, то репрезентативность выборки снижается, и наблюдается большая разница (более 15%) между оценками коэффициента равной вероятности ошибок (EER) до и после онлайн-обучения.

Процесс обучения адаптивной модели ИИ с учителем оказался достаточно устойчивым (наблюдалась незначительная склонность к переобучению при высоких значениях IBИ). В зависимости от используемого набора данных лучшие показатели ошибок составили: EER = 0.079, EER = 0.053, EER = 0.026.

Разработанная адаптивная модель и алгоритмы ее обучения удовлетворяют основным принципам построения ИИС:

1. Распределенный характер вычислений и проявление эмерджентности.
2. Достаточно устойчивый процесс обучения.
3. Способность к адаптации.
4. Взаимодействие – ВИ формирует параметры, которые влияют на механизм подкрепления детекторов ПИ.
5. Надежность решений зависит от объема и чувствительности популяции детекторов.
6. Формирование памяти при помощи механизмов ВИ и ПИ.

ВЫСОКОНАДЕЖНАЯ МНОГОФАКТОРНАЯ БИОМЕТРИЧЕСКАЯ АУТЕНТИФИКАЦИЯ

На данный момент не утверждено единого стандарта для безопасного объединения нескольких разнородных биометрических образов при защищенном исполнении процедур аутентификации (проект ГОСТ Р 52633.7 «Защита информации. Техника защиты информации. Высоконадежная мультибиометрическая аутентификация» находится на публичном обсуждении). Существующий стандарт ГОСТ Р 54411-2011 «Мультимодальные и другие мультибиометрические технологии» не рассматривает вопросы защищенного исполнения процедур биометрической аутентификации, а только варианты объединения классификационных решений и биометрических образов при их последовательном или одновременном представлении. В настоящей работе предлагается иной вариант объединения биометрических образов.

Представлены комплексный метод и алгоритм трехфакторной высоконадежной аутентификации с последовательным предоставлением образов (рис. 7). Первый фактор аутентификации – это тайный акустический образ уха [13]. Под тайным образом понимается, что он не компрометируется в естественной среде (фотография уха не информативна для создания состязательных примеров). Для защиты данных уха от компрометации при хранении и передаче по каналам связи используется разработанная модель НПБК. Второй и третий факторы могут быть открытыми (рукописная подпись, фиксированная фраза) или тайными (рукописный и голосовой пароли) [14]. Для этих типов биометрических данных важно поддерживать акту-

альность, так как они изменчивы, поэтому для их анализа применяется адаптивная нейро-иммунная модель ИИ. Для защиты знаний адаптивной модели ИИ ее параметры после обучения шифруются на ключе, формируемом НПБК. Небольшое число ошибочных бит в ключе, генерируемом НПБК, может быть скорректировано за счет использования алгоритмов помехоустойчивого кодирования и кодов, исправляющих ошибки. Это позволяет балансировать показатели FRR и FAR на выходе НПБК. Если число ошибочных бит в ключе больше исправляющей способности кода (когда на вход НПБК поступает образ «Чужой»), то параметры адаптивной модели ИИ дешифруются неверно, и доступ отклоняется; иначе исполняется алгоритм классификации и онлайн-обучения, который был проиллюстрирован на рис. 6, *б*.

