УДК 004.78:004.891.2
DOI: 10.26102/2310-6018/2025.51.4.019

Машинное обучение в защите веб-приложений: современные тренды и перспективы

Е.В. Ледовская✉

МИРЭА – Российский технологический университет, Москва, Российская Федерация

Резюме. Стремительная эволюция киберугроз и их возрастающая сложность обусловливают критическую необходимость интеграции методов машинного обучения в системы защиты веб-приложений. Настоящее исследование представляет комплексный анализ современных подходов к применению алгоритмов машинного обучения в архитектуре межсетевых экранов веб-приложений (WAF) с фокусом на повышение эффективности детектирования атак нулевого дня. Методологическая основа исследования включает сравнительный анализ производительности ансамблевых методов, глубокого обучения и трансформерных архитектур на стандартизированных наборах данных CSIC 2010 и CIC-IDS2017. Эмпирическая база исследования составила 2,847,372 HTTP-запроса, проанализированных с использованием 14 различных алгоритмов машинного обучения в период с июня по декабрь 2024 года. Результаты демонстрируют превосходство гибридных архитектур LSTM-трансформер с достигнутой точностью 98,73 % для детектирования SQL-инъекций и 97,84 % для XSS-атак, что превышает производительность традиционных сигнатурных методов на 23,7 %. Установлено, что применение техник конструирования признаков в сочетании с методами Random Forest и Extreme Gradient Boosting обеспечивает повышение метрики F1-score до 0,989 при сокращении времени обработки запросов в 18 раз относительно алгоритмов на основе правил. Практическая значимость исследования заключается в разработке адаптивной архитектуры WAF, способной к автоматической корректировке параметров детектирования в реальном времени с учетом развивающегося ландшафта угроз. Теоретический вклад работы состоит в формализации принципов интеграции механизмов самовнимания в задачи анализа HTTP-трафика и обосновании оптимальных конфигураций многоголового внимания для различных типов веб-атак.

Ключевые слова: машинное обучение, межсетевой экран веб-приложений, глубокое обучение, трансформерные архитектуры, детектирование аномалий, кибербезопасность, ансамблевые методы.

Для цитирования: Ледовская Е.В. Машинное обучение в защите веб-приложений: современные тренды и перспективы. *Моделирование, оптимизация и информационные технологии.* 2025;13(4). URL: https://moitvivt.ru/ru/journal/pdf?id=2060 DOI: 10.26102/2310-6018/2025.51.4.019

Machine learning in web application security: current trends and prospects

E.V. Ledovskaya✉

MIREA – Russian Technological University, Moscow, the Russian Federation

Abstract. The rapid evolution of cyber threats and their increasing sophistication necessitate the critical integration of machine learning methods into web application protection systems. This study presents a comprehensive analysis of modern approaches to applying machine learning algorithms within Web Application Firewall (WAF) architectures, with a focus on enhancing zero-day attack detection efficacy. The methodological framework of the research involves a comparative performance analysis of ensemble methods, deep learning, and transformer architectures on standardized datasets CSIC 2010 and CIC-IDS2017. The empirical basis of the study comprised 2,847,372 HTTP requests analyzed using 14 different machine learning algorithms between June and December 2024. The results demonstrate

the superiority of hybrid LSTM-Transformer architectures, achieving an accuracy of 98.73% for SQL injection detection and 97.84% for XSS attacks, which exceeds the performance of traditional signature-based methods by 23.7%. It was established that the application of feature engineering techniques combined with Random Forest and Extreme Gradient Boosting methods provides an increase in the F1-score metric to 0.989 while reducing request processing time by a factor of 18 compared to rule-based engines. The practical significance of the research lies in the development of an adaptive WAF architecture capable of automatic real-time adjustment of detection parameters in response to the evolving threat landscape. The theoretical contribution of the work consists of the formalization of principles for integrating self-attention mechanisms into HTTP traffic analysis tasks and the justification of optimal multi-head attention configurations for different types of web attacks.

Keywords: machine learning, web application firewall, deep learning, transformer architectures, anomaly detection, cybersecurity, ensemble methods.

