УДК 004.75
DOI: 10.26102/2310-6018/2022.37.2.020

Разработка системы стегоанализа цифровых изображений на основе нейросетевого классификатора

А.А. Минайчев[1]✉, А.О. Мезенцев[2], Э.А. Яндашевская[2]

[1]Московский государственный технический университет им. Н.Э. Баумана, Москва, Российская Федерация
[2]Академия Федеральной службы охраны Российской Федерации, Орёл, Российская Федерация
anton.minaichev@gmail.com✉

Резюме. В статье рассматривается подход к реализации системы стеганографического анализа цифровых изображений на основе нейросетевого классификатора, которая используется в рамках комплексной системы мониторинга событий информационной безопасности корпоративных инфокоммуникационных систем. В качестве базовой структуры нейросетевого классификатора предлагается использование модифицированного варианта сверточной нейронной сети, модуль преобработки которой реализует гистограммный метод анализа цвето-яркостных характеристик цифровых изображений. Для автоматизации процесса обучения нейросетевого классификатора в структуру разрабатываемой системы предлагается ввести модуль массовой генерации стегоконтейнеров с заранее заданными значениями типа и размера цифрового изображения, а также размера полезной нагрузки. На основе разработанной структуры системы стегоанализа цифровых изображений был спланирован и проведен факторный эксперимент по оцениванию качества функционирования предложенного нейросетевого классификатора в сравнении с известными решениями бинарных статистических классификаторов. Особенностью проведенного эксперимента является выбор в качестве метрики оценивания качества классификации площади под кривой ошибок (AUC ROC). Результаты эксперимента продемонстрировали возможность применения нейросетевых классификаторов для решения задач стегоанализа, в частности, применительно к их реализации в перспективных средствах защиты информации.

Ключевые слова: цифровая стеганография, цифровые изображения, сверточная нейронная сеть, бинарная классификация, стеганографический контейнер, точность классификации.

Для цитирования: Минайчев А.А., Мезенцев А.О., Яндашевская Э.А. Разработка системы стегоанализа цифровых изображений на основе нейросетевого классификатора. *Моделирование, оптимизация и информационные технологии.* 2022;10(2). Доступно по: https://moitvivt.ru/ru/journal/pdf?id=1196 DOI: 10.26102/2310-6018/2022.37.2.020

Development of a steganalysis system for digital images based on a neural network classifier

A.A. Minaychev[1]✉, A.O. Mezentsev[2], E.A. Yandashevskaya[2]

[1]Bauman Moscow State Technical University, Moscow, Russian Federation
[2]The Federal Guard Service Academy, Oryol, Russian Federation
anton.minaichev@gmail.com✉

Abstract. The article discusses an approach to the implementation of a system for steganographic analysis of digital images based on a neural network classifier. It is used as a part of an integrated system

for monitoring information security events of corporate infocommunication systems. As a basic structure for the neural network classifier, it is proposed to use a modified version of the convolutional neural network. Its preprocessing module implements the histogram method for analyzing the color and brightness characteristics of digital images. To automate the learning process of the neural network classifier, it is suggested to introduce a module for mass generation of stegocontainers with predefined values for the type and size of a digital image as well as for the size of the payload into the structure of the system being developed. Based on the developed structure of the steganalysis system for digital images, a factorial experiment was planned and conducted to evaluate the quality of the described neural network classifier in comparison with the known solutions of binary statistical classifiers. The choice of the area under the error curve (AUC ROC) as a metric for assessing the quality of classification is the main feature of the experiment. The results show that it is possible to use neural network classifiers to solve steganalysis problems, including their implementation in advanced information security tools.

Keywords: digital steganography, digital images, convolutional neural network, binary classification, steganographic container, classification accuracy.

For citation: Minaichev A.A., Mezentsev A.O., Yandashevskaya E.A. Development of a steganalysis system for digital images based on a neural network classifier. *Modeling, Optimization and Information Technology*. 2022;10(2). Available from: https://moitvivt.ru/ru/journal/pdf?id=1196 DOI: 10.26102/2310-6018/2022.37.2.020 (In Russ.).

Введение

Популярность цифровых изображений (ЦИ) как средств передачи информации активно используется злоумышленниками в качестве инструмента для реализации компьютерных атак с использованием стегоконтейнеров. Это обусловлено следующими причинами:

– большим количеством ЦИ, передаваемых в сети Интернет, которые позволяют добиваться высокой пропускной способности стегоканала;

– избыточностью ЦИ, которая позволяет встроить большой объем информации с сохранением малых значений полезной нагрузки;

– наличием большого разнообразия методов встраивания информации в ЦИ.

Широкое использование ЦИ делает их популярным средством реализации угроз информационной безопасности. В частности, злоумышленники активно используют форматы ЦИ в качестве контейнеров для создания скрытого канала передачи / приема, как похищаемой ими информации, так и вредоносного программного обеспечения (ВПО).

Наиболее очевидным видом подобных атак являются утечка информации различного уровня конфиденциальности, реализуемая внутренним нарушителем. Другой вид атак основан на внедрении в структуру ПО атакуемой инфокоммуникационной системы ВПО путем скрытой передачи его или его компонентов в стегоконтейнере на основе ЦИ.

