Трансформерная модель бинарной классификации временных рядов на данных инерциальных датчиков для детекции спуфинг-атак в БПЛА

В. И. Петренко, М. Х. Наджаджра, Ф. Б. Тебуева, Д. Г. Волошин, Н. Дибров

Аннотация*—В статье предложена новая модель бинарной классификации временных рядов BTSC-ISD-Transformer (Binary Time-Series Classification with Inertial Sensor Data Transformer), предназначенная для детекции спуфинг-атак на основе данных инерциальных датчиков (акселерометра и гироскопа). Модель адаптирует архитектуру трансформера для анализа временных рядов, используя механизм самовнимания для параллельного выявления сложных и протяженных во времени аномалий, в отличие от последовательной обработки в традиционных рекуррентных сетях. Проведенные эксперименты демонстрируют превосходство предложенного подхода по сравнению с моделью-аналогом на основе LSTM-RNN. Показатель точности (Accuracy) достиг 97,45%, что на 12% выше результата LSTM-RNN. Метрики F1-мера, Precision и Recall составили 96,41%, 97,03% и 95,79% соответственно, что свидетельствует о высокой сбалансированности модели, ее способности минимизировать как ложные срабатывания, так и пропуски атак. Результаты подтверждают перспективность использования трансформерных моделей в системах реального времени для обеспечения кибербезопасности БПЛА.

Ключевые слова*—беспилотные летательные аппараты (БПЛА), спуфинг-атаки, кибербезопасность, инерциальные датчики, временные ряды, бинарная классификация, трансформер, модель глубокого обучения, самовнимание, LSTM.

I. Введение

Защита беспилотных летательных аппаратов (БПЛА) от спуфинг-атак становится критически важной ввиду широкого распространения дронов в промышленности, обороне и сельском хозяйстве. Спуфинг представляет собой кибератаку, при которой передаются ложные навигационные сигналы, имитирующие сигналы глобальных навигационных спутниковых систем (GPS, ГЛОНАСС), что ведёт к искажению положения аппарата, нарушению его траектории полета и потере управления. В условиях растущей зависимости БПЛА от спутниковой навигации разработка своевременных и надёжных методов обнаружения таких атак является первоочередной задачей обеспечения безопасности.

Современные средства защиты базируются на многоуровневых системах сбора и анализа данных, включая обработку временных рядов с использованием передовых моделей машинного обучения. Трансформерные модели, изначально разработанные для обработки естественного языка, доказали высокую эффективность в анализе последовательных данных благодаря своей способности выявлять зависимости различной дальности и контекста. Их применение в кибербезопасности беспилотников существенно повышает качество детекции, учитывая сложные и динамичные изменения в сенсорных данных.

В настоящем исследовании представлена трансформерная модель бинарной классификации временных рядов, сформированных на основе данных инерциальных датчиков БПЛА. Такой подход позволяет обнаруживать признаки начала спуфинг-атаки в реальном времени на основе внутренней физической информации о движении аппарата, снижая зависимость от внешних источников. Предложенная модель основывается на современных методах глубокого обучения и демонстрирует потенциал для работы в различных условиях.

Отличительной чертой разработки является адаптация архитектуры трансформера к специфике данных инерциальных измерительных устройств, что обеспечивает эффективное выделение характерных паттернов спуфинг-атак. Особое внимание уделено минимизации числа ложных срабатываний, что имеет критическое значение для обеспечения безопасности и устойчивости эксплуатации БПЛА.

II. Анализ литературы

Критический анализ существующей литературы выявляет разнообразие подходов к проблемам детектирования спуфинг-атак беспилотных авиационных систем (БПЛА). Публикации [1, 2] представляют собой ключевые работы, определяющие базовые концепции проблемы. Первая работа посвящена исследованию уязвимости оборудования глобальных навигационных спутниковых систем (ГНСС), тогда как вторая публикация фокусируется на систематизации типов спуфинг-атак. Однако обе работы сосредоточены

В. И. Петренко, канд. техн. наук, доцент, заведующий кафедрой организации и технологии защиты информации, ФГАОУ ВО «Северо-Кавказский федеральный университет», г. Ставрополь, vipetrenko@ncfu.ru, http://orcid.org/0000-0003-4293-7013

М. Х. Наджаджра, PhD, ассоциированный профессор, Университет Аль-Истикляль, г. Иерихон, Палестина, mnajajra@pass.ps

Ф. Б. Тебуева, д-р физ.-мат. наук, доцент, профессор кафедры вычислительной математики и кибернетики, ФГАОУ ВО «Северо-Кавказский федеральный университет», г. Ставрополь, fariza.teb@gmail.com, ORCID: http://orcid.org/0000-0002-7373-4692

Д. Г. Волошин, аспирант, ФГАОУ ВО «Северо-Кавказский федеральный университет», г. Ставрополь, ultrageron@gmail.com

Н. Дибров, студент, ФГАОУ ВО «Северо-Кавказский федеральный университет», г. Ставрополь, dibrovnik@yandex.ru, http://orcid.org/0009-0006-4529-9634

исключительно на формировании теоретического фундамента и не затрагивают практические методы выявления угроз.

В настоящей работе рассмотрен ряд исследований, непосредственно посвящённых обработке сенсорных данных. В частности, в исследовании [3] описан алгоритм детекции спуфинга, базирующийся на объединении данных с различных сенсоров с применением архитектуры TimesNet. Работа [4] предлагает альтернативный метод прогнозного выявления аномалий в траекториях движения БПЛА. Кроме того, публикации [5, 6] охватывают широкий спектр методов машинного обучения, используемых для повышения безопасности беспилотных летательных аппаратов и выявления отклонений от нормальных режимов работы.