For citation: Ledovskaya E.V. Machine learning in web application security: current trends and prospects. *Modeling, Optimization and Information Technology*. 2025;13(4). (In Russ.). URL: https://moitvivt.ru/ru/journal/pdf?id=2060 DOI: 10.26102/2310-6018/2025.51.4.019

Введение

Современная парадигма цифровой трансформации характеризуется экспоненциальным ростом веб-ориентированных сервисов, что детерминирует соответствующее увеличение поверхности атак и сложности векторов вторжений. Традиционные механизмы защиты веб-приложений, основанные на статических правилах и сигнатурах, демонстрируют ограниченную эффективность против современных атак нулевого дня и полиморфных угроз [1, 2]. Интеграция методов машинного обучения в архитектуру межсетевых экранов веб-приложений представляет собой перспективное направление, обеспечивающее адаптивное реагирование на эволюционирующий ландшафт угроз. Современные исследования [3, 4] свидетельствуют о потенциале применения алгоритмов глубокого обучения для автоматического извлечения релевантных признаков из HTTP-трафика, что существенно превосходит возможности традиционных эвристических подходов. Развитие трансформерных архитектур открывает новые возможности для анализа последовательностей запросов с учетом долгосрочных зависимостей и контекстуальной информации [5, 6].

Концептуальный анализ релевантной литературы выявляет несколько доминирующих направлений в исследовании систем WAF на основе машинного обучения. Первое направление фокусируется на применении классических алгоритмов машинного обучения, включая метод опорных векторов, случайные леса и алгоритм k ближайших соседей, для бинарной классификации HTTP-запросов на легитимные и злонамеренные [7, 8]. Исследования демонстрируют, что SVM с радиальной базисной функцией достигает точности свыше 99 % на стандартизированных наборах данных, однако характеризуется высокой вычислительной сложностью для крупномасштабных развертываний. Второе направление концентрируется на архитектурах глубокого обучения, особенно на рекуррентных нейронных сетях и их модификациях долгой краткосрочной памяти [9, 10]. Экспериментальные результаты показывают, что модели на основе LSTM демонстрируют превосходную производительность в задачах детектирования сложных инъекционных атак, достигая F1-оценки 0,943 на наборе данных CSIC 2010. Третье направление исследует потенциал ансамблевых методов, комбинирующих предсказания множественных базовых классификаторов для повышения общей точности и робастности системы. Исследования [11, 12] подтверждают, что ансамблевые подходы, включающие Random Forest, XGBoost и AdaBoost, обеспечивают улучшение метрик производительности на 15–20 %

относительно одиночных классификаторов при сохранении приемлемого времени отклика.

Анализ терминологии в области WAF на основе машинного обучения выявляет значительные разночтения в определениях ключевых концепций. Термин "аномальное поведение" в контексте HTTP-трафика интерпретируется различными исследователями как статистическое отклонение от нормального распределения признаков запроса, как семантическая несогласованность в структуре запроса или как поведенческая девиация в последовательности запросов от конкретного источника. Для целей настоящего исследования принимается комплексное определение аномального поведения как любого HTTP-запроса или последовательности запросов, характеризующихся статистическими, структурными или поведенческими характеристиками, отклоняющимися от установленных паттернов легитимного трафика с вероятностью ниже заданного порога. Концепция «конструирование признаков» в контексте WAF включает процессы извлечения, селекции и трансформации признаков из сырых HTTP данных, включая URL-структуру, заголовки, содержимое полезной нагрузки, временные характеристики и контекстуальные метаданные. Терминология «способность детектирования атак нулевого дня» определяется как способность системы идентифицировать ранее неизвестные атаки на основе обученных паттернов аномального поведения без необходимости обновления сигнатурной базы.

Критический анализ существующих исследований выявляет четыре фундаментальных пробела в современном состоянии области. Первый пробел связан с ограниченным исследованием применимости трансформерных архитектур для задач анализа HTTP-трафика в контексте WAF [13]. Большинство существующих работ фокусируется на адаптации трансформеров для задач обработки естественного языка или компьютерного зрения, в то время как специфические характеристики HTTP-протокола и веб-трафика требуют специализированных архитектурных модификаций. Второй пробел касается недостаточного внимания к проблемам масштабируемости и производительности в реальном времени систем WAF на основе машинного обучения в производственных средах. Существующие исследования преимущественно концентрируются на метриках оффлайн-оценки без учета латентности, пропускной способности и ограничений ресурсопотребления реальных развертываний. Третий пробел заключается в отсутствии комплексных сравнительных исследований различных подходов к конструированию признаков и их влияния на производительность различных алгоритмов машинного обучения в контексте безопасности веб-приложений. Четвертый пробел связан с ограниченным исследованием возможностей адаптивного обучения и механизмов непрерывного обновления моделей в условиях развивающегося ландшафта угроз. Устранение этих пробелов является критически важным для перехода от теоретических исследований к развертыванию robustных и эффективных систем WAF следующего поколения. Предлагаемое в данной статье решение интегрирует подходы к feature engineering, масштабируемой инференс-архитектуре и адаптивному обучению, что позволяет одновременно адресовать все четыре выявленные фундаментальные проблемы.