Раздел стеганографии, связанный с выявлением факта передачи информации в анализируемом сообщении (например, в ЦИ) называется стеганографический анализ (стегоанализ) [1]. Исчерпывающий обзор методов стегоанализа ЦИ представлен в [2]. При этом рассматриваемые подходы к стегоанализу ЦИ носят в основном исследовательский характер. Существующие средства защиты информации (СЗИ) не в полной мере поддерживают средства и методы стегоанализа ЦИ, и расследование подобных инцидентов ИБ производится только по факту успешного завершения атак. Это определяет актуальность такого рода исследований и разработок в плане совершенствования СЗИ. Исходя из этого, целью исследования является определение возможности использования методов машинного обучения с учителем: в частности, статистических и нейросетевых классификаторов для выявления факта угроз

информационной безопасности на основе скрытых каналов, использующих стеганографическое преобразование ЦИ. Объектом исследования является система стегоанализа ЦИ. Предметом исследования являются методы и средства стеганографического преобразования и стегоанализа ЦИ.

Разработка системы стегоанализа цифровых изображений на основе сверточной нейронной сети

Анализ предметной области существующих систем стегоанализа ЦИ, основанных на статистических классификаторах, а также предметной области классификаторов ЦИ на основе сверточных нейронных сетей (СНС) позволил разработать функциональную схему системы стегоанализа ЦИ на основе нейросетевого классификатора (рисунок 1).

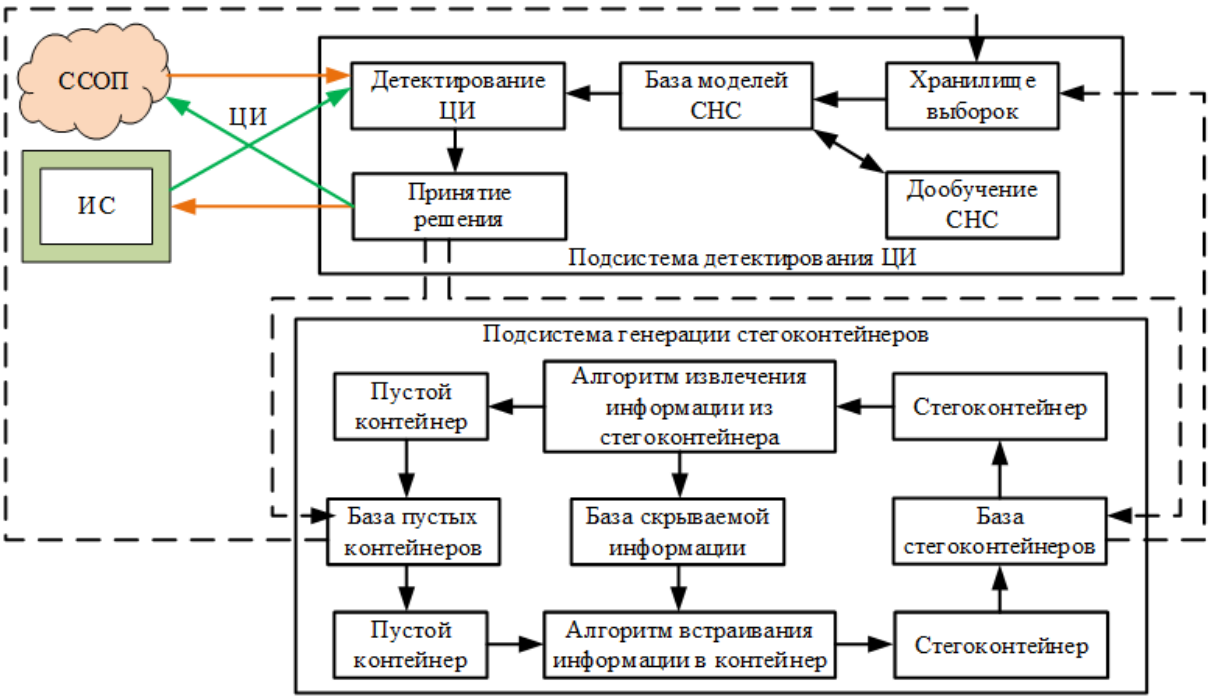

Рисунок 1 – Функциональная схема системы стегоанализа ЦИ
Figure 1 – Functional diagram of the DI stegoanalysis system

Из Рисунка 1 видно, предлагаемая система стегоанализа ЦИ состоит из двух подсистем:

1. Подсистема детектирования ЦИ, обеспечивающая основную функцию системы – детектирование ЦИ и принятия решения о его допуске в / из информационной системы (ИС). Поскольку данная подсистема базируется на классификаторе на основе СНС, то она дополнительно содержит:

– базу моделей СНС – предварительно обученных вариантов СНС различной структуры и параметрической настройки, реализующих классификаторы для различных методов встраивания информации в ЦИ;

– хранилище выборок (обучающей, тестовой и валидационной) для предварительного обучения варианта СНС из базы моделей СНС;

– модуль дообучения СНС, необходимый в случае выявления новых или модификации существующих методов встраивания.

2. Подсистема генерации стегоконтейнеров, необходимых для автоматизации формирования хранилища выборок СНС. Эта подсистема содержит:

– алгоритм автоматизации встраивания информации в ЦИ с использованием выбранных средств встраивания. Результатом его работы является множество ЦИ – база стегоконтейнеров;

– алгоритм извлечения информации из стегоконтейнера, обеспечивающий выборочную проверку качества встраивания.

Также из Рисунка 1 следует, что все ЦИ, участвующие в информационном обмене между ИС и сетью связи общего пользования (ССОП), перехватываются и поступают на вход подсистемы детектирования ЦИ, где производится их детектирование и отнесение к одному из двух классов (контейнер или стегоконтейнер). На основе оценки принадлежности ЦИ к стегоконтейнеру принимается решение о последующей передаче / приеме файла из / в ИС. В случае появления новых видов стегоконтейнеров, а также средств и методов встраивания информации, поступившие и отправляемые в / из ИС ЦИ сохраняются в зависимости от присвоенного класса в базы пустых и стегоконтейнеров, из которых в дальнейшем формируется выборка для дообучения СНС. Модифицированная модель СНС подается для дальнейшего использования в процессе детектирования ЦИ.