Ряд публикаций [7-11] рассматривают вспомогательные аспекты безопасности. В частности, исследования [7, 10] концентрируются на физических способах идентификации дронов, в статье [8] классифицируются сетевые атаки, тогда как в статье [9] демонстрируются возможности графа нейронных сетей для обнаружения аномалий. Статья [11] рассматривает проблему построения оптимальных маршрутов полетов в условиях наличия угрозы.

Статьи [12, 13] посвящены методам компьютерного зрения и обработки изображений, соответственно, глубоким свёрточным нейронным сетям (CNN) и сегментации изображений с использованием ансамблей нейронных сетей. Они решают важные задачи распознавания и классификации визуальной информации, что полезно для мониторинга воздушного пространства и предотвращения столкновений.

Особое внимание привлекают наиболее значимые работы [14-20], изучающие современные методики обработки данных. Среди них выделяются исследования, использующие преобразователи (transformer-based models) для выявления аномалий [14]. Другие работы [15] исследуют особенности навигации при отсутствии сигнала GPS. Наиболее близкие к заявленной тематике статьи [16-20] демонстрируют применение различных архитектур искусственных нейронных сетей, включая интерпретируемый искусственный интеллект (ИИ) [16], гибридные сети типа CNN+Transformer [18] и рекуррентные нейронные сети [20]. В работе [20] рассматривается задача бинарной классификации вторжений в сети БПЛА с использованием архитектуры LSTM-RNN, что служит эталоном для сравнения. Данный подход интересен представленными показателями качества и возможностью улучшения результатов посредством замены LSTM на более современную трансформерную архитектуру.

Цель настоящего исследования – разработка трансформерной модели бинарной классификации временных рядов инерциальных данных, способной повысить точность на 3-5% и F1-меру на 4-6% по сравнению с LSTM-моделями, одновременно сокращая время инференса на 10–15%.

III. Математическая постановка задачи детекции спуфинг-атак в БПЛА

Задача детекции спуфинг-атак в БПЛА формулируется как задача бинарной классификации временных рядов, полученных с инерциальных сенсоров аппарата, с целью определения факта наличия либо отсутствия атаки, воздействующей на навигационную систему.

Для математической постановки задачи детекции спуфинг-атак в БПЛА вводятся следующие обозначения и переменные:

$X = \{x_t\}_{t=1}^{T}$ – временной ряд, где $x_t \in \mathbb{R}^d$ – вектор данных инерциальных датчиков (акселерометров, гироскопов, магнитометров и др.) в момент времени t, $\mathbb{R}^d$ – d-мерное пространство действительных чисел, T – длина наблюдаемой последовательности;

$\mathbb{X} = \{X\}$ – пространство всех возможных временных рядов таких данных;

$y \in \{0,1\}$ – бинарная метка класса, где «0» означает нормальный, неатакующий режим работы БПЛА, а «1» – наличие спуфинг-атаки;

$f: \mathbb{X} \to \{0,1\}$ – функция классификации (детектор), которую необходимо построить.

θ – параметры модели, подлежащие обучению на тренировочной выборке;

$\{(X_i, y_i)\}_{i=1}^{N}$ – тренировочная выборка из N примеров временных рядов с соответствующими метками;

$L(f_\theta(X_i), y_i)$ – функция потерь, например бинарная кросс-энтропия, измеряющая расхождение предсказаний и истинных меток.

Функция обучения сводится к минимизации средней функции потерь:

$$\min_\theta \frac{1}{N} \sum_{i=1}^{N} L\left(f_\theta(X_i), y_i\right). \quad (1)$$

Задача сводится к построению модели f_θ, способной с максимальной точностью классифицировать временные ряды инерциальных данных на нормальные и подвергшиеся воздействию спуфинг-атак.

Для оценки качества модели используют метрики:

1) общая точность детектора или доля правильных решений

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (2)$$

где TP – количество истинно-положительных срабатываний детектора, TN – количество истинно-отрицательных срабатываний детектора, FP – количество ложноположительных срабатываний детектора, FN – количество ложноотрицательных срабатываний детектора;

2) точность обнаружения атаки, характеризующая насколько из всех срабатываний детектора именно атаки были распознаны верно

$$Precision = \frac{TP}{TP+FP}; \quad (3)$$

3) вероятность обнаружения атаки, отражающая способность модели находить все случаи атак, то есть долю правильно обнаруженных атак относительно их общего количества

$$Recall = \frac{TP}{TP+FN}; \quad (4)$$

4) гармоническое среднее точности и полноты (F1-мера)

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (5)$$

Временные ряды могут содержать шумы, искажения и динамические колебания, обусловленные эксплуатационными условиями. Поэтому задача требует применения предобработки и устойчивых алгоритмов выделения признаков, благодаря чему модель адекватно реагирует на характерные изменения, вызванные спуфинг-атаками.

IV. Описание аналогов

Модель обнаружения сетевых вторжений в БПЛА [20], выбранная в настоящем исследовании аналогом, представляет собой комбинацию рекуррентной нейронной сети (Recurrent Neural Network, RNN) с элементами долговременной кратковременной памяти (Long Short-Term Memory, LSTM).

Архитектура модели LSTM-RNN представляет собой последовательную обработку временных рядов данных с использованием рекуррентных слоев с долгосрочной кратковременной памятью. Модель принимает на вход последовательность векторов признаков сетевого трафика или данных инерциальных датчиков. Обработка осуществляется через несколько каскадно соединенных LSTM-слоев, где каждый слой состоит из множества LSTM-ячеек.

Ключевыми компонентами архитектуры являются:

– входной слой для приема временных последовательностей;

– стек LSTM-слоев с механизмами управления памятью (input, forget и output gates);

– полносвязный выходной слой с сигмовидной функцией активации.