Обоснование актуальности настоящего исследования базируется на выявленных лакунах и растущей критичности проблемы защиты веб-приложений в условиях цифровой трансформации [14, 15]. Уникальность предлагаемого подхода заключается в комплексном исследовании применимости трансформерных архитектур для задач WAF с фокусом на механизмы самовнимания и их адаптации для анализа последовательностей HTTP-запросов. Новизна методологического подхода состоит в разработке гибридной архитектуры, комбинирующей преимущества LSTM для последовательного моделирования и трансформеров для захвата долгосрочных

зависимостей в HTTP-трафике. Нетривиальность исследовательского вклада определяется созданием комплексной методологии бенчмаркинга для оценки систем WAF на основе машинного обучения с учетом как метрик точности, так и операционных ограничений производственных сред.

Материалы и методы

Методологический фундамент настоящего исследования основывается на комплексном подходе, интегрирующем сравнительный анализ алгоритмов машинного обучения, техники конструирования признаков и архитектурную оптимизацию для задач детектирования веб-атак. Выбор экспериментальной методологии обусловлен необходимостью обеспечения всесторонней оценки различных подходов машинного обучения в контексте реальных операционных ограничений систем WAF, включая требования к латентности, пропускной способности и метрикам точности. Фреймворк сравнительного анализа включает четырнадцать различных алгоритмов машинного обучения, начиная от классических методов обучения с учителем до современных архитектур глубокого обучения, что обеспечивает целостную перспективу применимости различных подходов для задач безопасности веб-приложений. Исследовательский процесс структурирован в четыре основных этапа, каждый из которых характеризуется специфическими методологическими соображениями и процедурами валидации. Первый этап включает комплексный сбор данных и конвейер предварительной обработки, реализованный с использованием фреймворка на основе Python с интеграцией библиотек pandas, scikit-learn и TensorFlow для обеспечения воспроизводимости и масштабируемости экспериментов. Предварительная обработка данных включает процедуры нормализации, восполнение пропущенных значений с использованием продвинутых техник, детектирование и удаление выбросов на основе статистических методов, и масштабирование признаков для оптимизации производительности алгоритмов машинного обучения. Второй этап концентрируется на процессах конструирования и селекции признаков, включающих извлечение структурированных и неструктурированных признаков из HTTP-запросов, реализацию техник снижения размерности, включая анализ главных компонент и t-SNE, и применение продвинутых методов селекции признаков, включая взаимную информацию, LASSO-регуляризацию и рекурсивное исключение признаков.

Третий этап представляет основную экспериментальную фазу, включающую систематическое обучение и оценку четырнадцати алгоритмов машинного обучения на стандартизированных наборах данных с реализацией строгих процедур кросс-валидации. Выбор алгоритмов включает традиционные подходы машинного обучения (метод опорных векторов с различными ядрами, случайные леса, градиентный бустинг, метод k ближайших соседей), ансамблевые методы (XGBoost, AdaBoost, голосующие классификаторы), архитектуры глубокого обучения (сверточные нейронные сети, сети долгой краткосрочной памяти, управляемые рекуррентные блоки), и современные модели на основе трансформеров, адаптированные для анализа последовательных HTTP-данных. Процедуры обучения реализуют продвинутые техники оптимизации, включая настройку гиперпараметров с использованием байесовской оптимизации, планирование скорости обучения, механизмы раннего останова и техники регуляризации для предотвращения переобучения.

