netskope

White Paper

Protecting Data Using Artificial Intelligence and Machine Learning

INTRODUCTION

Data is rapidly growing in volume, and it's becoming increasingly unstructured, such as in the form of images captured through screenshots or smartphone photos. As a result, sensitive data is becoming harder to detect and consequently protect.

Effectively, with the fast pace of modern business communication, users have adopted new data sharing behaviors that undermine the effectiveness of conventional data identification methods that data loss prevention (DLP) solutions mainly rely on. In fact, to facilitate frequent information sharing, users' communication has shifted toward image sharing. This practice serves to quickly convey concepts, present visual evidence, offer on-the-go diagrams and slides, exchange contact details, and much more. Consequently, while it's essential for DLP to have strong text analysis capabilities, as many and sophisticated as possible (such as thousands of regular patterns detection, precise content matching, data fingerprinting and optical character recognition [OCR]), the reliance on static text-based identification methods alone has become less effective in the contemporary data security landscape.

To address these modern challenges, DLP must leverage the power of artificial intelligence through dynamic deep learning, natural language processing (NLP), and new methods of machine learning-based image classification. DLP technology should be especially able to emulate human visual recognition, comprehending visual features and details to understand the image as a whole and enabling rapid and accurate identification of sensitive content even within low-quality images.

Netskope's Comprehensive Data Insights and Data Detection Accuracy

Every day the Netskope Security Cloud processes billions of events and files, capturing a wide variety of data and user activities across:

- SaaS applications, such as Microsoft 365, Box, Salesforce, Teams, and G Suite
- Public cloud infrastructure services, such as Amazon Web Services, Microsoft Azure, and Google Cloud Platform
- Websites that users have visited, unsanctioned SaaS apps, and personal instances of corporate SaaS apps
- Endpoint devices and their localized activities, such as USB and printing
- Email services on-premises and in the cloud

The Netskope solution assembles a comprehensive understanding of all enterprise data at rest and in motion. For example, Netskope has profound insights into the life cycle of every file, such as how a file is uploaded, downloaded, or shared within managed cloud storage apps, when it is transferred to shadow cloud applications, to personal SaaS instances, downloaded to personal devices, etc. This contextual understanding of user activities and corporate data, bolstered by a complete view of cloud and web traffic, enables the protection of all corporate sensitive data that employees access and use.

What sets Netskope apart is our remarkable proficiency in precisely differentiating specific sensitive data from the sea of information within a corporate environment. This precision ensures that our customers achieve the utmost accuracy in detection, sparing them the frustration of false positives. Netskope's unwavering commitment to enhancing data detection, classification accuracy, and efficiency has driven significant investments and advancements in artificial intelligence and machine learning.

Artificial Intelligence and Machine Learning 101

Artificial intelligence (AI) is an umbrella term for computer programs that mimic human brain functions. While AI has been under study since the 1950s, recent years have witnessed significant advancements in AI's ability to "learn" and "problem solve." These advancements have found applications across various technology sectors, including cybersecurity.

Machine learning (ML) is a subset of AI in which software programs are designed to learn from examples. In the realm of cybersecurity, an illustrative use case for ML is a program that can categorize spam emails after exposure to a substantial dataset of known spam emails. ML employs sample data or "training data" to construct a mathematical model for predicting or making decisions regarding future input data.

A subset of ML, known as deep learning, strives to emulate human brain processes. The term "deep" alludes to the multiple layers in its artificial neural network. Over the past decade, deep learning and artificial neural networks have achieved groundbreaking levels of accuracy in various critical domains of data protection, such as image recognition and natural language processing.

Artificial Intelligence and Machine Learning at Netskope

As a leader in cloud security, Netskope is at the forefront of developing and integrating the latest AI/ML technology into its data and threat protection capabilities, as well as its business operations. Our dedicated team comprises data scientists, security researchers, and product engineers, each with a proven track record of solving security and fraud challenges across diverse domains. Together, this team has collectively developed hundreds of patents.