Семантическая интерпретация input gate, forget gate и output gate:

– input gate выполняет функцию «рецептора» – решает, какая новая информация заслуживает сохранения;

– forget gate реализует «архивацию» – определяет, какие исторические данные остаются релевантными;

– output gate работает как «публицист» – управляет экспортом информации для последующих вычислений.

Математическая формализация модели LSTM-RNN описывается следующим образом. Пусть задана последовательность векторов признаков сетевого трафика $X = \{x_t\}_{t=1}^{T}$, где $x_t \in \mathbb{R}^d$ представляет собой d-мерный вектор в момент времени t, а T — длина временного горизонта. Пространство всех таких последовательностей обозначается как $\mathbb{X}$. Для каждой последовательности определена бинарная метка $y \in \{0,1\}$, где $y = 1$ соответствует наличию сетевой атаки.

Модель LSTM-RNN реализует отображение $f: \mathbb{X} \to \{0,1\}$, параметризованное вектором параметров θ. Архитектура сети основана на LSTM-ячейках, которые для каждого момента времени t вычисляют:

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \\ \tilde{C}_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \\ h_t &= o_t \odot \tanh(C_t), \end{aligned} \quad (6)$$

где i_t, f_t, o_t – векторы состояний input gate, forget gate и output gate соответственно; $\tilde{C}_t$ – вектор-кандидат состояния ячейки; C_t – обновленное состояние ячейки; h_t – скрытое состояние; W_{xi}, W_{hi}, b_i и аналогично для других gate – обучаемые параметры; σ – сигмовидная функция активации; $\odot$ – поэлементное умножение.

Векторы состояний input gate, forget gate и output gate в LSTM-архитектуре являются ключевыми управляющими механизмами, регулирующими поток информации через ячейку памяти. Их математическая сущность и функциональное назначение раскрываются следующим образом.

Вектор состояния input gate $i_t \in [0,1]^h$ определяет степень усвоения новой информации из входного сигнала x_t. Вычисляется по формуле:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \quad (7)$$

где σ – сигмовидная функция, отображающая значения в диапазон [0,1]. Компоненты вектора i_t интерпретируются как коэффициенты значимости соответствующих компонент вектора-кандидата $\tilde{C}_t$. Значение, близкое к 1, указывает на полное усвоение элемента, близкое к 0 – на игнорирование.

Вектор состояния forget gate $f_t \in [0,1]^h$ управляет сохранением или «забыванием» информации из предыдущего состояния ячейки C_{t-1}:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f). \quad (8)$$

Каждый элемент f_t выступает в роли коэффициента сохранения для соответствующей ячейки памяти. Значение 1 сохраняет информацию, 0 – обнуляет её. Этот механизм обеспечивает контролируемое «забывание» устаревших данных.

Вектор состояния output gate $o_t \in [0,1]^h$ регулирует влияние обновленного состояния ячейки C_t на формирование выходного скрытого состояния h_t:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o). \quad (9)$$

Компоненты o_t определяют степень экспорта информации из ячейки памяти во внешнее скрытое состояние. Фильтрация происходит через поэлементное умножение:

$$h_t = o_t \odot \tanh(C_t). \quad (10)$$

Размерность h этих векторов соответствует количеству нейронов в скрытом слое LSTM. Совместная работа трёх gate позволяет модели адаптивно управлять памятью, сохраняя долгосрочные зависимости и фильтруя шумовые компоненты временных рядов, что критически важно для задач анализа сетевого трафика БПЛА.

Для многослойной архитектуры скрытые состояния каждого слоя l вычисляются как:

$$h_t^{(l)} = \text{LSTM}^{(l)}(h_t^{(l-1)}, h_{t-1}^{(l)}), \quad (11)$$

где $h_t^{(0)} = x_t$, а $l = 1, \dots, L$.

Выход LSTM-сети формируется с помощью полносвязного слоя с сигмовидная функцией активации:

$$\hat{y} = \sigma(W_y h_T^{(L)} + b_y), \quad (12)$$

где $\hat{y} \in [0,1]$ интерпретируется как вероятность принадлежности к классу атак.

Обучение модели заключается в минимизации функции бинарной кросс-энтропии на обучающей выборке $\{(X_i, y_i)\}_{i=1}^{N}$:

$$\min_\theta \frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]. \quad (13)$$

Качество модели оценивается с помощью метрик, определенных в постановке задачи: *Accuracy* (2), *Precision* (3), *Recall* (4) и F1-мера (5).

V. Предлагаемая трансформерная модель бинарной классификации временных рядов на данных инерциальных датчиков для детекции спуфинг-атак в БПЛА

Для решения задачи бинарной классификации временных рядов, формализованной в разделе III, предлагается модель на основе архитектуры трансформер с названием «BTSC-ISD-Transformer» Binary Time-Series Classification with Inertial Sensor Data Transformer). В отличие от последовательной обработки данных в модели LSTM-RNN, предлагаемая модель BTSC-ISD-Transformer обеспечивает параллельный анализ всей временной последовательности за счет механизма самовнимания.

Пусть задана последовательность векторов признаков инерциальных датчиков $X = \{x_t\}_{t=1}^{T}$, где $x_t \in \mathbb{R}^d$, $d = 6$ (три оси акселерометра и три оси гироскопа). Как и в постановке задачи, требуется построить функцию классификации $f: \mathbb{X} \to \{0,1\}$, где $y = 1$ соответствует наличию спуфинг-атаки.