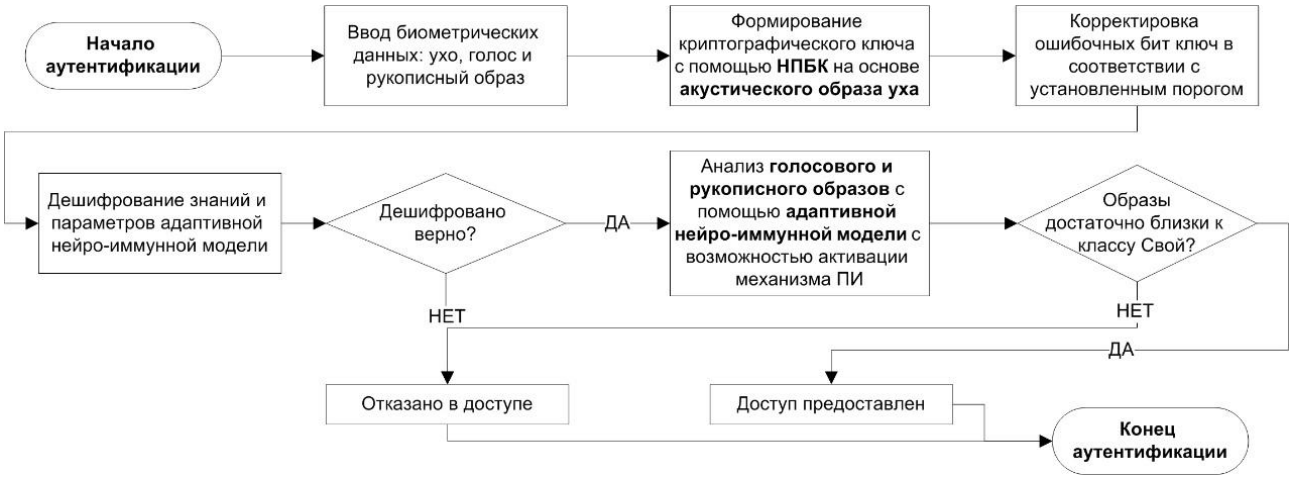


Рис. 7 Алгоритм трехфакторной аутентификации в защищенном режиме.

Рассмотрим каждый фактор (тип образов) по отдельности.

Строение уха полностью закладывается до 8 лет. Длина, толщина и форма ушного канала различаются у людей. Ушной канал создает резонанс (в среднем 2.5 кГц), чтобы получить информацию о его строении, можно воздействовать на него акустическими волнами, которые отражаются от стенок. Эхо-сигнал имеет отличия, обусловленные индивидуальными особенностями канала. Параметры эхо-сигнала или его передаточной функции можно воспринимать как вектор биометрических признаков.

Создано устройство для регистрации характеристик уха, которое состоит из двух электретных микрофонов, звукоизолирующего корпуса наушников, двух динамиков и звуковой карты CREATIVE. Для сбора биометрических образов привлечено 75 человек (мужчин и женщин в равном соотношении в возрасте от 18 до 40 лет без отологических патологий). Испытуемым предложено прослушать звуковой моно-сигнал возрастающей и убывающей частоты, получаемый путем линейной частотной модуляции. Частота сигнала варьировалась от 1 до 14 кГц, длительность составляла 10 сек, громкость – 80 дБ. Эхо-сигнал одновременно регистрировался монтированными в корпус наушников микрофонами. Все испытуемые прослушали сигнал через два динамика по 15 раз, каждый раз снимая и надевая наушники (чтобы учесть зависимость эхо-сигнала от монтажа). Регистрируемый эхо-сигнал можно назвать эхограммой или акустическим образом уха. Сформирован набор данных в виде совокупности wav-файлов.

Для анализа эхо-сигналов применялось быстрое оконное преобразование Фурье (STFT). Спектрограммы сигналов были преобразованы в усредненный по всем окнам амплитудный спектр $\bar{A}'$ (размер окна $W_{\text{size}} = 65536$, шаг $W_{\text{step}} = 16384$), из которого удалялись отчеты, соответствующие частотам менее 1 кГц и более 20 кГц (рис. 8). Применялись следующие оконные функции: Хэмминга, Блэкмана, Барлетта, прямоугольное, Гаусса, Лапласса.