Важной исследовательской задачей являлась разработка структуры СНС, реализующей функции нейросетевого классификатора для детектирования ЦИ, а также разработка алгоритма ее функционирования и его программная реализация на основе выбранного фреймворка *PyTorch*. Ниже на Рисунке 2 представлена модифицированная структура СНС для реализации поставленной задачи.

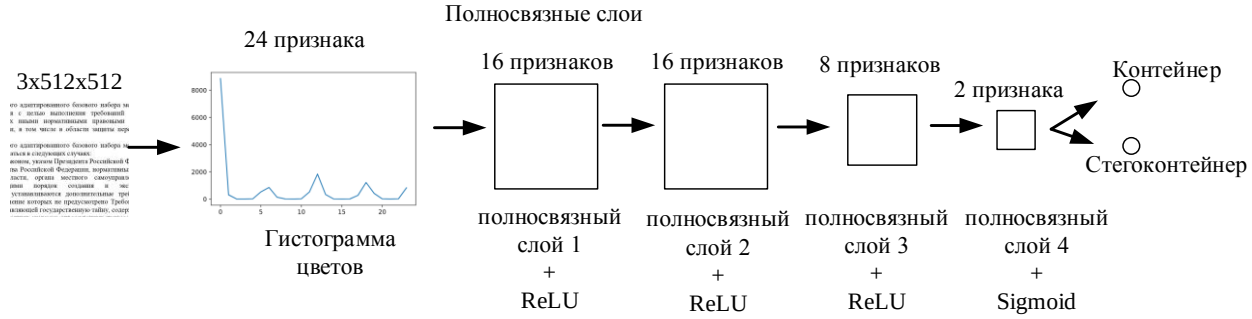

Рисунок 2 – Структурная схема системы стегоанализа ЦИ на основе СНС
Figure 2 – The structural diagram of the DI stegoanalysis system based on the CNN

Разработка этой схемы основана на следующих эмпирических гипотезах:

1. Количество нейронов в слое СНС зависит от объема обучающей выборки.

2. Выбор конкретной функции активации для конкретного слоя СНС зависит от особенностей этой функции применительно к этапу функционирования СНС.

3. С увеличением размера скрытого слоя увеличивается сложность реализующего его кода и время обработки ЦИ.

Из Рисунка 2 видно, что традиционный для СНС этап предобработки ЦИ модифицирован на процедуру получения гистограмм цвето-яркостных характеристик.

Традиционная для СНС предобработка ЦИ применяется с целью повышения показателя точности классификации, поскольку обеспечивает инвариантность СНС к несущественным искажениям или изменениям ЦИ. Обычно к ним относятся различного рода шумы, повороты, операции масштабирования.

Процедура предобработки содержит простые аффинные преобразования, устраняющие такие искажения, как и угловое смещение, поворот и сдвиг. Однако их применение существенно увеличивает время обучения, поскольку, в зависимости от

размера обучающей выборки, происходит формирование значительного количества промежуточных ЦИ.

С целью сокращения этого времени в работе предлагается на этапе предобработки использовать гистограммный метод. Он основан на получении гистограмм для каждого цветового и яркостного каналов ЦИ. Гистограмма при этом представляет собой перепады значений цвета (яркости) в некоторых местах ЦИ. Обычно первой формируется гистограмма яркостного канала по всем X-координатам и Y-координатам, как сумма значений яркости каждого пикселя. Аналогичным образом формируются гистограммы для трех цветовых каналов.

Также были модифицированы традиционные для СНС сверточные слои. В силу того, что при генерации кода нейронной сети в фреймворке *PyTorch* в качестве базового слоя применяется полносвязный слой с заданными характеристиками, то, с целью сокращения трудоемкости получаемого программного кода, было принято решение не создавать отдельную процедуру генерации сверточных слоев с заданными характеристиками, а использовать типовую функцию редукции связей в сгенерированном полносвязном слое, что обеспечивает его трансформацию в вариант сверточного слоя.

Выбор и обоснование метрик оценивания качества нейросетевого стегоанализатора ЦИ

В литературе [2, 6], посвященной теории машинного обучения, в качестве метрик оценивания качества процесса классификации, в основном, рассматриваются метрики *accuracy* (доля правильных ответов), *precision* (точность) и *recall* (полнота).

В работе для оценки качества предложенной модели классификации будем использовать такие метрики, как *accuracy* (доля правильных ответов) – точность классификации, *AUC ROC* (*Area Under Curve Receiver Operator Characteristic*, оценка площади под кривой) [3].

Для перехода к самим метрикам необходимо ввести важную концепцию их описания в терминах ошибок классификации – матрицу ошибок (*confusion matrix*), представленную Таблицей 1.

Таблица 1 – Общий вид матрицы ошибок как результат классификации
Table 1 – General view of the error matrix as a result of classification

		Верная гипотеза	
		$y=1$	$y=0$
Результат работы СНС	$\hat{y}=1$	TP	FP
	$\hat{y}=0$	FN	TN

Результат отнесения СНС поступившего на вход неизвестного класса контейнера может быть одним из четырех ниже перечисленных понятий, которые будут использованы для дальнейшего расчета метрик:

– *True Positive (TP)* – истинно-положительный результат (поступивший на вход стегоконтейнер был принят СНС в качестве стегоконтейнера).

– *False Positive (FP)* – ложно-положительный результат (ошибка 1 рода) (поступивший на вход пустой контейнер был принят СНС в качестве стегоконтейнера).