Архитектура предложенной модели включает следующие компоненты:

1. Входное проектирование. В отличие от модели LSTM-RNN, где входные данные подаются последовательно (6), модель BTSC-ISD-Transformer одновременно проецирует все элементы последовательности:

$$z_t = x_t W_e + b_e, \quad W_e \in \mathbb{R}^{d \times d_{model}}, \quad b_e \in \mathbb{R}^{d_{model}}. \quad (14)$$

2. Позиционное кодирование. Для учета временного порядка, который в LSTM моделируется рекуррентными связями (7)-(10), в трансформерной модели добавляется позиционное кодирование:

$$h_t^{(0)} = z_t + p_t, \quad (15)$$

где p_t вычисляется по синусоидальным функциям.

3. Механизм самовнимания. В отличие от gate-механизмов модели LSTM-RNN (7)-(9), модель BTSC-ISD-Transformer использует многокомпонентный механизм внимание:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (16)$$

где Q, K, V – матрицы запроса, ключа и значения, получаемые линейными преобразованиями входных данных.

4. Нормализация и позиционно-ориентированная полносвязная Feed-Forward сеть. Каждый слой модели BTSC-ISD-Transformer содержит два ключевых блока, следующих за механизмом внимания:

1) слой нормализации, который стабилизирует активации для каждого примера последовательности, обеспечивая лучшую устойчивость при обработке временных рядов по сравнению с пакетной нормализацией;

2) позиционно-ориентированную полносвязную сеть, которая выполняет независимое нелинейное преобразование признаков для каждой позиции последовательности через два линейных слоя с активацией ReLU.

5. Классификатор. Аналогично модели LSTM-RNN (12), используется сигмовидная функция активации:

$$\hat{y} = \sigma(h_{[CLS]}^{(L)} W_c + b_c), \quad (17)$$

где $h_{[CLS]}^{(L)}$ – выходное представление специального токена классификации.

Обучение модели проводится минимизацией функции бинарной кросс-энтропии (формула (13)) на обучающей выборке $\{(X_i, y_i)\}_{i=1}^{N}$. Для оценки качества модели используются метрики, определенные в разделе III (формулы (2)-(5)).

Ключевые преимущества предложенной архитектуры по сравнению с LSTM-RNN (раздел IV):

1) параллельная обработка временных последовательностей;

2) способность моделировать зависимости любой длины напрямую;

3) более эффективное выделение значимых временных паттернов;

4) устойчивость к проблеме исчезающих градиентов.

VI. Реализация и эксперимент

Для обучения и оценки модели использовались данные с инерциальных датчиков БПЛА – гироскопа и акселерометра, по трём осям каждого из них (6 признаков). Входной датасет [21] содержит два файла: нормальные полёты (norm.csv) и записи спуфинг-атак GPS (spo.csv). Для ускорения обучения использована случайная выборка 10% от исходных данных.

На рисунках 1 и 2 представлены временные ряды акселерометра и гироскопа для нормального полёта и спуфинг-сигналов. Отмечается наличие событий клиппинга в данных спуфинга, что свидетельствует о перегрузках сенсоров и ключевых паттернах для классификатора.

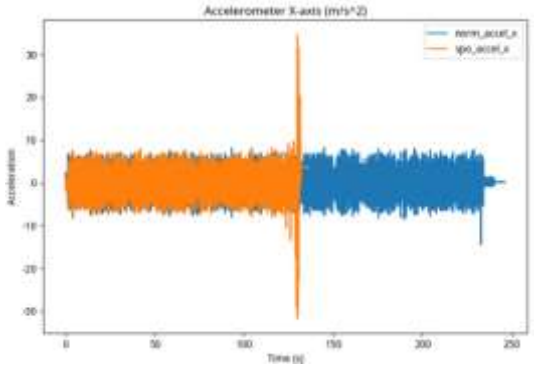

Рис. 1. Временной ряд данных акселерометра для нормального и спуфинг-сигнала

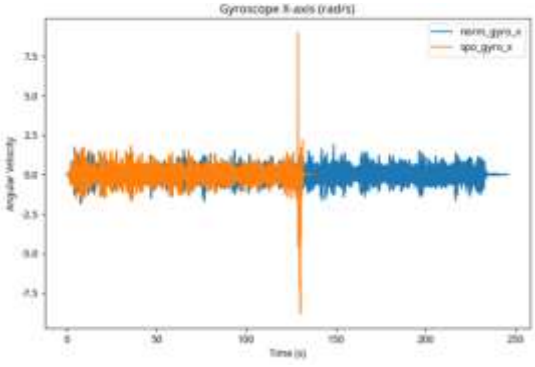

Рис. 2. Временной ряд данных гироскопа для нормального и спуфинг-сигнала

Данные приведены к единому масштабу с использованием стандартного метода предобработки данных (нормализации) в машинном обучении StandardScaler, которая выполняет преобразование по формуле

$$\tilde{\mathbf{x}}_t = \frac{\mathbf{x}_t - \mu}{\sigma},$$

где $\mathbf{x}_t$ – исходное значение признака, μ – среднее значение признака по обучающей выборке, σ – стандартное отклонение признака. Такая стандартизация обеспечивает центральное выравнивание данных со средним 0 и единичной дисперсией.

Для формирования обучающих примеров исходные временные ряды данных инерциальных датчиков преобразовывались с помощью метода скользящего окна. Каждое окно длиной в 10 отсчетов (временных шагов) формирует один обучающий пример (sample), где первые 9 отсчетов служат признаками (фичами), а последний (10-й) отсчет содержит метку класса (нормальный режим / спуфинг-атака). Этот подход соответствует формуле (1) из раздела IV и позволяет учитывать локальные временные зависимости. Полученный набор примеров разбит на обучающую (60%), валидационную (20%) и тестовую (20%) выборки с сохранением баланса классов.»