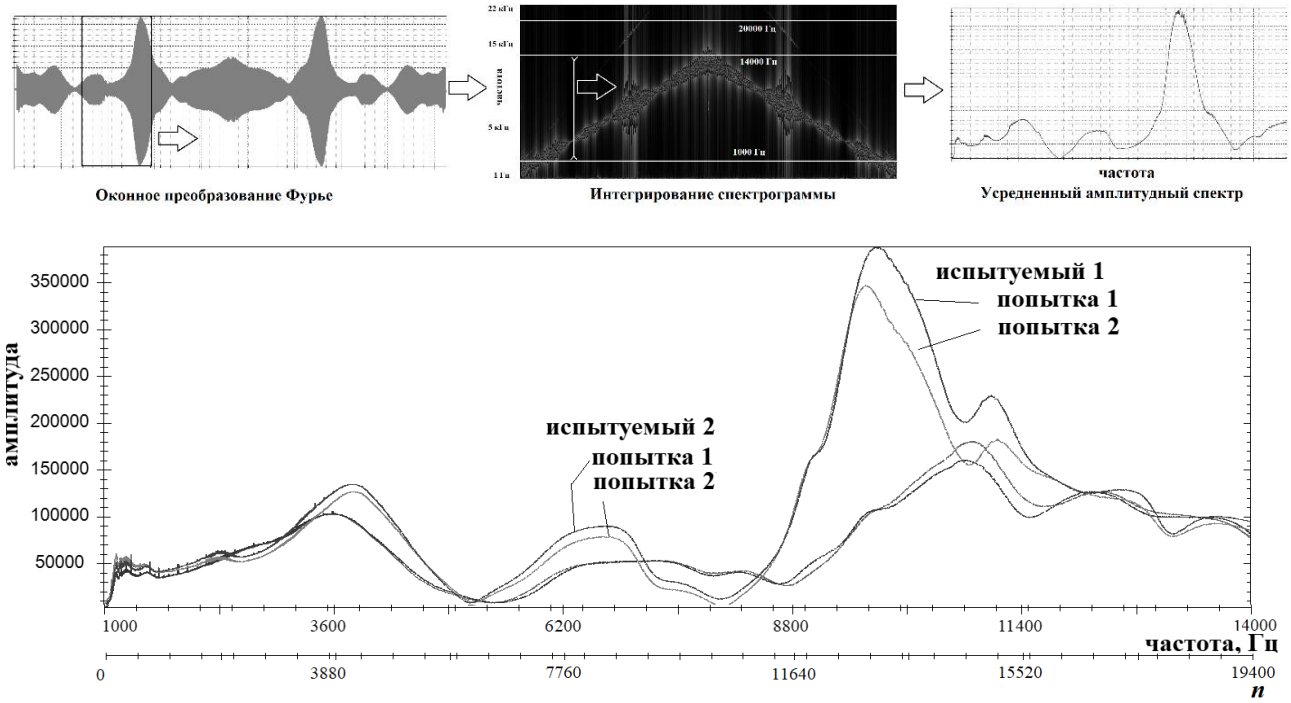


Рис. 8 Получение усредненного амплитудного спектра (вверху) и различия спектров у испытуемых (внизу).

Чтобы выявить локальные особенности усредненного спектра, построены кепстрограммы $K_{s,k}$ путем применения STFT к $\bar{A}'_{s,k}$ без операции логарифмирования. Так, частотная шкала v спектральной функции $A'_{s,k}$ (ν) принималась за временную шкалу, $\bar{A}'_{s,k}$ делился на частотные интервалы Δv в соответствии с размером окна W^*_{size} и шагом W^*_{step}. Далее каждый интервал раскладывался в ряд Фурье, и для интервалов строились спектры кепстральных коэффициентов. Комбинируя разные типы окон на этапе вычисления $\bar{A}'_{s,k}$ и $K_{s,k,}$ можно получить больше информации. Проведен эксперимент по классификации образов испытуемых на основе спектральных и кепстральных признаков с помощью «наивного» байесовского классификатора и ИНС (восемь архитектур, из них шесть – сверточные сети). По результатам тестирования байесовский классификатор в сочетании с кепстральными признаками показал меньше ошибок (EER = 0.0053 против EER = 0.0266 для лучшей ИНС). Установлено, что кепстральные признаки являются более информативными, однако для обучения ИНС на кепстрограммах требуется больший объем обучающей выборки, что нереализуемо на имеющемся наборе данных.

Апробированы две схожие архитектуры нейросетевых автокодировщиков (на базе кодировщиков из 8 и 9 одномерных сверточных слоев и одного полносвязного, а также идентичных декодировщиков из 10 сверточных слоев) для построения на их основе блоков извлечения признаков из усредненных спектров. Коррелированность признаков, извлекаемых разными нейронными сетями, в данном случае является желательным свойством.

Акустический образ уха имеет схожесть с голосовым сигналом, как и их усредненные спектры. Решено использовать следующую схему переноса обучения. Из речевых наборов данных TIMIT и VoxCeleb1 в совокупности было извлечено 71264 голосовых образа дикторов. Эти образы были преобразованы в усредненные спектры (W_{size} = 4096 и W_{step} = 2048). Для аугментации данных использовались четыре оконные функции: Хэмминга, Блэкмана, Барлетта, прямоугольная, в результате получено 285056 усредненных голосовых спектров, которые использовались для обучения двух автокодировщиков. ИНС обучались (оптимизатор Adam) с возрастающим объемом батчей (далее размер батча/количество эпох): 256/10, 512/10, 1024/3, 2048/3, 4096/2, 8192/2.

Проведен эксперимент по верификации личности испытуемых с помощью двух различных моделей НПБК – на базе корреляционных нейронов (№ 1) и ГОСТ Р 52633.5 (№ 2). Все образы были обработаны кодировщиками. Наилучшими результатами аутентификации субъектов являются следующие:

НПБК № 1: EER = 0.0238 (FRR = 0.093, FAR < 0.0001), L = 8192, KG = 6, KI = 49.
НПБК № 2: EER = 0.03136 (FRR = 0.2342, FAR < 0.0001), L = 716, KG = 8, KI = 49.