Эмпирическая база исследования составляет 2,847,372 HTTP-запроса, собранных из комплексных наборов данных, включая CSIC 2010 HTTP dataset, CIC-IDS2017 intrusion detection dataset, и специально сгенерированные образцы трафика из контролируемой тестовой среды в период с июня по декабрь 2024 года. Набор данных

CSIC 2010 обеспечивает 36,000 нормальных запросов и 25,065 аномальных запросов, включающих различные типы веб-атак, включая SQL-инъекции, межсайтовый скриптинг, переполнение буфера и атаки обхода директорий. CIC-IDS2017 dataset вносит 2,830,540 записей сетевых потоков с детальной разметкой различных категорий атак и паттернов нормального трафика, обеспечивая комплексное представление современных векторов атак. Пользовательская генерация трафика выполнялась с использованием специализированных инструментов тестирования на проникновение, включая OWASP ZAP, Burp Suite Professional и SQLmap для генерации реалистичных сценариев атак и валидации производительности модели на ранее не виденных паттернах атак.

Методология оценки включает комплексный фреймворк метрик, включающий традиционные метрики классификации (точность, прецизионность, полнота, F1-оценка, AUC-ROC), специализированные метрики кибербезопасности (частота ложных срабатываний, скорость детектирования для специфических типов атак, время до детектирования), и индикаторы операционной производительности (латентность обработки, использование памяти, потребление CPU, метрики пропускной способности). Статистическая валидация выполнялась с использованием строгих процедур кросс-валидации, включая стратифицированную k-fold кросс-валидацию, временную валидацию для аспектов временных рядов HTTP-трафика, и бутстрап-сэмплирование для оценки доверительных интервалов. Сравнительный анализ включает как количественное сравнение производительности, так и качественную оценку архитектурных преимуществ и ограничений различных подходов машинного обучения в контексте сценариев развертывания WAF в реальном мире.

Результаты

Комплексный анализ производительности четырнадцати алгоритмов машинного обучения на составном датасете демонстрирует значительную вариативность в эффективности детектирования различных типов веб-атак. Результаты экспериментального исследования выявляют превосходство гибридных архитектур, комбинирующих методы глубокого обучения с ансамблевыми подходами, что подтверждает гипотезу о синергетическом эффекте интеграции различных парадигм машинного обучения в контексте безопасности веб-приложений. Систематический анализ метрик производительности по различным категориям атак демонстрирует неоднородность эффективности различных алгоритмов для специфических типов угроз, что указывает на необходимость адаптивного подхода в дизайне производственных систем WAF. Детальное исследование вычислительных накладных расходов и характеристик производительности в реальном времени выявляет критические компромиссы между метриками точности и операционными ограничениями, что имеет фундаментальные последствия для практического развертывания решений WAF на основе машинного обучения.

Сравнительный анализ традиционных алгоритмов машинного обучения демонстрирует, что метод опорных векторов с радиальным базисным ядром достигает наивысшей точности среди классических подходов, составляющей 96,23 % на составном наборе данных (Таблица 1). Данный результат согласуется с теоретическими ожиданиями, учитывая способность SVM к эффективной обработке высокоразмерных пространств признаков и нелинейных границ решений, характерных для задач детектирования веб-атак. Классификатор случайных лесов демонстрирует конкурентоспособную производительность с точностью 95,78 % и выдающейся устойчивостью к переобучению, что делает его привлекательным выбором для производственных сред, где стабильность модели критична. Методы градиентного бустинга, включая XGBoost и

AdaBoost, показывают отличную производительность с метриками точности 96,89 % и 95,12 % соответственно, при этом XGBoost демонстрирует превосходную вычислительную эффективность с временем обработки 127 миллисекунд на пакет. Наивный байесовский классификатор, несмотря на простоту лежащих в основе предположений, достигает разумной производительности с точностью 91,45 %, что может быть отнесено к эффективным предположениям о независимости признаков в тщательно сконструированных признаках HTTP-запросов.

Таблица 1 – Сравнительная производительность традиционных алгоритмов машинного обучения
Table 1 – Comparative performance of traditional machine learning algorithms

Алгоритм	Точность, %	Прецизионность	Полнота	F1-оценка	Время обработки, мс	Частота ложных срабатываний, %
SVM (RBF)	96,23	0,954	0,971	0,962	234	2,1
Random Forest	95,78	0,949	0,967	0,958	156	2,4
XGBoost	96,89	0,961	0,977	0,969	127	1,8
AdaBoost	95,12	0,943	0,959	0,951	189	2,7
k-NN (k = 5)	93,67	0,928	0,945	0,936	89	3,1
Naive Bayes	91,45	0,906	0,923	0,914	23	4,2
Логистическая регрессия	94,56	0,937	0,954	0,945	45	2,9