Архитектура трансформерной модели BTSC-ISD-Transformer построена на 3 слоях с размерностью скрытого пространства $d_{model} = 128$, восьмью компонентами механизма самовнимания и размером внутреннего полносвязного слоя 512, что соответствует формулам (14)-(17). Код модели доступен в репозитории [22]. Для повышения устойчивости использованы пакетная нормализация, регуляризация методом отключения нейронов (dropout) и градиентное отсечение с порогом 0,5. Входная последовательность дополнена специальным токеном классификации [CLS], а для учёта порядка элементов применяется позиционное кодирование. Для предотвращения переобучения применён механизм ранней остановки с параметром patience=20 и адаптивным снижением скорости обучения.

Для сравнительного анализа реализована модель LSTM-RNN с двумя слоями по 128 нейронов в скрытом состоянии. Код реализации доступен в [22]. Классификация производилась по последнему временному шагу последовательности. Обучение модели проводилось методом минимизации функции бинарной кросс-энтропии (13) с использованием алгоритма оптимизации Adam [23] и скоростью обучения 1×10^{-4}. Дополнительные методы регуляризации не применялись.

Рисунки 3, 4 демонстрирует динамику функции потерь и динамику точности на обучающей и валидационной выборках при обучении моделей LSTM-RNN и BTSC-ISD-Transformer.

В процесс обучения модели LSTM-RNN (рис. 2) выявлена менее стабильная сходимость и признаки начала переобучения. Процесс обучения модели BTSC-ISD-Transformer (рис. 3) наблюдается более быстрая сходимость и стабильность на протяжении всего процесса обучения. В таблице 1 представлены рассчитанные ключевые метрики качества моделей LSTM-RNN и BTSC-ISD-Transformer, рассчитанные по формулам (2)-(5).

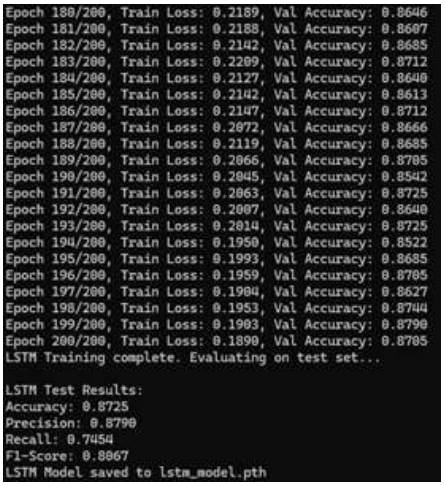

Рис. 3. Процесс обучения модели LSTM-RNN [22]

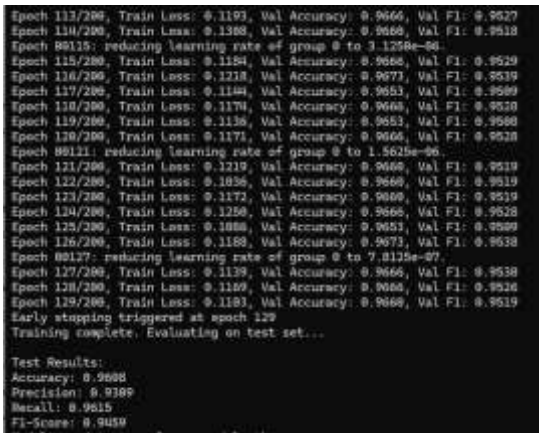

Рис. 4. Процесс обучения модели BTSC-ISD-Transformer [22]

Таблица 1 – Ключевые метрики качества моделей LSTM-RNN и BTSC-ISD-Transformer

Модели	*Accuracy*, %	*Precision*, %	*Recall*, %	F1-мера, %
LSTM-RNN	85,35	87,10	69,23	77,14
BTSC-ISD-Transformer	97,45	97,03	95,79	96,41

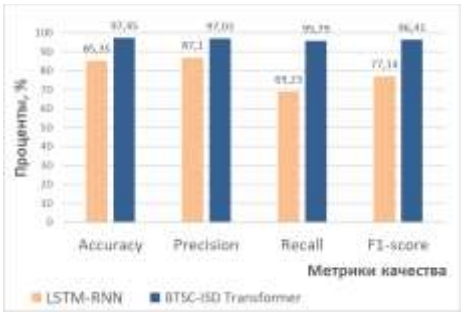

Рис. 5. Сравнение ключевых метрик качества моделей LSTM-RNN и BTSC-ISD-Transformer

Значения ключевых метрик из таблицы 1 и рисунка 5 свидетельствуют о высокой эффективности модели BTSC-ISD-Transformer по сравнению с классической LSTM-сетью.

Показатель точности (*Accuracy*) трансформера достигает 97,45%, что на 12% выше результата LSTM-RNN (85,35%). Это указывает на улучшение общего уровня правильной классификации.

Точность распознавания положительного класса (*Precision*) у модели BTSC-ISD-Transformer составляет 97,03%, превышая аналогичный показатель LSTM-RNN (87,10%), что демонстрирует уменьшение числа ложных срабатываний.

Чувствительность (*Recall*), критически важная для задач обнаружения спуфинг-атак, у модели BTSC-ISD-Transformer равна 95,79%, что значительно выше 69,23%, чем у модели LSTM-RNN. Это свидетельствует о существенном снижении количества пропущенных атак.

Гармоническое среднее *Precision* и *Recall* (F1-мера) у модели BTSC-ISD-Transformer достигает 96,41%, в то время как у модели LSTM-RNN этот показатель составляет 77,14%, что отражает лучшую сбалансированность и надёжность модели.

На рисунке 6 представлена динамика значений функции потерь на обучающей выборке для моделей LSTM-RNN и BTSC-ISD-Transformer. Функция потерь вычисляется по формуле бинарной кросс-энтропии (13).