Leveraging this deep expertise in AI/ML, Netskope is actively developing large-scale AI/ML solutions tailored for data protection and security. Some of the customer use cases for which Netskope employs AI and ML include:

- Discovering and governing the responsible use of generative AI amidst the rapid proliferation of novel SaaS apps, with automatic protection of sensitive data across generative AI apps and real-time user coaching
- User and entity behavior analytics (UEBA) for detecting malicious insiders, compromised accounts, brute-force attacks, and data exfiltration incidents
- Detection of malware using ML models, complementing traditional antivirus signatures, threat intelligence, heuristics, and sandboxes
- Categorization and identification of malicious web domains, URLs, and web content

Lastly, the primary focus of this paper is:

- Assisting organizations in protecting all their sensitive data (both structured and unstructured) and adhering to compliance regulations such as GDPR, CCPA, PCI, HIPAA, or their own internal data protection policies. This is achieved through patented AI/ML automated DLP detection of sensitive information in documents, images, and other modern forms of data across any communication vector.

Data Loss Prevention

Netskope DLP is the industry's most advanced cloud data loss prevention solution, ensuring comprehensive protection for sensitive data across clouds, networks, email, endpoints, and users anywhere. It seamlessly integrates with Netskope's Security Service Edge solution, providing unified policy enforcement across all critical channels.

Netskope DLP consistently discovers, monitors, and safeguards sensitive data across various environments and all critical channels, including SaaS applications, IaaS, corporate networks, mobile workforce's devices, and email services. It offers unified data protection policies and centralized management through a single console.

Our DLP solution achieves high data detection and classification accuracy while minimizing false positives, thanks to a wide range of advanced data detection technologies and classification algorithms.

Netskope's cloud-based DLP is context-aware data protection. It granularly collects and identifies data risks, including cloud and web access, users, devices, apps, app instances, activities, and content involved, allowing to dynamically take appropriate actions, such as allowing, blocking, notifying, coaching, quarantining, encrypting, applying legal holds, etc.

Whether inspecting data in motion in real time or at rest, Netskope offers precise inspection through a comprehensive set of thousands of global data identifiers, compliance templates, various content matching techniques, and OCR.

MACHINE LEARNING FOR DLP

File classification through machine learning offers a rapid and effective means of identifying sensitive information and is a reliable method of identification for unstructured data. This enables users to work seamlessly with precise, real-time DLP policy controls. ML classifiers excel at accurately categorizing documents and images, based on similarities, into various categories, such as tax forms, patent documents, source code, passports, driver licenses, payment cards, screenshots, and more. Importantly, this classification process doesn't necessarily require specific character extraction and identification of sensitive textual information within these files.

Security administrators can then create DLP policies based on these categories. In this manner, ML file classification serves as a complementary approach to conventional DLP textual rules, enhancing the security of sensitive data like information regulated by PII, PCI, and HIPAA.

Image classification

Image classification is a crucial component of DLP, enabling organizations to track and protect sensitive data within images, whether such data is stored or shared. Traditionally, sensitive data in images, such as patient information in medical X-rays, is identified using OCR to extract text from the image. However, OCR is resource-intensive and can cause delays and errors in sensitive data detection and identification of security violations, especially with low-quality images.

In many cases, it would be sufficient to recognize an image without always extracting every textual information. For example, to determine whether an image is a passport or not, it would be enough to

recognize common visual characteristics of the image itself without needing to extract personal details, such as the person's name. ML-based image classification, utilizing deep learning and convolutional neural networks (CNNs), provides a more accurate, resource-efficient, and secure alternative.

Image classification doesn't always have to replace textual analysis but can also be combined with OCR in order to complement textual analysis and therefore enhance data detection, leading to very accurate policy violations.

CNN Architecture

Deep learning and CNNs marked significant breakthroughs in image classification in the early 2010s. medicine, autonomous vehicles, and cybersecurity, achieving accuracy levels approaching those of humans.

Inspired by the functioning of the human visual cortex, a CNN is able to effectively capture shapes, objects, and other characteristics to better comprehend an image's contents. A typical CNN, illustrated in Figure 1, consists of two main parts: the convolutional base and the classifier.