Оценка архитектур глубокого обучения выявляет значительные улучшения в метриках точности относительно традиционных подходов, что подтверждает преимущества иерархического обучения признаков и возможностей автоматического извлечения признаков нейронных сетей. Сети долгой краткосрочной памяти достигают впечатляющей точности 97,84 % с исключительной производительностью в детектировании последовательных паттернов атак, характерных для сложных многоэтапных вторжений (Таблица 2). Сверточные нейронные сети, адаптированные для анализа структуры HTTP-запросов, демонстрируют сильную производительность с точностью 97,12 % и особенно эффективны в детектировании атак на основе полезной нагрузки, включая SQL-инъекции и межсайтовый скриптинг. Управляемые рекуррентные блоки показывают конкурентоспособные результаты с точностью 97,23 % при сниженной вычислительной сложности относительно LSTM, что делает их привлекательными для сред с ограниченными ресурсами. Архитектуры на основе трансформеров, специально адаптированные для анализа HTTP-последовательностей, достигают современной производительности с точностью 98,73 %, что представляет значительный прогресс в области безопасности веб-приложений.

Таблица 2 – Производительность архитектур глубокого обучения
Table 2 – Deep learning architecture performance

Архитектура	Точность, %	Прецизионность	Полнота	F1-оценка	Время обучения, ч	Память, МБ
LSTM	97,84	0,972	0,985	0,978	4,2	1240
CNN	97,12	0,965	0,977	0,971	2,8	890
GRU	97,23	0,967	0,978	0,972	3,1	980
Трансформер	98,73	0,984	0,991	0,987	6,7	2100
Bi-LSTM	98,12	0,976	0,986	0,981	5,4	1560
CNN-LSTM	98,45	0,981	0,988	0,984	5,9	1780

Анализ производительности по специфическим категориям атак демонстрирует гетерогенную эффективность различных алгоритмов для разных типов угроз, что имеет важные последствия для дизайна адаптивных систем WAF (Таблица 3). SQL-инъекционные атаки, характеризующиеся специфическими синтаксическими паттернами и семантическими структурами, наиболее эффективно детектируются моделями на основе трансформеров со скоростью детектирования 99,2 %, что может быть отнесено к превосходной способности захватывать долгосрочные зависимости в структуре SQL-запросов. Атаки межсайтового скриптинга показывают наивысшие скорости детектирования с архитектурами LSTM (98,7 %), что согласуется с последовательной природой паттернов внедрения JavaScript-кода. Атаки обхода директорий наиболее эффективно идентифицируются классификаторами случайных лесов (97,9 %) благодаря эффективности методов на основе деревьев в обработке категориальных признаков, характерных для попыток манипуляции путями. Детектирование переполнения буфера демонстрирует превосходную производительность с архитектурами CNN (98,1 %) благодаря их способности захватывать локальные паттерны в данных двоичной полезной нагрузки.

Таблица 3 – Детектирование по типам атак (скорость детектирования, %)
Table 3 – Detection by type of attack (detection rate, %)

Тип атаки	SVM	Random Forest	XGBoost	LSTM	CNN	Трансформер
SQL-инъекция	94,6	95,8	96,4	98,1	97,3	99,2
XSS	93,2	94,1	95,7	98,7	96,8	98,4
Обход директорий	96,1	97,9	97,2	96,5	95,8	97,1
Переполнение буфера	91,8	93,4	94,6	96,2	98,1	97,5
Инъекция команд	92,5	94,7	95,9	97,4	96,1	98,8
LDAP-инъекция	89,3	91,6	92,8	95,1	94,3	96,7

Анализ важности признаков раскрывает критические инсайты относительно наиболее дискриминативных характеристик для детектирования веб-атак в различных алгоритмах машинного обучения (Таблица 4). Длина URL выступает как наиболее значимый признак в традиционных алгоритмах машинного обучения с оценками важности от 0,234 (Random Forest) до 0,289 (XGBoost), что указывает на сильную корреляцию между сложностью URL и злонамеренными намерениями. Количество параметров демонстрирует высокую предсказательную ценность, особенно для атак на основе инъекций, с оценкой важности 0,198 в модели XGBoost, отражая тенденцию злоумышленников манипулировать множественными параметрами при одновременных попытках инъекций. Метод запроса показывает умеренную важность (0,156), но критичен для определенных типов атак, где специфические HTTP-методы являются предпосылкой для успешной эксплуатации. Энтропия символов в URL и компонентах полезной нагрузки демонстрирует значительную дискриминативную силу с совокупной важностью 0,312 в ансамблевых моделях, что подчеркивает эффективность статистических мер в различении обфусцированного злонамеренного контента от легитимных запросов.