– *True Negative (TN)* – истинно-отрицательный результат (поступивший на вход пустой контейнер был принят СНС в качестве пустого контейнера).

– *False Negative (FN)* – ложно-отрицательный результат (ошибка 2 рода) (поступивший на вход стегоконтейнер был принят СНС в качестве пустого контейнера).

Соответственно используемые метрики можно представить следующими выражениями:

1. *accuracy* показывает отношение числа правильно классифицированных контейнеров к общему числу поданных на вход модели контейнеров

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. *ROC*-кривая *(Receiver Operator Characteristic)* – инструмент для визуализации и оценки качества моделей обучения, который наиболее часто используется для представления результатов бинарной классификации в машинном обучении. Кривая представляет собой зависимость доли истинно-положительных результатов от доли ложно-положительных результатов классификации при варьировании решающего правила, именуемым порогом.

При анализе *ROC*-кривых оперируют не абсолютными, а относительными величинами *TPR (True Positive Rate)* и *FPR (False Negative Rate)*, представленными выражениями (3.4) и (3.5) соответственно:

$$TPR = \frac{TP}{TP+FN} \quad (2)$$

$$FPR = \frac{FP}{FP+TN} \quad (3)$$

а. Чувствительность *(sensitivity) (True Positive Rate, TPR)* равна доле правильно идентифицированных положительных результатов (вероятности того, что стегоконтейнер будет классифицирован как стегоконтейнер) и выражается формулой:

$$Se = TPR = \frac{TP}{TP+FN} \quad (4)$$

б. Специфичность *(specificity) (True Negative Rate, TNR)* равна доле правильно идентифицированных отрицательных результатов (вероятности того, что контейнер будет классифицирован как контейнер) и выражается формулой:

$$Sp = 1 - FPR = 1 - \frac{FP}{FP+TN} = \frac{TN}{FP+TN} \quad (5)$$

ROC-кривая строится следующим образом:

1. Задается ось абсцисс – вероятность ложно-положительная.

2. Классификации и ось ординат – вероятность истинно-положительных результатов. График, построенный в таких осях позволяет оценить эффективность классификации при варьировании порога классификации.

3. Рассчитываются значения *TPR* и *FPR* для каждого значения порога, которое меняется от 0 до 1 с некоторым шагом.

4. Строится график зависимости двух величин, в результате которого получается *ROC*-кривая.

ROC-кривая для идеального классификатора проходит через левый верхний угол, где доля истинно-положительных случаев составляет 100 %, а доля ложно положительных случаев равна нулю. Поэтому чем ближе полученная кривая к верхнему левому углу, тем больше предсказательная способность модели.

В качестве сравнения алгоритмов с помощью *ROC*-кривых предлагается использовать оценку площади под кривой *AUC ROC* (6).

$$AUC = \int_0^1 f(x)dx \quad (6)$$

В качестве нижней границы можно взять 0.5, как для «бесполезного» классификатора и в качестве верхней 1.0, как для «идеального» классификатора.

Показатель *AUC* предназначен для сравнения моделей, и чем он больше, тем больше предсказательная способность модели (Рисунок 3).

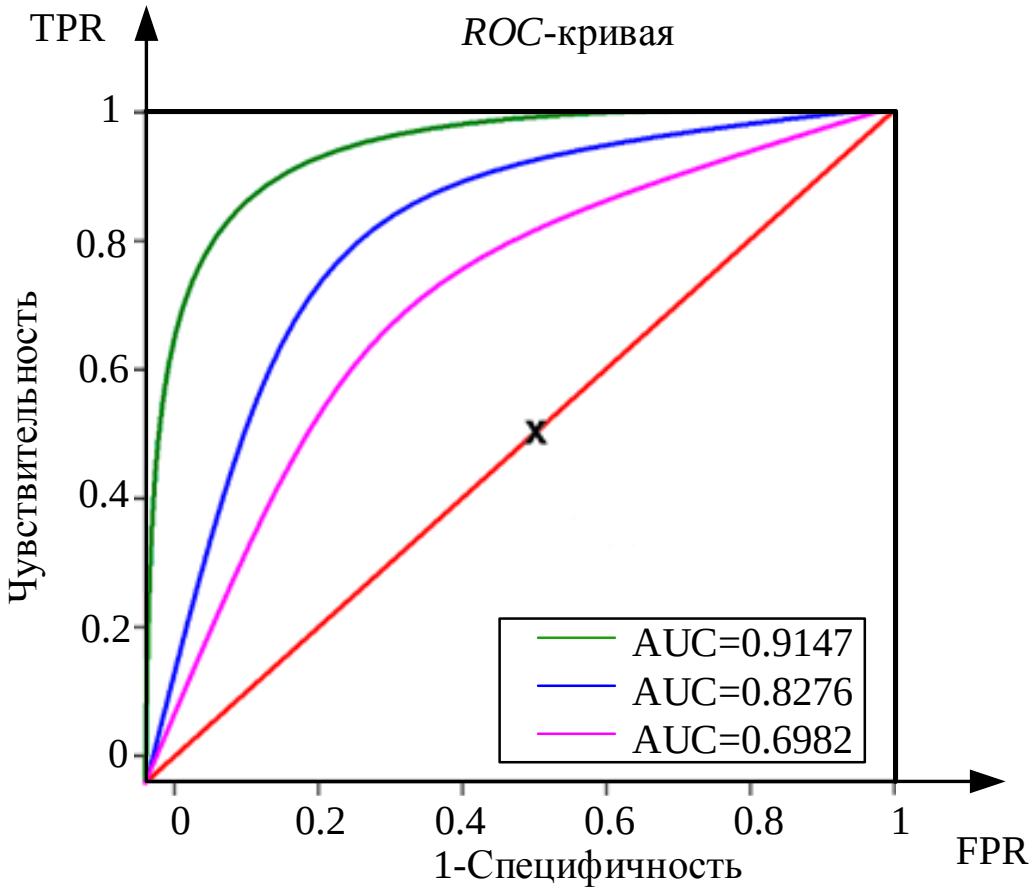

Рисунок 3 – Оценка эффективности модели через значения *AUC* (площади под *ROC*-кривой)
Figure 3 – Assessment of model effectiveness through AUC values (area under the ROC curve)

В литературе [4] приводится следующая таблица 2 для оценки эффективности модели через значения *AUC*.