Достоверность оценок для вероятностей FRR и FAR составляет 0.99 и 0.96. Предложенная модель НПБК при меньшей тренировочной выборке дает меньше ошибок и более, чем в 10 раз увеличивает длину ключа, чем НПБК, обучаемый по ГОСТ Р 52633.5-2011. Метод аутентификации по акустическим параметрам уха на основе предложенной модели НПБК можно применять как в составе алгоритма (см. рис. 7), так и отдельно. Для анализа рукописных и голосовых образов сформирован набор данных. Рукописные образы собраны с использованием планшета Wacom (частота опроса 200 Гц, 1024 уровня давления), голосовые – с использованием микрофонов Pioneer, Sony (диапазон частот 70 – 12000 Гц). Из открытых источников взяты примеры для расширения тестовой выборки «Чужих». Набор данных можно разделить на выборки:

– «Все Свои»: 260 подписантов и 260 дикторов (пол и возраст распределены равномерно от 18 до 35 лет) воспроизвели образ 90 раз, данные собраны в три этапа с интервалом в несколько недель, на каждом этапе испытуемый ввел 30 примеров, на третьем этапе испытуемые находились в сонном ПФС;

– «Неизвестные Чужие»: по одному тестовому примеру других 6500 рукописных и 6500 голосовых образов, воспроизведенных другими субъектами.

Оцифрованный рукописный образ состоит из функций координат x_coord (t), y_coord (t) и давления пера на планшет pressure (t), где t – это время в дискретной форме. Рассмотрено два набора признаков (табл. 1): с поддержкой давления и без (n = 782 / n = 521). Для голосовых сигналов также апробировано два набора признаков (табл. 2): при частоте дискретизации сигнала 24 кГц и 8 кГц (n = 570 / n = 350) для низко информативных каналов передачи данных.

Метод двухфакторной аутентификации по голосовым и рукописным образам можно также использовать как отдельно, так и в составе алгоритма на рис. 7. Экспериментальная оценка надежности двухфакторной аутентификации показала следующие результаты: FRR = 0.03 при FAR < 10^{-10}.

Использование трех факторов аутентификации, а также предложенный метод и алгоритм (см. рис. 7) многократно повышают надежность и защищенность биометрической системы от деструктивных воздействий. Экспериментальная оценка надежности трехфакторной аутентификации показала следующие результаты: FRR = 0.12 при FAR < 10^{-14}.

Таблица 1

Краткое описание методик извлечения признаков рукописного образа

Группа признаков рукописного образа	n
Образ делится на 16 равных по числу точек отрезков, строится матрица расстояний между их краями в 2- и 3-мерном пространстве (*pressure* (t) – третье измерение)	240 / 120
Вычисление коэффициентов корреляции между *x_coord*(t), *y_coord*(t), *pressure*(t), их производных и функцией скорости пера $v_{xy}(t)$, производной от *x_coord*(t), *y_coord*(t)	21 / 10
Вычисление параметров внешнего вида образа – угол наклона, отношение длины к ширине, центр в 2- и 3-мерном пространстве, описываемый 2 или 3 координатами	5 / 4
Вычисление средних значений фрагментов функций *pressure*(t), *x_coord'*(t), *y_coord'*(t), $v_{xy}(t)$ (образ делится на 5 равных по числу точек отрезков)	20 / 15
Вычисление детализирующих коэффициентов вейвлет преобразования Хаара (алгоритм Малла), полученных на 4 нижних уровнях разложения для *x_coord*(t), *y_coord*(t), *pressure*(t), $v_{xy}(t)$ (функции сначала приводились к 128 отчетам (интерполяция))	240 / 180
Вычисление усредненного амплитудного спектра с помощью STFT (размер окна – 128 отчетов, шаг – 16 отчетов) для *x_coord*(t), *y_coord*(t), *pressure*(t), $v_{xy}(t)$	256 / 192

Таблица 2

Краткое описание методик извлечения признаков голосовых паролей

Группа признаков голосовых паролей	*n*
Вычисление усредненного по всем окнам амплитудного спектра низких частот речевого сигнала, вычисленного с помощью STFT (размер окна – 2048 (512) при *F* =24 (8) кГц, шаг – 16). Предварительно речевой сигнал нормируется по энергии, удаляется тишина	40 / 20
Вычисление коэффициентов нижних частот кепстра, который берется от полного усредненного по всем окнам амплитудного спектра, получаемого в соответствии с 2.1	40 / 20
Вычисление коэффициентов нижних частот кепстра, который берется от полного логарифмированного усредненного амплитудного спектра	40 / 20
Вычисление кепстра второго порядка от полных кепстров 2.1 и 2.2 (кепстры 2.1 и 2.2 повторно подвергаются прямому преобразованию Фурье)	256 / 128
Подсчет частоты переходов сигнала через нулевое деление окном (размер окна – 2048 (512) при *F* =24 (8) кГц, шаг – 16), грубо характеризуют ЧОТ (нулевую форманту)	64 / 32
Вычисление амплитудного спектра функции автокорреляции речевого сигнала	128
Вычисление частоты переходов через «ноль» и экстремумов функции автокорреляции	2