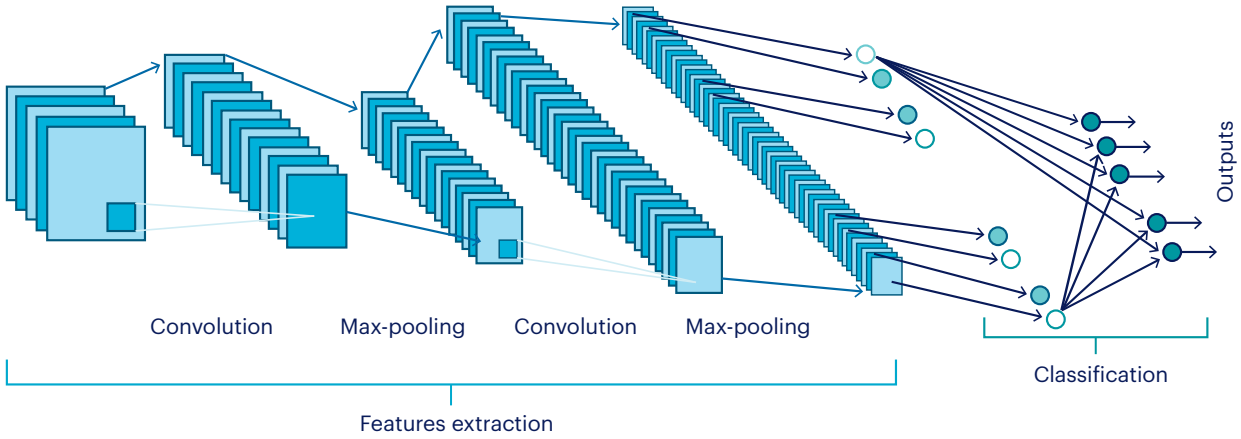


Figure 1: A diagram of a typical CNN

To illustrate, consider how a human would identify a passport. When presented with an image, a human's eye would search for distinctive attributes and features that define a passport, such as a headshot, the word "passport", a name, dates, barcodes, etc. The convolutional base operates similarly, utilizing a stack of convolutional and pooling layers.

The primary objective of the convolutional base is to generate progressively higher-level features from the analyzed image. The early layers refer to general features like edges, lines, and dots, while the later layers focus on task-specific features, which are more interpretable to humans, such as credit card logos or application windows in a screenshot.

The classifier is usually composed of fully connected layers. Think of the classifier as a machine that sorts the features identified in the convolutional base. The classifier will tell whether the identified features represent a cat, dog, driver license, or X-ray.

By employing a technique known as transfer learning, we take advantage of pre-existing CNNs that have been trained on large-scale datasets and fine-tune them using a smaller dataset of labeled positive images that contain sensitive information and negative samples. As a result, our classifiers are able to quickly identify the unique patterns associated with the sensitive information, with high accuracy and reduced training time.

Training Data

We collect positive and negative sample images from various sources to train our classifiers, without using our customers' data. In order to minimize false positives, it is important for our image classifiers to be exposed to a wide variety of realistic negative samples. To achieve this, we have sourced tens of thousands of actual cloud images from our own corporate data. This approach enables us to collect a substantial number of genuine training images, while simultaneously maintaining our commitment to customer privacy. These images were labeled by hand, with the majority of them being either negative examples or screenshots typical of real-world cloud data.

In addition to these random negative examples, we have also incorporated several thousand carefully curated adversarial samples, further bolstering our classifiers' resilience against false positives. One interesting type of adversarial sample was labels for electronics (as shown in Figure 2 below). Due to their bold fonts and high contrast coloring, they can be mistaken for sensitive documents. By training our classifiers on these adversarial examples, we can effectively prevent such misclassifications in the production environment.