Таблица 2 – Оценка эффективности модели через значения *AUC*
Table 2 – Assessing the effectiveness of the model through AUC values

Интервал *AUC*	Качество модели
0,9-1,0	отличное
0,8-0,9	Очень хорошее
0,7-0,8	Хорошее
0,6-0,7	Среднее
0,5-0,6	Удовлетворительное

При анализе графиков *ROC*-кривых оперируют следующими показателями [7, 9]:
1. Оценка эффективности метода обучения.

Чем больше площадь под *ROC*-кривой на проверочной выборке и чем ближе кривая проходит к левому верхнему углу, тем лучше данный метод обучения подходит для решения поставленной задачи.

2. Оценка переобучения алгоритма.

При большой разнице между кривыми по обучающей выборке и по проверочной выборке можно говорить об эффекте переобучения.

На графике тонкие линии показывают доверительные интервалы: розового цвета – для контрольной выборки и синего цвета – для обучающей выборки. Если между верхней синей кривой и нижней розовой кривой есть значительный промежуток, то можно говорить, что в данном случае имеет место эффект переобучения.

3. Сравнение *ROC*-кривых для разных классов позволяет оценить отделимость классов.

У надежно отделимого класса *ROC*-кривая на проверочной выборке расположена ближе к левому верхнему углу. В этом случае почти все объекты из этого класса распознаются верно, и совершается минимальное количество ложных обнаружений (когда объекты не из данного класса приписываются к нему).

Плохо отделимый класс имеет кривую, близкую к диагонали (это «бесполезный» классификатор, который не видит разницы между классами).

4. Доверительные интервалы для *ROC*-кривой на контроле позволяют оценить устойчивость метода обучения.

На графике тонкие красные линии показывают доверительные интервалы для *ROC*-кривой на проверочной выборке. Если доверительные интервалы достаточно велики, то это говорит о неустойчивости метода обучения, то есть о сильной зависимости от состава обучающей выборки.

Экспериментальная оценка эффективности нейросетевого стегоанализатора ЦИ

Для получения оценки эффективности предлагаемого нейросетевого классификатора ЦИ был спланирован факторный эксперимент. Начальный этап планирования эксперимента для получения коэффициентов линейной модели основан на варьировании факторов на двух уровнях: нижнем x_{iH} и верхнем x_{iB}, симметрично расположенных относительно основного уровня x_{i0}, $x = 1, \ldots, k$.

Стратегическое планирование заключается в выборе факторов, определении диапазона их изменения (уровни факторов) и построение матрицы плана.

А. Выбор факторов:

Переменные:

1. Независимые варьируемые:

– *x1:* размер группы ЦИ (батч);

– *x2:* число эпох;

– *x3:* скорость обучения.

2. Независимые фиксируемые:

Метод формирования стегоконтейнеров: F5 (ДКП, как один из самых трудно реализуемых).

2. Зависимые:

– *y:* точность (*accuracy*).

Б. Определение диапазона изменения факторов.

Пусть заданы факторы с соответствующими уровнями *x1* = {8, 12, 16}, *x2* = {60,70,80}, *x3* = {0.0004, 0.0005, 0.0006}.

В. Построение матрицы плана.

При числе факторов $k = 3$ и уровнях $l = 2$ общее число точек факторного пространства (уровни факторов) $p = l^k = 2^3 = 8$.

Для ПФЭ выбирается число испытаний $q = 2$ в каждой точке.

Таким образом, общее число испытаний: $N = p \cdot q = 8 \cdot 2 = 16$, результаты которых отображены в Таблице 3.

Таблица 3 – Полный факторный эксперимент посредством двух испытаний
Table 3 – A complete factorial experiment through two trials

Факторы				1-е испытание	2-е испытание	$y_{cp\,i}$	D
№	x_1 *{8, 12, 16}*	x_2 *{60,70,80}*	x_3 *{0.0004, 0.0005, 0.0006}*				
1	(12) –1	(70) –1	(0.0005) –1	70.66	70.66	70.66	0
2	(8, 16) +1	(70) –1	(0.0005) –1	72.34	67.55	83.22	5.74
3	(12) –1	(60, 80) +1	(0.0005) –1	69.11	75.77	73.44	11.09
4	(8,16) +1	(60, 80) +1	(0.0005) –1	86.22	80.66	77.33	7.73
5	(12) –1	(70) –1	(0.0004, 0.0006) +1	68.44	71.22	72.33	1.93
6	(8, 16) +1	(70) –1	(0.0004, 0.0006) +1	78	74.88	74.44	2.44
7	(12) –1	(60, 80) +1	(0.0004, 0.0006) +1	65.88	69.31	78.44	2.95
8	(8, 16) +1	(60, 80) +1	(0.0004, 0.0006) +1	73.77	72.88	73.33	0.19

Вычисление средней точности (математического ожидания) осуществляется по следующей формуле:

$$M(y) = y_{cp\ i} = \frac{1}{q}\sum_{j=1}^{q} y_{i,j} \quad (7)$$

Вычисление дисперсии осуществляется по следующей формуле:

$$D = M\{(y - M(y))^2\} = M(y^2) - M^2(y) \quad (8)$$

$$M(y^2) = \frac{1}{q}\sum_{j=1}^{q}\left(y_{i,j}\right)^2 \quad (9)$$

Для проверки однородности дисперсий выдвигаются следующие гипотезы:

1. H_0 – дисперсия однородна.
2. H_1 – дисперсия неоднородна.

Критерий Кохрена используется при проведении двух или более измерений одних и тех же субъектов для установки наличия или отсутствия различия результатов.