Рис. 6. Распределение значений функции потерь на обучающей выборке (train loss) для моделей LSTM-RNN и BTSC-ISD-Transformer

Из рисунка 6 видно, что модель BTSC-ISD-Transformer обеспечивает более быстрое снижение значения функции потерь и лучшую стабильность в процессе обучения по сравнению с моделью LSTM-RNN. Это свидетельствует о более эффективной оптимизации параметров модели и лучшей способности к обобщению. Более низкое значение функции потерь указывает на более точное приближение модели к правильным ответам и, как следствие, на улучшение классификационных показателей.

VII. Заключение

В статье предложена трансформерная модель бинарной классификации временных рядов, получаемых с инерциальных датчиков БПЛА, для детекции спуфинг-атак. Ключевым преимуществом предложенной модели BTSC-ISD-Transformer является её способность к параллельному анализу всей временной последовательности благодаря механизму самовнимания, который позволяет эффективно выявлять взаимосвязи между элементами последовательности, независимо от их расстояния во времени.

В отличие от модели-аналога LSTM-RNN, которая обрабатывает данные последовательно и может «забывать» критически важные ранние паттерны из-за проблем с долгосрочными зависимостями, предложенная модель BTSC-ISD-Transformer напрямую моделирует взаимосвязи между любыми точками временного ряда. Это позволило более эффективно выделять сложные и протяженные во времени аномалии, характерные для спуфинг-атак, такие как события клиппинга, отмеченные на рисунке 4.

Результаты экспериментов подтвердили высокую эффективность предложенного подхода в выявлении атак с минимальным числом ложных срабатываний и пропусков, что особенно важно в системах обеспечения безопасности беспилотных платформ. Использование небольшого набора инерциальных сенсоров позволяет оптимизировать вычислительные ресурсы при сохранении качества детекции.

Полученные результаты имеют ряд ограничений. Модель тестировалась на данных с одного датасета и для одного типа спуфинг-атаки. Ее обобщающая способность на другие типы БПЛА и сценарии атак требует дальнейшего исследования. Кроме того, вычислительная сложность трансформерной архитектуры может стать ограничением для развертывания на бортовых системах с жесткими требованиями к задержкам, что также является предметом будущих работ

В качестве направлений для будущих исследований можно выделить тестирование модели на более обширных и разнообразных наборах данных, включающих различные типы БПЛА и сценарии атак, а также оптимизацию вычислительной сложности модели для ее развертывания на бортовом оборудовании с ограниченными ресурсами.

Библиография

[1] Неровный В.В., Коратаев П.Д., Облов П.С., Толстых М.Ю. Характеристики уязвимости аппаратуры потребителей глобальных навигационных спутниковых систем к спуфинг-атакам // Труды учебных заведений связи. 2023. Т. 9. № 6. С. 95-100. DOI: 10.31854/1813-324X-2023-9-6-95-100

[2] Воловик Е.В. Типы имитационного воздействия (спуфинга) // Труды ГосНИИАС. Серия Авиационные системы. 2024. № 3 (66). С. 19-31.

[3] Wang J., Nie L., Gu Z., Wang J., Tan R. and Kumari S. Real-Time GPS Spoofing Detection in Consumer Drones Through Multi-Sensor Data Fusion and TimesNet // IEEE Transactions on Consumer Electronics. 2025. Vol. 71. № 2, pp. 5569-5583, doi: 10.1109/TCE.2025.3560264.

[4] Huang T., Wu H., Tao X., Wei Z. Prediction-based trajectory anomaly detection in UAV system with GPS spoofing attack // Chinese Journal of Aeronautics. 2025. Vol. 38, Issue 10. Article № 103478; https://doi.org/10.1016/j.cja.2025.103478.

[5] Yang Z., Zhang Y., Zeng J., Yang Y., Jia Y., Song H., Lv T., Sun Q., An J. AI-Driven Safety and Security for UAVs: From Machine Learning to Large Language Models // Drones. 2025. Vol. 9(6). Article №392; https://doi.org/10.3390/drones9060392

[6] Zhen Li, Kamarudin N.H., Kok V.J., Qamar F. Anomaly Detection Model in Network Security Situational Awareness Based on Machine Learning: Limitation, Techniques, and Future Trends // IEEE Access. 2025. Vol.13. Pp.126084-126129; DOI: 10.1109/ACCESS.2025.3589620.

[7] Аверина М.Д., Леванова О.А., Грушевская Д.В., Кухарев К.А., Мурин Д.М., Калинин М.А. Детекция БПЛА при помощи нейронных сетей // Моделирование и анализ информационных систем. 2024. Т. 31. № 2. С. 182-193. DOI: 10.18255/1818-1015-2024-2-182-193.

[8] Балыбердин А.В. Мультимодальный метод извлечения признаков данных для классификации сетевых атак // Известия ЮФУ. Технические науки. 2025. № 3 (245). С. 6-16. https://doi.org/10.18522/2311-3103-2025-3-6-16.

[9] Feng C., Fan J., Liu Z., Jin G., Chen S. Unmanned Aerial Vehicle Anomaly Detection Based on Causality-Enhanced Graph Neural Networks // Drones. 2025. Vol. 9(6). Article № 408; https://doi.org/10.3390/drones9060408.

[10] Semenyuk V., Kurmashev I., Lupidi A., Alyoshin D., Kurmasheva L., Cantelli-Forti A. Advances in UAV detection: integrating multi-sensor systems and AI for enhanced accuracy and efficiency // International Journal of Critical Infrastructure Protection. 2025. Vol. 49. Article №100744; https://doi.org/10.1016/j.ijcip.2025.100744.