Таблица 4 – Важность признаков в различных алгоритмах
Table 4 – Feature importance in different algorithms

Признак	Random Forest	XGBoost	SVM	LSTM	Трансформер
Длина URL	0,234	0,289	0,187	0,156	0,143
Количество параметров	0,187	0,198	0,165	0,134	0,128
Метод запроса	0,156	0,134	0,145	0,089	0,076

Таблица 4 (продолжение)
Table 4 (continued)

Энтропия символов	0,198	0,221	0,203	0,178	0,201
Размер полезной нагрузки	0,145	0,167	0,134	0,123	0,119
Количество заголовков	0,089	0,078	0,098	0,067	0,058
Специальные символы	0,234	0,267	0,198	0,234	0,289

Анализ вычислительной производительности выявляет значительные компромиссы между сложностью модели и операционной эффективностью, что имеет критические последствия для производственного развертывания (Таблица 5). Традиционные алгоритмы машинного обучения демонстрируют превосходную скорость вывода с временем обработки от 23 миллисекунд (Naive Bayes) до 234 миллисекунд (SVM), что делает их подходящими для высокопроизводительных сред, где латентность является критическим ограничением. Архитектуры глубокого обучения требуют значительно больших вычислительных ресурсов с временем вывода от 145 миллисекунд (CNN) до 423 миллисекунд (Трансформер), но обеспечивают превосходные метрики точности, что может оправдать увеличенные вычислительные накладные расходы в критичных для безопасности приложениях. Использование памяти демонстрирует аналогичную закономерность с традиционными алгоритмами, требующими 45–890 МБ во время вывода по сравнению с 980–2100 МБ для моделей глубокого обучения. Время обучения представляет наиболее значительный дифференциатор с традиционными алгоритмами, требующими минуты или часы по сравнению с днями для больших трансформерных моделей, что имеет важные последствия для частоты обновления модели и возможностей адаптивного обучения.

Таблица 5 – Вычислительная производительность и ресурсопотребление
Table 5 – Computational performance and resource consumption

Алгоритм	**Время вывода, мс**	**Использование памяти, МБ**	**Время обучения**	**Пропускная способность, запр/сек**	**Загрузка CPU, %**
Naive Bayes	23	45	12 мин	4347	15
Логистическая регрессия	45	67	28 мин	2222	22
Random Forest	156	234	1,2 ч	641	45
XGBoost	127	189	2,1 ч	787	38
SVM	234	356	3,8 ч	427	67
LSTM	289	1240	18 ч	346	82
CNN	145	890	12 ч	689	71
Трансформер	423	2100	48 ч	236	94

Оценка ансамблевых методов демонстрирует значительные улучшения как в метриках точности, так и в характеристиках робастности относительно индивидуальных алгоритмов (Таблица 6). Голосующий классификатор, комбинирующий Random Forest, XGBoost и SVM, достигает точности 97,89 % со сниженной частотой ложных срабатываний 1,2 %, демонстрируя эффективность принятия решений на основе консенсуса в снижении предвзятостей индивидуальных алгоритмов. Стекинг-ансамбль с мета-обучающимся, обученным на предсказаниях базовых алгоритмов, достигает точности 98,34 %, что представляет существенное улучшение по сравнению с лучшим индивидуальным исполнителем. Взвешенный ансамбль с оптимизированными весами на основе специфических сильных сторон алгоритмов для различных типов атак достигает наивысшей точности 98,67 % среди ансамблевых подходов. Багинг-ансамбль

демонстрирует отличные характеристики робастности с согласованной производительностью по различным разбиениям данных и сниженной вариативностью в предсказаниях, что указывает на превосходные возможности обобщения. Бутстрап-агрегирование с базовыми обучающимися Random Forest показывает особенную эффективность в обработке несбалансированных наборов данных, характерных для приложений кибербезопасности, где нормальный трафик значительно превосходит по числу экземпляры атак.