Критерий применяется для выборок одинакового объема. В случае использования критерия Кохрена вычисляют отношение максимальной дисперсии к сумме всех дисперсий:

$$G_{набл} = \frac{max_i D_i}{\sum_{i=1}^{N} D_i} \quad (10)$$

Если

$$G_{набл} > G_{\alpha}(f, n) \quad (11)$$

то принимают гипотезу об отсутствии однородности дисперсий с вероятностью α совершить ошибку.

Процедура сравнения заключается в том, что мы последовательно исключаем наибольшие дисперсии из рассмотрения в случае неоднородности, пока не находим хотя бы две однородные дисперсии или полное отсутствие таковых.

Наблюдение показывает, что $G_{набл} = \frac{11.09}{32.07} = 0.3458$

Находим табличное значение $G_{\alpha}(f, n) = G_{0.05}(f, n)(2; 7) = 0.8332$, где n – число сравниваемых испытаний $(n = 2)$, $f = m - 1 = 7$ – число степеней свободы, $\alpha = 0{,}05$ – уровень значимости.

Следовательно, неравенство (3.13) нарушается, значит, мы принимаем гипотезу об однородности дисперсий, так как 0.3458>0.8332.

Нахождение коэффициентов уравнения регрессии выполняется по следующим формулам:

$$\beta_j = \frac{1}{p}\sum_{i=1}^{p} x_{i,j} \cdot y_i \quad (12)$$

$$\beta_0 = \frac{1}{p}\sum_{i=1}^{p} x_{i,0} \cdot y_i \quad (13)$$

Уравнение регрессии по заданным коэффициентам составляется по следующей формуле:

$$y = b_0 + \sum_{i=1}^{k} b_i \cdot x_i = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 \quad (14)$$

$$y = 72.96 + 2.83 \cdot x_1 + 1.24 \cdot x_2 - 1.16 \cdot x_3$$

Для проверки значимости коэффициентов уравнения регрессии выдвигаются гипотезы:

1. H_0 – коэффициент уравнения регрессии $b_j = 0$, т. е. незначим.
2. H_1 – коэффициент уравнения регрессии $b_j \neq 0$, т. е. значим.

Вычисляем наблюдаемое значение коэффициента регрессии по следующей формуле:

$$b_{набл}, j = \frac{|b_j|}{\sqrt{\frac{\sum_{i=1}^{N} D_i}{N^2 \cdot q}}} \quad (15)$$

$$b_{набл}, 0 = \frac{72.96}{\sqrt{\frac{32.07}{8^2 \cdot 2}}} = \frac{72.96}{0.50054} = 145.76; b_{набл}, 1 = \frac{2.83}{0.50054} = 5.6538; \quad b_{набл}, 2 = \frac{1.24}{0.50054} = 2.47729; b_{набл}, 3 = \frac{1.16}{0.50054} = 2.3174$$

Критерий Стьюдента предназначен для определения статистической значимости различий средних величин. Согласно этому критерию, рассчитываем $f_{крит} = qt(1 - \alpha, N \cdot q - N) = qt(0{,}95; 8) = 2.3060$ и выполняем проверку значимости коэффициентов:

$b_{набл,}0 > f_{крит}$, следовательно, коэффициент значим (принимаем H1);

$b_{набл,}1 > f_{крит}$, следовательно, коэффициент значим (принимаем H1);

$b_{набл,}2 > f_{крит}$, следовательно, коэффициент значим (принимаем H1);

$b_{набл,}3 > f_{крит}$, следовательно, коэффициент значим (принимаем H1).

Следовательно, все коэффициенты уравнения регрессии значимы.

Далее решается задача оптимизации, которая сводится к нахождению сочетания таких параметров обучения СНС (независимых переменных), при которых значение метрики *accuracy* достигает экстремума (максимизируется значение зависимой переменной). Для нахождения оптимального значения точности используется метод Гаусса-Зейделя. Таким образом, решается задача определения значений переменных *x1, x2, x3* (значений факторов), при которых значение *y* стремится к максимальному значению.

При оптимизации по данному методу последовательное продвижение к экстремуму осуществляется путем поочередного варьирования каждым фактором до достижения частного экстремума функции отклика.

Зададим вектор начальных значений системы, соответствующий центру факторного пространства (8; 60; 0.0004) и являющийся исходной точкой.

Анализ влияния факторов начат с переменной *x1*, фиксируя значения *x2, x3* в соответствии с полученным уравнением регрессии (3.16), где коэффициент переменной *x1* имеет большее значение, а соответственно большее влияние на зависимую переменную. Затем производим последовательный поиск максимального значения *y* приемлемого при решении настоящей задачи, аналогично с другими переменными.

Результаты, полученные в ходе эксперимента, представлены в Таблице 4.

Таким образом, в найденных точках функция достигает свое максимально приемлемое значение, следовательно, полученные значения *x1*=8, *x2*=70, *x3*=0.0004 являются точкой оптимума для исследуемого объекта.