[11] Петренко В.И., Тебуева Ф.Б., Волошин Д.Г. Метод построения траектории полета беспилотных летательных аппаратов в условиях кибератак GPS/ГЛОНАСС спуфинга // Прикаспийский журнал: управление и высокие технологии. 2024. № 2 (66). С. 71-80.

[12] Корчагин В.Д. Анализ современных SOTA-архитектур искусственных нейронных сетей для решения задач классификации изображений и детекции объектов // Программные системы и вычислительные методы. 2023. № 4. С. 73-87. DOI: 10.7256/2454-0714.2023.4.69306.

[13] Котенко И.В., Саенко И.Б., Лаута О.С., Васильев Н.А., Садовников В.Е. Метод противодействия состязательным атакам на системы классификации изображений // Вопросы кибербезопасности. 2025. № 2 (66). С. 114-123. DOI: 10.21681/2311-3456-2025-2-114-123.

[14] Dilek E., Dener M. An overview of transformers for video anomaly detection // Neural Computing and Applications. 2025. Vol. 37. Pp. 17825-17857; https://doi.org/10.1007/s00521-025-11218-1.

[15] Jarraya I., Al-Batati A., Kadri M.B., Abdelkader M., Ammar A., Boulila W., Koubaa A. Gnss-denied unmanned aerial vehicle navigation: analyzing computational complexity, sensor fusion, and localization methodologies // Satellite Navigation. 2025. Vol. 6. №9. https://doi.org/10.1186/s43020-025-00162-z.

[16] Devkota B.P., Kandel L.N. Applying Deep Learning Approach for GPS Spoofing Detection of UAV with Explainable AI // 44th Digital Avionics Systems Conference. 2025. Montreal, Canada. September 14-18. URL: https://www.researchgate.net/publication/395112491_Applying_Deep_Learning_Approach_for_GPS_Spoofing_Detection_of_UAV_with_Explainable_AI.

[17] Korium M.S., Saber M., Ahmed A.M., Narayanan A., Nardelli P.H.J. Image-based intrusion detection system for GPS spoofing cyberattacks in unmanned aerial vehicles // Ad Hoc Networks. 2024. Vol. 163. Article № 103597 https://doi.org/10.1016/j.adhoc.2024.103597.

[18] Han Y., Jia Z., He S., Zhang Y., Wu Q. CNN+Transformer Based Anomaly Traffic Detection in UAV Networks for Emergency Rescue // https://arxiv.org/html/2503.20355v1.

[19] Ahmad M.W., Akram M.U., Mohsan M.M., Kashif Saghar, Ahmad R., Butt W.H. Transformer-based sensor failure prediction and classification framework for UAVs // Expert Systems with Applications. 2024. Volume 248, article 123415. https://doi.org/10.1016/j.eswa.2024.123415.

[20] Gamal M., Elhamahmy M., Taha S., Elmahdy H. Improving intrusion detection using LSTM-RNN to protect drones' networks // Egyptian Informatics Journal. 2024. Vol. 27. Article №100501; https://doi.org/10.1016/j.eij.2024.100501.

[21] University of New Brunswick. CIC IoT Dataset 2023 [Электронный ресурс]. URL: https://www.unb.ca/cic/datasets/iotdataset-2023.html (дата обращения: 08.10.2025).

[22] Волошин Д.Г., Дибров Н. Репозиторий с исходным кодом модели LSTM и трансформера: lstm-transformer-compare [Электронный ресурс]. URL: https://github.com/dibrovnik/lstm-transformer-compare (дата обращения: 08.10.2025).

[23] Kingma D.P., Ba J. Adam: A Method for Stochastic Optimization // arXiv:1412.6980v9; https://doi.org/10.48550/arXiv.1412.6980

Transformer-based binary classification model for time series using inertial sensor data for spoofing attack detection in UAVs

V. I. Petrenko, M. Kh. Nadzhadzhra, F. B. Tebueva, D. G. Voloshin, N. Dibrov

***Abstract*—This paper proposes a novel binary time-series classification model, BTSC-ISD-Transformer (Binary Time-Series Classification with Inertial Sensor Data Transformer), designed for spoofing attack detection based on inertial sensor data (accelerometer and gyroscope). The model adapts the transformer architecture for time-series analysis, leveraging the self-attention mechanism to enable parallel detection of complex and long-term anomalies, in contrast to the sequential processing employed by traditional recurrent neural networks. Experimental results demonstrate the superiority of the proposed approach compared to an LSTM-RNN-based baseline model. The classification accuracy reached 97.45%, which is 12% higher than that of the LSTM-RNN model. The F1-score, Precision, and Recall achieved 96.41%, 97.03%, and 95.79%, respectively, indicating a high level of model balance and its ability to minimize both false positives and missed attacks. The results confirm the strong potential of transformer-based models for real-time cybersecurity systems in UAV applications.**

Keywords—unmanned aerial vehicles (UAVs), spoofing attacks, cybersecurity, inertial sensors, time series, binary classification, transformer, deep learning model, self-attention, LSTM.

REFERENCES

[1] Неровный V. V., Korataev P. D., Oblov P. S., Tolstykh M. Yu. Vulnerability characteristics of global navigation satellite system consumer equipment to spoofing attacks. *Proceedings of Educational Institutions of Communications*, 2023, vol. 9, no. 6, pp. 95–100. DOI: 10.31854/1813-324X-2023-9-6-95-100.

[2] Volovik E. V. Types of spoofing (imitation interference). *Proceedings of GosNIIAS. Aviation Systems Series*, 2024, no. 3 (66), pp. 19–31.

[3] Wang J., Nie L., Gu Z., Wang J., Tan R., Kumari S. Real-time GPS spoofing detection in consumer drones through multi-sensor data fusion and TimesNet. *IEEE Transactions on Consumer Electronics*, 2025, vol. 71, no. 2, pp. 5569–5583. DOI: 10.1109/TCE.2025.3560264.

