Проверка подлинности документов с использованием нейросетевых технологий

Тарасов Денис Андреевич
Студент ФГБОУ ВО «Поволжский Государственный университет телекоммуникаций и информатики, г. Самара

Аннотация. Статья знакомит читателей с различными методами проверки подлинности электронных документов, используя нейросетевые технологии. Предложены различные способы проверки подлинности документов с помощью специально обученной нейросети. Выводы сделаны на основе использования различных методов и способов проверки подлинности электронных документов. Приведены сравнения различных способов проверки подлинности и распознавания электронных документов. Приведены их достоинства и недостатки.

Ключевые слова: нейросеть, проверка подлинности, электронные документы.

Annotation. The article acquaints readers with various methods of recognition and authentication of electronic documents using neural network technologies. Various methods of document authentication using a specially trained neural network are proposed. Conclusions are made based on the use of various methods and methods for verifying the authenticity of electronic documents. Comparisons of various methods of authentication and recognition of electronic documents are given. Their advantages and disadvantages are given.

Keywords: neural network, authentication, electronic documents.

В современном обществе, когда многие действия совершаются с помощью электронных документов, возникает риск подделки электронных документов. Подделка электронных документов несёт негативные последствия как для организации, так и для экономики страны в целом. Чтобы избежать негативные последствия использования поддельных электронных документов, используются различные методы проверки подлинности, среди которых наиболее распространённым является, использование нейро-сетевых технологий.

Принцип проверки подлинности заключается в создании специально обучаемой нейросети, в составе которой лежит способность распознавать сложные зависимости между входными и выходными данными, а также выполнять обобщение и сравнивать полученный образ с непроверенными на подлинность документами [2]. Если образ совпадает с непроверенным документом, то непроверенный документ подлинный, если образ не совпадает с непроверенным документом, то непроверенный документ поддельный. С использованием нейро-сетевых технологий можно проверять любые документы, как электронные, например: электронные трудовые книжки, банковские карты, медицинские справки, счета и накладные в электронном виде, так и бумажные, например: наличные деньги, дипломы об образовании, медицинские справки, счета, накладные в бумажном виде и т. д.

Нейросеть будет реализована на языке программирования Python через интегрированную среду PyCharm. Для обучения и тестирования системы создан специальный набор обучающих данных, который состоит из документов, находящихся в электронном документообороте. В составе документов будет находиться 100 файлов. Изображения документов обработаны с помощью специальной программы BeFunky. В ходе разработки нейросети использованы библиотеки Darknet и PyTesseract.

Для обучения модуля используется собственный набор обучающих данных, который состоит

из 100 документов. 10 документов используются для тестирования. Формат документов — jpg и pdf. На рисунке 1 представлен документ, подлежащий проверке.

Рис. 1 Образец документа, подлежащий проверке

В качестве объектов, по которому будет проверяться документ, изображённый на рисунке 1, будут использоваться водяные знаки, изображённые на банковской купюре. На рисунке 2 изображён образ, по которому будет проверяться документ, изображённый на рисунке 1. В основе проверки будет происходить сравнение водяных знаков на образе и на документе. Если на документе будут находиться водяные знаки, находящиеся на образе, то документ подлинный. Если на документе не будет находиться водяных знаков, находящихся на образе, то документ поддельный.

Рис. 2 Образ, по которому проводится проверка документа

На данном рисунке выделены водяные знаки, по которым проводится проверка документа. Координаты выделенных участков сохраняются в виде текстовых документов.

При обучении модели использовались все 100 файлов.

Несмотря на все преимущества, которые даёт проверка подлинности документов с использованием нейросетей, существует один серьёзный недостаток — неправильное распознавание образов и непроверенных документов [3]. Неправильное распознавание образов может привести к использованию поддельных документов, что может привести к негативным последствиям как для организации, так и для экономики в целом. Чтобы избежать неправильного распознавания образов и непроверенных документов, нужно периодически контролировать работу нейросети, 1 раз в месяц производить проверку нейросети при помощи сравнения действующего образа с непроверенным документом, который нигде не будет использоваться. Затем проверить непроверенный документ любыми другими способами. Если проверки непроверенного документа

другими способами дали другой результат, значит проверка документов с помощью нейросети даёт неверный результат и такая нейросеть для проверки подлинности документов использоваться не будет.