Пользователям рекомендуется периодически (не реже, чем раз в полгода) проходить процедуру аутентификации, чтобы нейро-иммунная модель адаптировалась к дрейфу, и точность не снижалась.

Технология автоматического синтеза и обучения нейросетевых моделей

Разработана технология автоматического синтеза и обучения нейросетевых моделей доверенного ИИ. Разработана первая редакция стандарта [15], основанного на данной технологии, который в значительной степени расширяет уже имеющийся стандарт ГОСТ Р 52633.5-2011 и отличается от него тем, что он распространяется на нейросетевые преобразователи не только биометрических, но и любых иных образов в код, а также на задачи идентификации образов на открытом множестве. Стандарт позволяет повысить длину пароля и криптографического ключа, связываемого с классом «Свой»; снизить вероятности ошибок 1-го и 2-го рода; улучшить хэширующие (перемешивающие) свойства НПБК; повысить уровень защищенности знаний НПБК от компрометации; сделать невозможным или, по крайней мере, усложнить несанкционированное управление ИИ путем манипуляций с моделями машинного обучения, анализ логики работы ИИ с целью повлиять на его решения, извлечение знаний из памяти ИИ (в том числе путем зондирования модели), а также их интерпретацию; снизить вероятность успеха состязательных атак, реализуемых злоумышленником путем наложения шумов на исходный образ «Свой» или путем создания синтетических примеров образов «Чужой». Первая редакция стандарта ГОСТ Р «Искусственный интеллект. Нейросетевые алгоритмы в защищённом исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации» прошла публичное обсуждение, получены ряд замечаний, которые могут быть устранены во второй редакции стандарта. На данный момент ведется дискуссия относительно необходимости устранения всех представленных замечаний, и принимается решение относительно необходимости разработки второй редакции стандарта, учитывающей комментарии членов технического комитета «Искусственный интеллект» (ТК 164) к тексту стандарта. Также 21 марта 2024 г. на конференции РусКрипто 2024 был заслушан доклад (название доклада: «Защищенное исполнение нейросетевых алгоритмов классификации образов для задач биометрической аутентификации на базе сетей корреляционных нейронов»), в котором обоснована несостоятельность атаки, направленной на взлом нейро-корреляционного преобразователя, описанного в первой редакции стандарта, представленной годом ранее на РусКрипто 2023.

Результаты работы внедрены в ряде предприятий (восемь внедрений результатов на семи предприятиях), в том числе использованы при разработке линейки программных продуктов AIConstructor (AIC). AIC desktop ориентирован на исследования по машинному обучению и классификации образов в области ИБ. Программный комплекс поддерживает множество форматов данных (xml, txt, bmp, wav, edf, csv и др.), имеет конструктор глубоких нейронных сетей. AIC ModelOps Platform – это система управления жизненным циклом ИИ для цифровой трансформации предприятий. Программный продукт включает следующие модули: управление экспериментом (интеграция с фреймворками, отслеживание экспериментов, инструменты для командной работы исследователей, редактор кода эксперимента); конструктор конвейеров обработки данных; диагностика (мониторинг модели, определение дрейфа). Описана разработанная библиотека машинного обучения на базе корреляционных нейронов, интегрированная в AIC ModelOps Platform.

На базе результатов разработаны образовательные программы и внедрены в учебный процесс ФГАОУ ВО СПбГЭТУ «ЛЭТИ» и ФГАОУ ВО ОмГТУ (по дисциплинам «Распознавание образов», «Машинное обучение в приложениях биометрии», «Биометрия и защита информации», «Этика и правовые проблемы искусственного интеллекта», «Защищенное исполнение искусственного интеллекта», «Доверенный искусственный интеллект»).

Заключение

В работе на основе выполненного исследования представлено решение актуальной научной проблемы повышения надежности многофакторной биометрической аутентификации и защищенности биометрических систем от компьютерных атак на основе технологии защищенного исполнения нейросетевых моделей и алгоритмов ИИ, имеющей важное хозяйственное значение в плане обеспечения информационной безопасности компьютерных ресурсов и конфиденциальных данных, а также знаний, моделей и алгоритмов ИИ.