Таблица 6 – Производительность ансамблевых методов
Table 6 – Ensemble methods performance

Ансамблевый метод	Базовые алгоритмы	Точность, %	F1-оценка	Время обработки, мс	Частота ложных срабатываний, %
Голосующий классификатор	RF + XGBoost + SVM	97,89	0,978	198	1,2
Стекинг	RF + XGB + LSTM + CNN	98,34	0,983	267	0,9
Взвешенный ансамбль	Все алгоритмы	98,67	0,986	234	0,8
Багинг	Random Forest	96,78	0,967	167	1,8
AdaBoost	Деревья решений	95,12	0,951	189	2,7
Градиентный бустинг	Слабые обучающиеся	97,23	0,972	156	1,5

Оценка производительности в реальном времени в симулированной производственной среде демонстрирует практическую жизнеспособность решений WAF на основе машинного обучения с приемлемыми характеристиками латентности для большинства случаев использования (Таблица 7). Традиционные алгоритмы поддерживают постоянное время отклика менее 100 мс в условиях высокой нагрузки с возможностями пропускной способности, превышающими 1000 запросов в секунду для легковесных моделей. Архитектуры глубокого обучения показывают увеличенную вариативность латентности под нагрузкой, но поддерживают приемлемую производительность для большинства приложений, где преимущества безопасности перевешивают затраты на латентность. Трансформерные модели, несмотря на наивысшие вычислительные требования, демонстрируют согласованные характеристики производительности с предсказуемыми паттернами использования ресурсов, что позволяет эффективное планирование ресурсов в производственных развертываниях.

Таблица 7 – Производительность в реальном времени под нагрузкой
Table 7 – Real-time performance under load

Модель	Средняя латентность, мс	P95 латентность, мс	Пропускная способность, запр/сек	Использование CPU, %	Память, ГБ
Random Forest	78	156	1282	45	0,23
XGBoost	92	187	1087	52	0,19
LSTM	234	445	427	78	1,24
CNN	156	298	641	67	0,89
Трансформер	378	623	264	89	2,10
Ансамбль	198	367	505	71	1,45

Результаты проведенного исследования демонстрируют революционный потенциал интеграции методов машинного обучения в архитектуру межсетевых экранов веб-приложений, представляя собой парадигматический сдвиг от статических сигнатурных подходов к адаптивным интеллектуальным системам защиты.

Заключение

Экспериментальная валидация четырнадцати различных алгоритмов машинного обучения на масштабном датасете из 2,847,372 HTTP-запросов выявляет превосходство гибридных архитектур, где трансформерные модели достигают точности 98,73 % для детектирования SQL-инъекций, превышая производительность традиционных методов на 23,7 %. Ансамблевые подходы демонстрируют оптимальное соотношение точности и вычислительной эффективности с F1-оценкой 0,989 при сокращении времени обработки в 18 раз относительно rule-based систем. Статистический анализ производительности по категориям атак подтверждает гипотезу о специализации различных алгоритмов для специфических типов угроз, где LSTM-архитектуры показывают 98,7 % эффективность для XSS-атак, а CNN-модели достигают 98,1 % для детектирования переполнения буфера. Комплексная оценка вычислительных характеристик выявляет критические компромиссы между точностью и операционными ограничениями, где традиционные алгоритмы обеспечивают латентность 23–234 мс против 145–423 мс для архитектур глубокого обучения. Анализ важности признаков подтверждает доминирующую роль энтропии символов (важность 0,289) и длины URL (важность 0,267) в дискриминации злонамеренного трафика.

Динамика развития данной области характеризуется экспоненциальным ростом исследовательского интереса к применению продвинутых методов машинного обучения в кибербезопасности, при этом наблюдается устойчивый тренд к интеграции transformer-based архитектур и механизмов внимания в системы реального времени. Эволюция от традиционных эвристических подходов к адаптивным системам на основе искусственного интеллекта представляет фундаментальную трансформацию парадигмы защиты веб-приложений, где способность к обучению и адаптации становится критическим фактором конкурентного преимущества. Прогнозируемое развитие области включает конвергенцию методов федеративного обучения для распределенного совершенствования моделей, интеграцию объяснимого искусственного интеллекта для повышения интерпретируемости решений системы безопасности, и развитие автономных адаптивных архитектур, способных к самостоятельной эволюции в ответ на новые векторы угроз. Стратегическая значимость результатов исследования простирается за рамки технических аспектов, предоставляя основу для формирования политики кибербезопасности и стандартов отрасли, где интеллектуальные адаптивные системы защиты становятся базовым требованием для критической инфраструктуры и коммерческих веб-сервисов.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Román-Gallego J.-A., Pérez-Delgado M.-L., Viñuela M.L., Vega-Hernández M.-C. Artificial Intelligence Web Application Firewall for Advanced Detection of Web Injection Attacks. *Expert Systems.* 2023;42(1). https://doi.org/10.1111/exsy.13505
2. Shaheed A., Kurdy M.H.D.B. Web Application Firewall Using Machine Learning and Features Engineering. *Security and Communication Networks.* 2022;2022. https://doi.org/10.1155/2022/5280158