Таблица 4 – Нахождение оптимума точности обучения СНС
Table 4 – Finding the optimum of SNN learning accuracy

Влияние переменной *x1*			
x1	*x2*	*x3*	*y*
8	60	0.0004	**73.77**
12	60	0.0004	65.88
Влияние переменной *x2*			
x1	*x2*	*x3*	*y*
8	60	0.0004	73.77
8	70	0.0004	**78**
8	80	0.0004	93.77
Влияние переменной *x3*			
x1	*x2*	*x3*	*y*
8	70	0.0004	**78**
8	70	0.0005	98.88

В ходе проведения исследования был выполнен сравнительный эксперимент по оцениванию качества стегоанализа выбранными статистическими классификаторами и

разработанной СНС. В обобщенном виде результаты экспериментальной оценки по метрике *accuracy* представлены на Рисунке 4 и в Таблице 5. В таблице цветом выделены значения показателей, отражающих наилучшие и наихудшие результаты классификации для конкретных программных средств встраивания информации в ЦИ. Критерием принятия решения является значение точности распознавания 0,6. Выбор этого значения обусловлен тем, что равновероятная точность распознавания соответствует так называемому «бесполезному» классификатору [8, 10].

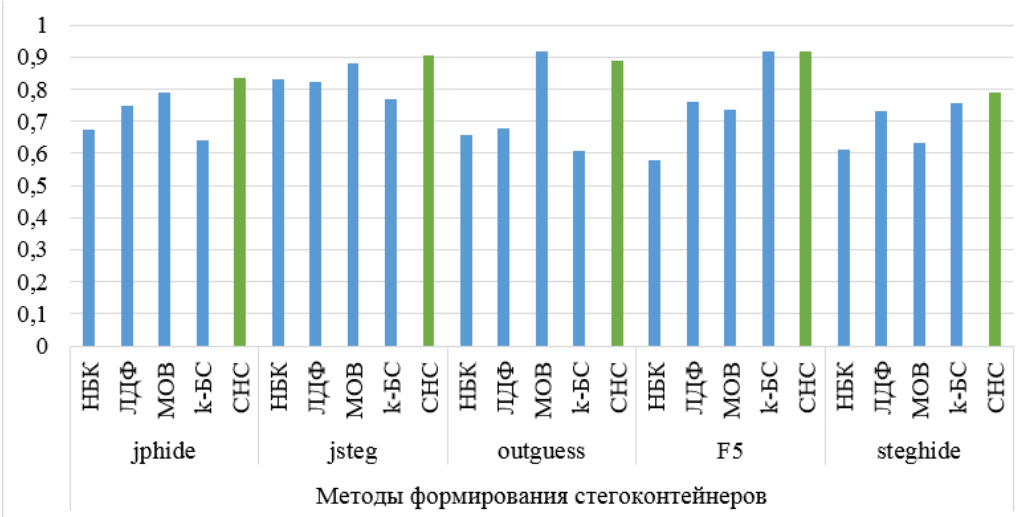

Рисунок 4 – Результаты экспериментальной оценки по метрике accuracy
Figure 4 – Results of the experimental evaluation using the accuracy metric

Таблица 5 – Результаты экспериментальной оценки по метрике accuracy
Table 5 – Results of the experimental evaluation using the accuracy metric

Тестовая выборка по accuracy	Классификатор				
Методы формирования стегоконтейнеров	НБК	ЛДФ	МОВ	k-БС	СНС
jphide	0.6724	0.7476	0.7899	0.6394	0.8372
jsteg	0.8322	0.8322	0.8829	0.7689	0.9068
outguess	0.6598	0.6771	0.9181	0.6097	0.8902
F5	0.5774	0.7601	0.7382	0.9202	0.9201
steghide	0.6106	0.7304	0.6334	0.7570	0.7917

Из Рисунка 4 и Таблицы 5 видно, что полученные значения точности классификации (метрика *accuracy*) для разработанной СНС в общем случае превышают аналогичные значения для статистических классификаторов и превышают выбранный порог точности 0,6, что означает выполнение заданного критерия эффективности ее функционирования.

Заключение

В статье представлены результаты исследования в области стегоанализа ЦИ и совершенствования существующих СЗИ. В ходе решения этой задачи были исследованы актуальные методы стеганографического преобразования ЦИ, и методы стегоанализа

ЦИ, основанные на статистических и нейросетевых классификаторах. В ходе исследования были получены следующие результаты:

1. Предложена структурная схема системы стегоанализа ЦИ на основе методов машинного обучения с учителем.

2. Разработаны и программно реализованы алгоритмы системы стегоанализа ЦИ на основе СНС.

3. Проведен эксперимент по оцениванию эффективности системы стегоанализа ЦИ на основе СНС.

Основным результатом исследования является комплекс программ, обеспечивающих как поддержку процесса обучения классификаторов за счет автоматизированной генерации, тестовой и обучающей выборок, так и поддержку конфигурирования параметров разработанной СНС.

Полученные экспериментальные результаты свидетельствуют об эффективности разработанных алгоритмов и программных решений.

Направлениями дальнейших исследований являются:

– проведение исследований в предметных областях, актуальных с точки рения угроз ИБ методов и средств стеганографического преобразования ЦИ, а также методов и средств стегоанализа ЦИ;

– исследование возможности применения нейронных сетей других классов, например, генеративно-состязательных нейронных сетей;

– совершенствование алгоритмов для модулей генерации стегоконтейнеров и конфигурирования параметров СНС.

Полученные в ходе исследования результаты можно использовать для совершенствования существующих и разработки перспективных систем комплексного мониторинга и управления ИБ.