Разработана концепция защищенного исполнения нейросетевых алгоритмов ИИ, которая позволяет сформировать устойчивость модели к извлечению знаний. Это достигается путем преобразования корреляционных связей между признаками в высокоинформативные мета-признаки Байеса – Минковского, которые сложно фальсифицировать. Установлено, что один мета-признак может содержать в 2 – 3 раза больше информации, чем содержится суммарно в паре исходных признаков, от которых он порожден. Доказано, что корреляция между признаками увеличивает количество информации о классифицируемом образе. Предложены отображения для перехода в пространства мета-признаков Байеса – Минковского, что не требует хранения какой-либо дифференциальной информации о параметрах классов образов. Свойства пространств мета-признаков исследованы экспериментально.

Разработаны модель корреляционных нейронов и модель нейросетевого преобразователя биометрия – код на их основе, анализирующие корреляционные связи между признаками вместо признаков, а также робастной алгоритм их автоматического синтеза и обучения на малых выборках. Это позволило повысить защищенность биометрических данных от компрометации и длину ключа, связываемого с биометрическими образами субъектов, снизить вероятность ошибок биометрической аутентификации в защищенном режиме исполнения и повысить устойчивость ИИ к состязательным атакам.

Разработаны адаптивная нейро-иммунная модель и алгоритмы ее пакетного обучения с учителем и онлайн-обучения с подкреплением, позволяющие предупредить или снизить влияние концептуального дрейфа, даже если исходная обучающая выборка недостаточно репрезентативна или незначительна в объеме. Алгоритм обучения с учителем позволяет сформировать врожденный иммунитет модели, позволяющий разделять входные образы на два класса. В процессе функционирования модель адаптируется к изменению данных, используя алгоритм онлайн-обучения с подкреплением, в результате чего формируется приобретенный иммунитет, корректирующий решения модели в спорных случаях.

Разработаны методы и алгоритм высоконадежной многофакторной аутентификации на основе рукописных и голосовых образов, а также акустических образов уха с обеспечением защиты биометрических данных от компрометации. Новизна заключается в использовании нового типа биометрических данных – акустических параметров ушного канала, получаемых методом эхолокации, а также учете информативности, стабильности и приоритезации признаков, совместным использованием НПБК и нейро-иммунной модели, способе кепстрального анализа сигналов. Акустические образы уха не компрометируются в естественной среде, так как фотография уха неинформативна для синтеза состязательных примеров. Предложенные методы и алгоритм дают более низкий процент ошибок по сравнению с известными мировыми аналогами: FRR = 0.12 при FAR < 10^{-14} и FRR = 0.03 при FAR < 10^{-10}.

Разработана технология автоматического синтеза и обучения нейросетевых моделей доверенного ИИ для высоконадежной биометрической аутентификации и других ответственных приложений ИИ. На базе технологии разработаны линейка программных продуктов AIConstructor и первая редакция национального стандарта. На данный момент принимается решение относительно необходимости разработки второй редакции стандарта, учитывающей полученные замечания и комментарии членов ТК 164 к тексту первой редакции. Выполнено восемь внедрений результатов на семи предприятиях.

Все более актуальными становятся вопросы создания систем доверенного ИИ, минимизации рисков, связанных с внедрением и поддержкой систем ИИ, защиты от компьютерных атак и обеспечения функциональной безопасности ИИ. Основные перспективы развития темы состоят именно в данном направлении. Защищенное исполнение алгоритмов ИИ необходимо на объектах критической информационной инфраструктуры. Перспективным направлением для развития темы также является управление жизненным циклом ИИ. Разработанный аппарат может применяться для сокращения объемов обучающей выборки, формирования инструментов повышения объяснимости решений, обнаружения и корректировки дрейфа моделей ИИ.