[4] Huang T., Wu H., Tao X., Wei Z. Prediction-based trajectory anomaly detection in UAV systems under GPS spoofing attacks. *Chinese Journal of Aeronautics*, 2025, vol. 38, issue 10, article no. 103478. https://doi.org/10.1016/j.cja.2025.103478.

[5] Yang Z., Zhang Y., Zeng J., Yang Y., Jia Y., Song H., Lv T., Sun Q., An J. AI-driven safety and security for UAVs: from machine learning to large language models. *Drones*, 2025, vol. 9, no. 6, article no. 392. https://doi.org/10.3390/drones9060392.

[6] Li Z., Kamarudin N. H., Kok V. J., Qamar F. Anomaly detection models in network security situational awareness based on machine learning: limitations, techniques, and future trends. *IEEE Access*, 2025, vol. 13, pp. 126084–126129. DOI: 10.1109/ACCESS.2025.3589620.

[7] Averina M. D., Levanova O. A., Grushevskaya D. V., Kukharev K. A., Murin D. M., Kalinin M. A. UAV detection using neural networks. *Modeling and Analysis of Information Systems*, 2024, vol. 31, no. 2, pp. 182–193. DOI: 10.18255/1818-1015-2024-2-182-193.

[8] Balyberdin A. V. Multimodal feature extraction method for network attack classification. *Proceedings of Southern Federal University. Engineering Sciences*, 2025, no. 3 (245), pp. 6–16. https://doi.org/10.18522/2311-3103-2025-3-6-16.

[9] Feng C., Fan J., Liu Z., Jin G., Chen S. Unmanned aerial vehicle anomaly detection based on causality-enhanced graph neural networks. *Drones*, 2025, vol. 9, no. 6, article no. 408. https://doi.org/10.3390/drones9060408.

[10] Semenyuk V., Kurmashev I., Lupidi A., Alyoshin D., Kurmasheva L., Cantelli-Forti A. Advances in UAV detection: integrating multi-sensor systems and AI for enhanced accuracy and efficiency. *International Journal of Critical Infrastructure Protection*, 2025, vol. 49, article no. 100744. https://doi.org/10.1016/j.ijcip.2025.100744.

[11] Petrenko V. I., Tebueva F. B., Voloshin D. G. A method for UAV flight trajectory generation under GPS/GLONASS spoofing cyberattacks. *Caspian Journal: Management and High Technologies*, 2024, no. 2 (66), pp. 71–80.

[12] Korchagin V. D. Analysis of modern state-of-the-art neural network architectures for image classification and object detection. *Software Systems and Computational Methods*, 2023, no. 4, pp. 73–87. DOI: 10.7256/2454-0714.2023.4.69306.

[13] Kotenko I. V., Saenko I. B., Lauta O. S., Vasiliev N. A., Sadovnikov V. E. A method for counteracting adversarial attacks on image classification systems. *Cybersecurity Issues*, 2025, no. 2 (66), pp. 114–123. DOI: 10.21681/2311-3456-2025-2-114-123.

[14] Dilek E., Dener M. An overview of transformers for video anomaly detection. *Neural Computing and Applications*, 2025, vol. 37, pp. 17825–17857. https://doi.org/10.1007/s00521-025-11218-1.

[15] Jarraya I., Al-Batati A., Kadri M. B., Abdelkader M., Ammar A., Boulila W., Koubaa A. GNSS-denied unmanned aerial vehicle navigation: analysis of computational complexity, sensor fusion, and localization methodologies. *Satellite Navigation*, 2025, vol. 6, no. 9. https://doi.org/10.1186/s43020-025-00162-z.

[16] Devkota B. P., Kandel L. N. Applying deep learning approaches for GPS spoofing detection in UAVs with explainable AI. In *Proceedings of the 44th Digital Avionics Systems Conference (DASC)*, Montreal, Canada, Sept. 14–18, 2025. Available: ResearchGate.

[17] Korium M. S., Saber M., Ahmed A. M., Narayanan A., Nardelli P. H. J. Image-based intrusion detection system for GPS spoofing cyberattacks in unmanned aerial vehicles. *Ad Hoc Networks*, 2024, vol. 163, article no. 103597. https://doi.org/10.1016/j.adhoc.2024.103597.

[18] Han Y., Jia Z., He S., Zhang Y., Wu Q. CNN + transformer-based anomaly traffic detection in UAV networks for emergency rescue. *arXiv preprint*, 2025. https://arxiv.org/abs/2503.20355.

[19] Ahmad M. W., Akram M. U., Mohsan M. M., Saghar K., Ahmad R., Butt W. H. Transformer-based sensor failure prediction and classification framework for UAVs. *Expert Systems with Applications*, 2024, vol. 248, article no. 123415. https://doi.org/10.1016/j.eswa.2024.123415.

[20] Gamal M., Elhamahmy M., Taha S., Elmahdy H. Improving intrusion detection using LSTM-RNN to protect drone networks. *Egyptian Informatics Journal*, 2024, vol. 27, article no. 100501. https://doi.org/10.1016/j.eij.2024.100501.

[21] University of New Brunswick. CIC IoT Dataset 2023 [Online]. Available: https://www.unb.ca/cic/datasets/iotdataset-2023.html (accessed: Oct. 8, 2025).

[22] Voloshin D. G., Dibrov N. Source code repository for LSTM and transformer models: *lstm-transformer-compare* [Online]. Available: https://github.com/dibrovnik/lstm-transformer-compare (accessed: Oct. 8, 2025).

[23] Kingma D. P., Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980v9*, 2015. https://doi.org/10.48550/arXiv.1412.6980.