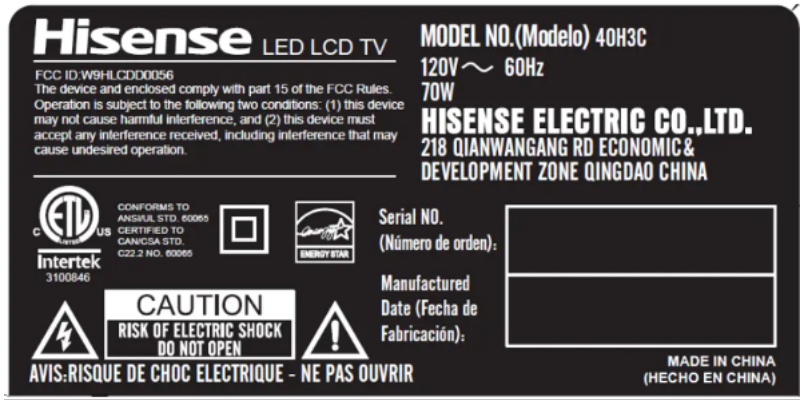


Figure 2: Example of an adversarial sample

In addition to sourcing real cloud data, we employ a comprehensive suite of data augmentation techniques specifically designed for computer vision applications, such as rotation and cropping. What sets our approach apart is the customization of these augmentations to ensure maximum fidelity with the image data encountered in real cloud environments. One example is our custom augmentation that seamlessly integrates documents onto realistic backgrounds, such as a driver license pasted on a screenshot. This enables our classifiers to train on documents in a diverse range of settings, significantly boosting its versatility and performance on real-world data. Figure 3 below shows a training sample of a driver license pasted on a realistic background, in this case a screenshot.

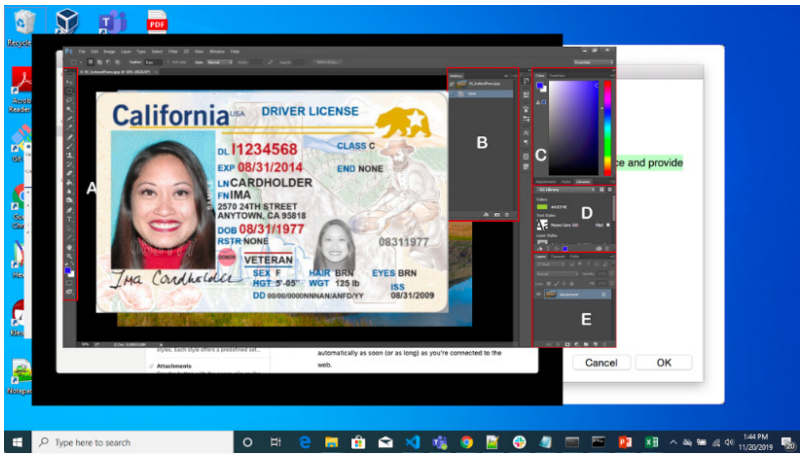


Figure 3: Synthetic screenshot used as a training sample

Netskope has developed CNN-based image classifiers for use with its NG SWG and Cloud Inline solutions. The classifiers can accurately identify images containing sensitive information, including passports, driver licenses, U.S. Social Security cards, credit and debit cards, and full-screen and application screenshots. These inline classifiers allow for granular policy controls to be applied in real time.