Благодарности и поддержка

Работа выполнена в Омском государственном техническом университете в рамках Государственного задания Минобрнауки России на 2023–2025 годы № FSGF-2023-0004.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Вульфин А. М. Модели и методы комплексной оценки рисков безопасности объектов критической информационной инфраструктуры на основе интеллектуального анализа данных // Системная инженерия и информационные технологии. 2023. Т. 5. № 4(13). С. 50–76. EDN NNZWLV. [[Vulfin A. M. “Models and methods for comprehensive assessment of security risks of critical information infrastructure objects based on intelligent data analysis” // System Engineering and Information Technologies. 2023. Vol. 5, No. 4(13), pp. 50–76. EDN NNZWLV. (In Russian).]]
2. Иванов А. И., Сомкин С. А., Андреев Д. Ю., Малыгина Е. А. О многообразии метрик, позволяющих наблюдать реальные статистики распределения биометрических данных «нечетких экстракторов» при их защите наложением гаммы // Вестник УрФО. Безопасность в информационной сфере. 2014. № 2 (12). С. 16–23. [[Ivanov A. I., Somkin S. A., Andreev D. Yu., Malygina E. A. “On the variety of metrics that allow one to observe real statistics of the distribution of biometric data of “fuzzy extractors” when they are protected by gamma imposition” // Bulletin of the Urals Federal District. Security in the Information Sphere. 2014. No. 2 (12), pp. 16–23. (In Russian).]]
3. Малыгина Е. А. Биометрико-нейросетевая аутентификация: перспективы применения сетей квадратичных нейронов с многоуровневым квантованием биометрических данных: препринт / Пензенский гос. ун-т. Пенза: Изд-во ПГУ, 2020. [[Malygina E. A. Biometric-neural network authentication: prospects for using quadratic neuron networks with multi-level quantization of biometric data: preprint / Penza State University. univ. Penza: PSU Publishing House, 2020. (In Russian).]]
4. Жумажанова С. С., Сулавко А. Е., Ложников П. С. Распознавание психофизиологического состояния субъектов-операторов на основе анализа термографических изображений лица с применением сверточных нейронных сетей // Системная инженерия и информационные технологии. 2023. Т. 5. № 2(11). С. 41–55. EDN NNZWLV. [[Zhumazhanova S. S., Sulavko A. E., Lozhnikov P. S. “Recognition of the psychophysiological state of operator subjects based on the analysis of thermographic images of the face using convolutional neural networks” // System Engineering and Information Technologies. 2023. Vol. 5, No. 2(11), pp. 41-55. EDN NNZWLV. (In Russian).]]
5. Sulavko A. E. “Biometric-based key generation and user authentication using acoustic characteristics of the outer ear and a network of correlation neurons” // Sensors. 2022. Vol. 22. Pp. 9551. DOI: 10.3390/s22239551.

6. Sulavko A. E. “Bayes-Minkowski measure and building on its basis immune machine learning algorithms for biometric facial identification” // Journal of Physics: Conf. Series. 2020. Vol. 1546. Pp. 012103-1–012103-7. DOI: 10.1088/1742-6596/1546/1/012103.
7. Брюхомицкий Ю. А. Клавиатурный мониторинг на основе иммунологического клонирования // Безопасность информационных технологий. 2016. № 4. С. 5–11. [[Bryukhomitsky Yu. A. “Keyboard monitoring based on immunological cloning” // Security of Information Technologies. 2016. No. 4, pp. 5–11. (In Russian).]]
8. Брюхомицкий Ю. А., Федоров В. М. Иммунологический метод текстонезависимой верификации личности по голосу // Известия ЮФУ. Технические науки. 2019. № 5 (207). С. 123–134. [[Bryukhomitsky Yu. A., Fedorov V. M. “Immunological method of text-independent personality verification by voice” // News of the Southern Federal University. Technical Science. 2019. No. 5 (207), pp. 123–134. (In Russian).]]
9. Самотуга А. Е. Распознавание субъектов и их психофизиологических состояний на основе параметров подписи для защиты документооборота // Системная инженерия и информационные технологии. 2023. Т. 5. № 2(11). С. 56–65. EDN JCFRBU. [[Samotuga A. E. “Recognition of subjects and their psychophysiological states based on signature parameters for document management” // System Engineering and Information Technologies. 2023. Vol. 5, No. 2(11), pp. 56–65. EDN JCFRBU. (In Russian).]]
10. Брюхомицкий Ю. А. Модель искусственной иммунной системы с двойной пластичностью // Информационное противодействие угрозам терроризма. 2013. № 20. С. 76–83. [[Bryukhomitsky Yu. A. “Model of an artificial immune system with double plasticity” // Information Counteraction to the Threats of Terrorism. 2013. No. 20, pp. 76–83. (In Russian).]]
11. Сулавко А. Е., Шалина Е. В., Стадников Д. Г., Чобан А. Г. Иммунные алгоритмы распознавания образов и их применение в биометрических системах (обзор) // Вопросы защиты информации. 2019. № 1. С. 38–46. [[Sulavko A. E., Shalina E. V., Stadnikov D. G., Choban A. G. “Immune algorithms for pattern recognition and their application in biometric systems (review)” // Issues of Information Security. 2019. No. 1, pp. 38–46. (In Russian).]]
12. Сулавко А. Е. Абстрактная модель искусственной иммунной сети на основе комитета классификаторов и ее использование для распознавания образов клавиатурного почерка // Компьютерная оптика. 2020. Т. 44. № 5. С. 830–842. DOI: 10.18287/2412-6179-CO-717. [[Sulavko A. E. “Abstract model of an artificial immune network based on a committee of classifiers and its use for recognizing keyboard handwriting patterns” // Computer Optics. 2020. Vol. 44, No. 5, pp. 830–842. DOI: 10.18287/2412-6179-CO-717. (In Russian).]]
13. Sulavko A. E., Samotuga A. E., Kuprik I. A. “Personal identification based on acoustic characteristics of the outer ear using cepstral analysis, bayesian classifier and artificial neural networks” // IET Biometrics. 2021. Vol. 10. No. 6. Pp. 692–705. DOI: 10.1049/bme2.12037.
14. Сулавко А. Е. Высоконадежная двухфакторная биометрическая аутентификация по рукописным и голосовым паролям на основе гибких нейронных сетей // Компьютерная оптика. 2020. Т. 44. № 1. С. 82–91. DOI: 10.18287/2412-6179-CO-567. [[Sulavko A. E. “Highly reliable two-factor biometric authentication using handwritten and voice passwords based on flexible neural networks” // Computer Optics. 2020. Vol. 44, No. 1, pp. 82–91. DOI: 10.18287/2412-6179-CO-567. (In Russian).]]
15. Иванов А. И., Сулавко А. Е. Проект третьего национального стандарта России по быстрому автоматическому обучению больших сетей корреляционных нейронов на малых обучающих выборках биометрических данных // Вопросы кибербезопасности. 2021. № 3. С. 84–93. DOI: 10.21681/2311-3456-2021-3-84-93. [[Ivanov A. I., Sulavko A. E. “Project of the third national standard of Russia for fast automatic training of large networks of correlation neurons on small training samples of biometric data” // Issues of Cyber Security. 2021. No. 3, pp. 84–93. DOI: 10.21681/2311-3456-2021-3-84-93. (In Russian).]]