Для того чтобы нейросеть работала качественно и без сбоев организации нужно нанимать специальных программистов, которые будут контролировать работу нейросети, а в случае непредвиденных сбоев и неправильной работы нейросети — быстро корректировать её работу [1].

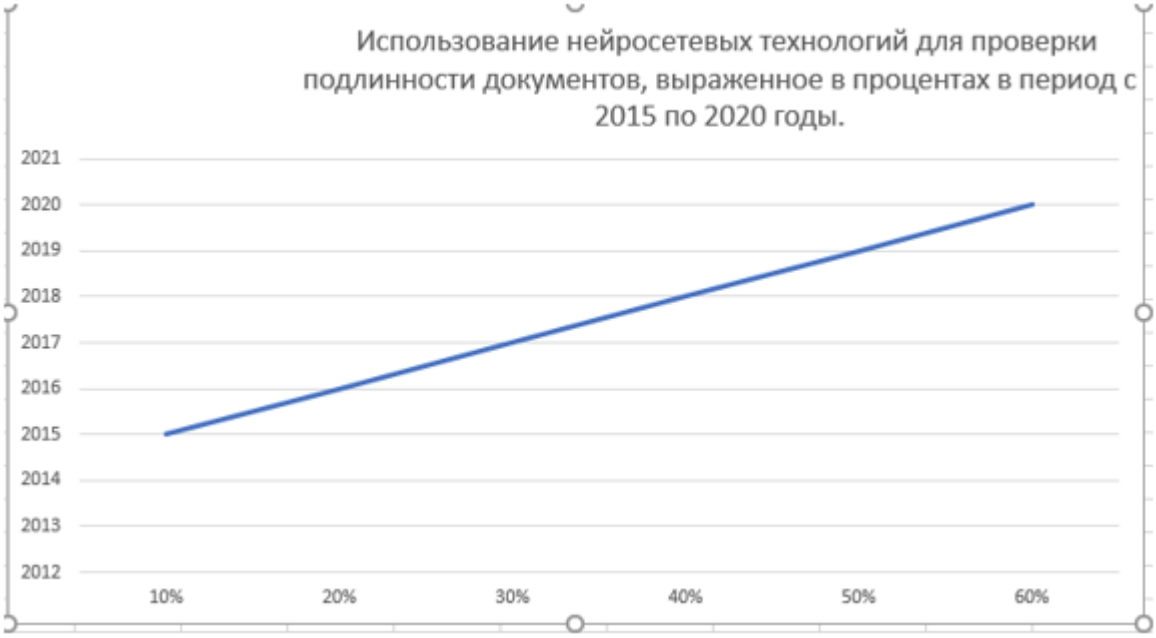

Рис. 3 График зависимости использования нейросетевых технологий в период с 2015 по 2020 годы.

Из графика видно, что с каждым годом использование нейросетевых технологий для проверки документов становится всё более популярным. Например, в 2015 году, когда использование нейросетевых технологий в программировании только зарождалось, использовали нейросетевые технологии для проверки подлинности документов только 10 % различных организаций. Тогда как, в 2016 году нейросетевые технологии стали использовать уже 20% организаций. В 2020 году этот процент вырос до 60. Планируется, что к 2024 году все документы будут проверяться на подлинность с использованием нейросетевых технологий.

Причина такого подъёма в использовании нейросетевых технологий для проверки документов заключается в постоянном совершенствовании нейросетей, быстрота использования, высокое качество проверки документов и минимальный процент сбоев в работе нейросети. Также основной причиной использования нейросетей является стремительный переход на электронные документы во всех сферах человеческой деятельности. Нейросети позволяют быстро проверить на подлинность любые документы (электронные и бумажные), тем самым ускорив выполнение различных операций, например, трудоустройство на работу и оформление банковского кредита.

Список используемой литературы

1. Как работает нейронная сеть: алгоритмы, обучение, функции активации и потери [электронный ресурс] // Режим доступа: https://neurohive.io/ru/osnovy-data-science/osnovy-nejronnyh-setej-algoritmy-obuchenie-funkcii-aktivacii-i-poteri/

2. Типы нейронных сетей. Принцип их работы и сфера применения [электронный ресурс] // Режим доступа: https://otus.ru/nest/post/1263/

3. Что такое нейронные сети и как они работают? Классификация искусственных нейросетей

[электронный ресурс] // Режим доступа: https://reobzor.ru/reviews/2020/12/chto-takoe-nejroseti-ins-i-kak-oni-rabotayut/