СПИСОК ИСТОЧНИКОВ

1. Шипулин П. Стеганография. СФУ; 2017. Режим доступа: http://security.pmkb.sfu-kras.ru/blog/steganografiya/ (дата обращения: 20.02.2022).
2. Генне О.В. Основные положения стеганографии. *Защита информации. Конфидент*. 2000;(3):36–39.
3. Колесников А.А., Яндашевская Э.А. Теоретико-информационный подход к моделированию распределенной стеганографической системы с пассивным противником. *Системы управления и информационные технологии*. 2020;3(81):19–23.
4. Башмаков Д.А. *Методы и алгоритмы выявления встроенных сообщений в пространственной области неподвижных изображений при малой полезной нагрузке: дис. на соискание ученой степени канд. техн. наук*. Санкт-Петербург; 2018. 150 с.
5. Гребенников В.В. *Стеганография. История тайнописи.* М.: ЛитРес: Самиздат; 2019. 160 с.
6. Яндашевская Э.А. Разработка подсистемы стегоанализа цифровых изображений на основе сверточной нейронной сети для обнаружения и предовращения атак, использующих скрытые стеганографические каналы. *Доклады ТУСУР (ВАК).* 2021;24(2):29–33.
7. Фукунага К. *Введение в статистическую теорию распознавания образов.* Пер. с англ. М.: Наука. Главная редакция физико-математической литературы; 1979. 368 с.
8. Яндашевская Э.А., Полунин А.А. Использование аппарата свёрточных нейронных сетей для стегоанализа цифровых изображений. *Сборник материалов*

Международной конференции «Иванниковские чтения», Труды ИСП РАН. 2020;32(4):155–164.
9. Килбас И.А., Паринger Р.А. Сравнение точности распознавания сцен и производительности свёрточных нейронных сетей. *Науки о данных: Сборник трудов V Международной конференции и молодёжной школы «Информационные технологии и нанотехнологии»*. 2019;740–747.
10. Сикорский О.С Обзор свёрточных нейронных сетей для задачи классификации изображений. *Новые информационные технологии в автоматизированных системах.* 2017;(20):45–53.

REFERENCES

1. Shipulin P. Steganografiya. Siberian Federal University; 2017. Available by: http://security.pmkb.sfu-kras.ru/blog/steganografiya/ (accessed on 20.02.2022). (In Russ.)
2. Genne O.V. Osnovnye polozheniya steganografii. *Zashita informacii. Confident.* 2000;(3):36–39. (In Russ.)
3. Kolesnikov A.A., Yandashevskaya E.A. Teoretiko-informatsionnyi podkhod k modelirovaniyu raspredelennoi steganograficheskoi sistemy s passivnym protivnikom. *Sistemy upravleniya i informatsionnye tekhnologii.* 2020;3(81):19–23. (In Russ.)
4. Bashmakov D.A. *Metody i algoritmy vyyavleniya vstroennykh soobshchenii v prostranstvennoi oblasti nepodvizhnykh izobrazhenii pri maloi poleznoi nagruzke: dis. na soiskanie uchenoi stepeni kand. tekhn. nauk*. Saint-Petersburg; 2018. 150 p. (In Russ.)
5. Grebennikov V.V. Steganografiya. Istoriya tainopisi. Moscow: LitRes: Samizdat; 2019. 160 p. (In Russ.)
6. Yandashevskaya E.A. Razrabotka podsistemy stegoanaliza tsifrovykh izobrazhenii na osnove svertochnoi neironnoi seti dlya obnaruzheniya i predovrashcheniya atak, ispol'zuyushchikh skrytye steganograficheskie kanaly. *Doklady TUSUR (VAK).* 2021;24(2):29–33. (In Russ.)
7. Fukunaga K. *Introduction to statistical pattern recognition.* Purdue university; 1972. 368 p.
8. Yandashevskaya E.A., Polunin A.A. Ispol'zovanie apparata svertochnykh neironnykh setei dlya stegoanaliza tsifrovykh izobrazhenii. *Sbornik materialov Mezhdunarodnoi konferentsii «Ivannikovskie chteniya», Trudy ISP RAN*. 2020;32(4):155–164. (In Russ.)
9. Kilbas I.A., Paringer R.A. Sravnenie tochnosti raspoznavaniya stsen i proizvoditel'nosti svertochnykh neironnykh setei. *Nauki o dannykh: Sbornik trudov V Mezhdunarodnoi konferentsii i molodezhnoi shkoly «Informatsionnye tekhnologii i nanotekhnologii»*. 2019;740–747. (In Russ.)
10. Sikorskii O.S Obzor svertochnykh neironnykh setei dlya zadachi klassifikatsii izobrazhenii. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh.* 2017;(20):45–53. (In Russ.)

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Минайчев Антон Андреевич, МГТУ им. Н.Э. Баумана, НТЦ «Орион», Москва, Российская Федерация.
e-mail: anton.minaichev@gmail.com
ORCID: 0000-0003-2201-5337

Anton Andreevich Minaichev, Bauman Moscow State Technical University, Science and Technology Center “Orion”, Moscow, Russian Federation.

Мезенцев Александр Олегович, Академия Федеральной службы охраны, Орёл, Российская Федерация.

Aleksandr Olegovich Mezentsev, Federal Guard Service Academy, Oryol, Russian Federation.

Яндашевская Элина Андреевна, Академия Федеральной службы охраны, Орёл, Российская Федерация.

Elina Andreevna Yandashevskaya, Federal Guard Service Academy, Oryol, Russian Federation.

Статья поступила в редакцию 03.06.2022; одобрена после рецензирования 14.06.2022; принята к публикации 28.06.2022.

The article was submitted 03.06.2022; approved after reviewing 14.06.2022; accepted for publication 28.06.2022.