Поступила в редакцию 17 февраля 2024 г.

МЕТАДАННЫЕ / METADATA

Title: Highly reliable biometric authentication based on secured execution of neural network models and artificial intelligence algorithms.

Abstract: The article presents the results of a study devoted to solving the scientific and technical problem of increasing the reliability of multifactor biometric authentication and the security of biometric systems from computer attacks. The object of research is biometric authentication systems based on methods, models and algorithms of trusted AI. The subject of the research is neural network models and machine learning algorithms on small samples for highly reliable biometric authentication and protection of biometric data from compromise. The goal of the work is to increase the reliability of multifactor biometric authentication based on the secure execution of neural network models of trusted AI and algorithms for their automatic synthesis and training on small samples of biometric data. To achieve the goal, the following tasks were completed: 1. Development of the concept of secure execution of neural network AI algorithms. 2. Development of models of artificial neurons and a neural network biometrics-code converter, potentially resistant to destructive influences, and algorithms for their robust automatic learning on small samples. 3. Development of an adaptive AI model and its training algorithms to prevent or reduce the impact of conceptual data drift in biometric authentication systems. 4. Development of multi-factor authentication methods based on secret biometric images while ensuring the confidentiality of biometric data. 5. Development of technology for automatic synthesis and training of neural network models for highly reliable multi-factor biometric authentication.

Key words: protected execution of artificial intelligence, neural network biometrics-to-code converters, biometric authentication, biometric voice parameters, features of reproducing handwritten passwords, echograms of the ear canal, correlation between features, correlation neurons, automatic machine learning, reinforcement learning

Язык статьи / Language: русский / Russian.

Об авторе / About the author:

СУЛАВКО Алексей Евгеньевич
Омский государственный технический университет, Россия.
Проф. каф. комплексной защиты информации. Д-р техн. наук (Уфимск. ун-т науки и технологий, 2023). Иссл. в обл. биометрической идентификации, информационных систем, искусственных нейронных сетей.
E-mail: sulavich@mail.ru
ORCID: https://orcid.org/0000-0002-9029-8028
URL: https://elibrary.ru/author_profile.asp?authorid=660485

SULAVKO Alexey Evgenievich
Omsk State Technical University, Russia.
Professor, Dept. of Integrated Information Security. Dr. Techn. Science (Ufa Univ. of Science & Technology, 2023). Research in biometric identification, information systems, artificial neural networks.
E-mail: sulavich@mail.ru
ORCID: https://orcid.org/0000-0002-9029-8028
URL: https://elibrary.ru/author_profile.asp?authorid=660485