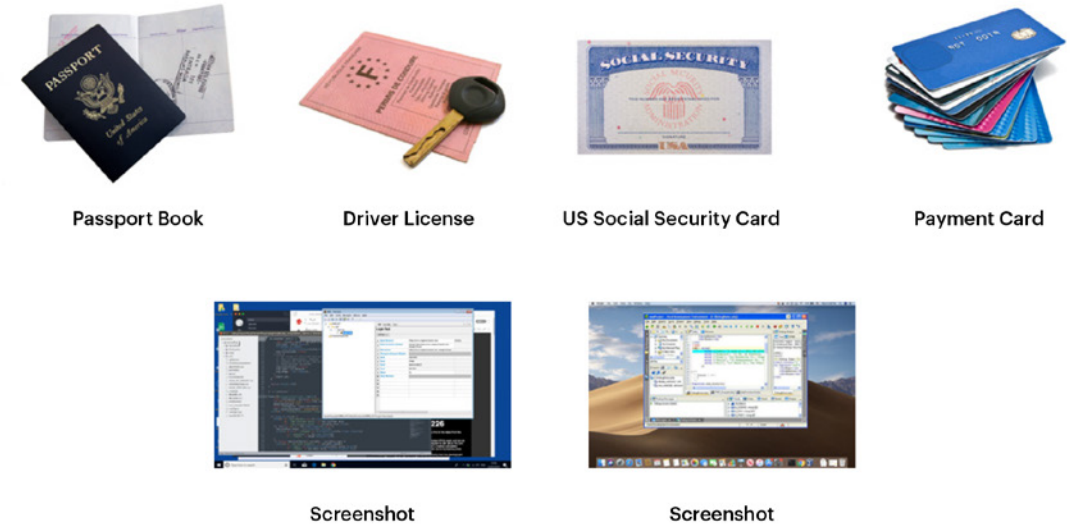


Figure 4: Examples of sensitive data that can be identified using image classification

Document classification

Similar to image classification, document classification is a key part of DLP that uses ML to improve accuracy and therefore reduce false positives. While the process to identify sensitive data is similar, there are some key differences to note.

Many of the documents that an organization's users transfer through or store in the cloud contain sensitive information, including confidential legal and financial documents, intellectual property, and employee or user personally identifiable information (PII). ML-based document classifiers automatically classify documents into different categories, including source code, tax forms, patents, and bank statements. ML classifiers complement the more traditional text-matching or regex-based DLP rules. In many cases, manually configured regex rules can generate excessive false positives or false negatives when looking for specific patterns in documents. In comparison, ML classifiers automatically learn the patterns that identify sensitive data in real time, without the need for traditional DLP rules.

Text classification is one of the standard natural language processing (NLP) tasks. As illustrated in Figure 5, text content is extracted from documents and a pre-trained language model is used as an encoder to convert documents into numeric values that capture the contextual and semantic information of the documents' words. Based on these document encodings, document classifiers are trained using fully connected neural network layers.

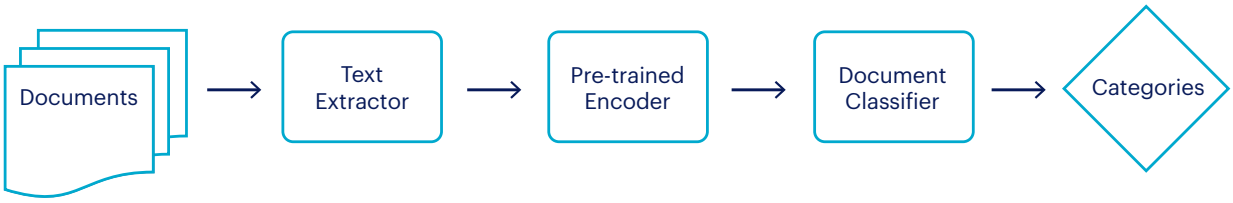
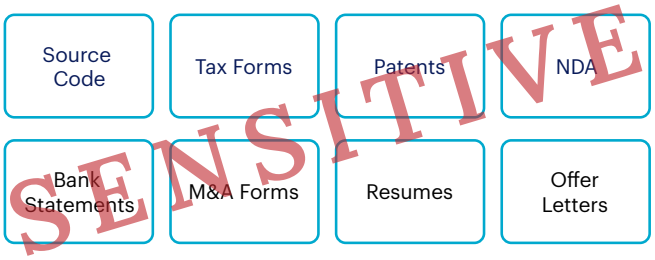


Figure 5: The NLP process of document classification

Currently, Netskope document classifiers can accurately identify more than 12 types of documents containing sensitive information, including:

- Source code
- IRS tax forms
- M&A forms
- Resumes
- U.S. patent files
- Offer letters
- Bank statements
- Nondisclosure agreements
- Consulting agreements
- Partner agreements
- Stock agreements
- Medical power of attorney forms



These lightweight, high-speed document classifiers are integrated into the Netskope DLP offering and can be used in conjunction with inline or API-based policies to provide data protection for Netskope customers.