3. Dawadi B.R., Adhikari B., Srivastava D.K. Deep Learning Technique-Enabled Web Application Firewall for the Detection of Web Attacks. *Sensors.* 2023;23(4). https://doi.org/10.3390/s23042073
4. Vartouni A.M., Teshnehlab M., Kashi S.S. Leveraging Deep Neural Networks for Anomaly-Based Web Application Firewall. *IET Information Security.* 2019;13(4). https://doi.org/10.1049/iet-ifs.2018.5404
5. Hartono B., Silalahi F.D., Muthohir M. Transformers in Cybersecurity: Advancing Threat Detection and Response Through Machine Learning Architectures. *Journal of Technology Informatics and Engineering.* 2024;3(3):382–396. https://doi.org/10.51903/jtie.v3i3.211
6. Avci C., Tekinerdogan B., Catal C. Design Tactics for Tailoring Transformer Architectures to Cybersecurity Challenges. *Cluster Computing.* 2024;27:9587–9613. https://doi.org/10.1007/s10586-024-04355-0
7. Junior M.D., Ebecken N.F.F. A New WAF Architecture with Machine Learning for Resource-Efficient Use. *Computers & Security.* 2021;106. https://doi.org/10.1016/j.cose.2021.102290
8. Applebaum S., Gaber T., Ahmed A. Signature-Based and Machine-Learning-Based Web Application Firewalls: A Short Survey. *Procedia Computer Science.* 2021;189:359–367. https://doi.org/10.1016/j.procs.2021.05.105
9. Belavagi M.C., Muniyal B. Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection. *Procedia Computer Science.* 2016;89:117–123. https://doi.org/10.1016/j.procs.2016.06.016
10. Urda D., Martínez B., Basurto N., Kull M., Arroyo Á., Herrero Á. Enhancing Web Traffic Attacks Identification Through Ensemble Methods and Feature Selection. arXiv. URL: https://arxiv.org/abs/2412.16791 [Accessed 15th July 2025].
11. Franklin J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *The Mathematical Intelligencer.* 2005;27:83–85. https://doi.org/10.1007/BF02985802
12. Sukumar J.V.A., Pranav I., Neetish M.M., Narayanan J. Network Intrusion Detection Using Improved Genetic k-Means Algorithm. In: *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 19–22 September 2018, Bangalore, India.* IEEE; 2018. P. 2441–2446. https://doi.org/10.1109/ICACCI.2018.8554710
13. Vaswani A., Shazeer N., Parmar N., et al. Attention Is All You Need. arXiv. URL: https://arxiv.org/abs/1706.03762 [Accessed 15th July 2025].
14. Tavallaee M., Bagheri E., Lu W., Ghorbani A.A. A Detailed Analysis of the KDD CUP 99 Data Set. In: *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, 08–10 July 2009, Ottawa, ON, Canada.* IEEE; 2009. P. 1–6. https://doi.org/10.1109/CISDA.2009.5356528
15. Shiravi A., Shiravi H., Tavallaee M., Ghorbani A.A. Toward Developing a Systematic Approach to Generate Benchmark Datasets for Intrusion Detection. *Computers & Security.* 2012;31(3):357–374. https://doi.org/10.1016/j.cose.2011.12.012

ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

Ледовская Екатерина Валерьевна, кандидат технических наук, доцент кафедры прикладной математики, МИРЭА – Российский технологический университет, Москва, Российская Федерация.
e-mail: ekvaled@mail.ru

Ekaterina V. Ledovskaya, Candidate of Engineering Sciences, Associate Professor at the Applied Mathematics Department, MIREA – Russian Technological University, Moscow, the Russian Federation.

Статья поступила в редакцию 29.08.2025; одобрена после рецензирования 03.10.2025; принята к публикации 16.10.2025.

The article was submitted 29.08.2025; approved after reviewing 03.10.2025; accepted for publication 16.10.2025.