TYOC: Train Your Own Classifiers

Different industries and organizations deal with different types of sensitive data. This could include things like identity cards, HR documents, or critical infrastructure images. Netskope has come up with a new way for customers to train their own classifiers for images or documents, all while keeping their data private. This method allows organizations to focus on the information that is most important to them.

Netskope uses advanced contrastive learning techniques to make this possible. Customers can upload a small number of example documents (around 20-30) to the Netskope Security Cloud. From these examples, important features are extracted and used to train a customized classifier with Netskope's ML engine. The training process is done in a way that doesn't require a large amount of labeled data or a time-consuming retraining of a supervised classification model.

Once the custom classifier is trained, it is deployed into the customer's own tenant and used to detect sensitive information in their web and cloud traffic. It's important to note that the uploaded sample data and trained classifier are not shared with any other customers. This ensures that the customer's sensitive data remains protected throughout the process.

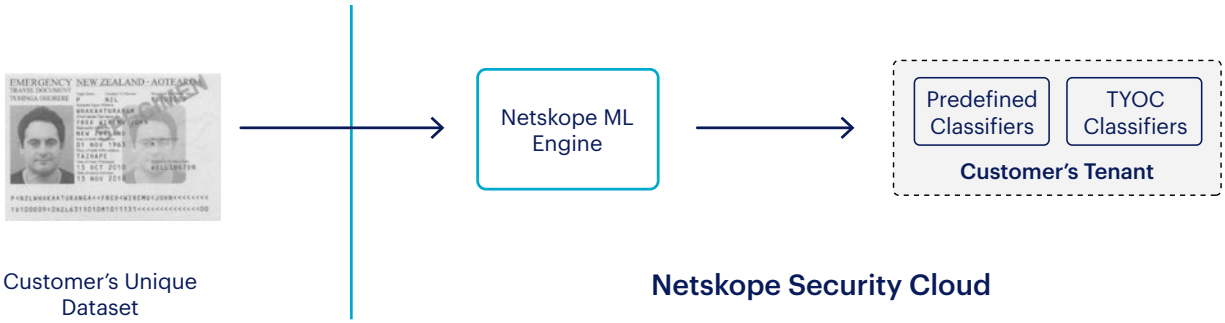


Figure 6: Overview of TYOC process

Not only does this system of "training your own classifiers" allow organizations to get the most from their Netskope DLP capability, but it also avoids any risk of exposing sensitive customer information to Netskope at any stage of the process.

Document Classification — Case Study

A real estate company uses Microsoft Office 365 OneDrive for Business as corporate storage for all its documents, including sensitive tax and HR forms. It needs to ensure that these documents are not leaked and shared publicly.

A DLP policy was configured within Netskope to detect the download or share activities of Schedule K-1 (Form 1065) tax forms, which report on a partner's share of the income, deductions, credits, and more of its business's information. Initially, the DLP policy used document fingerprints to identify K-1 forms for the past 10 years. However, only 70 percent of the K-1 forms were successfully detected by this policy due to different variations of the form. The DLP policy was updated to use the tax form document classifier combined with keywords such as "Schedule K-1" and "Form 1065." As a result of the ML classifier being used, 100 percent of the K-1 forms were identified with zero false positives, effectively preventing sensitive tax forms from being leaked.

SUMMARY

At Netskope, we embrace AI and ML technology to create future-looking, more efficient and secure products. Our native integration of highly differentiated ML-based image and document classification into our DLP solution demonstrates our commitment to continuous innovation in data protection. This approach provides more resource efficiency, higher accuracy, and better data security superior to traditional and conventional DLP solutions of today. Netskope also provides organizations the ability to train the solution with their own ML classifiers for proprietary documents and images. We are continuously expanding our portfolio of inline file classifiers to meet our customers' needs.

Netskope's ML-based classification is available as part of the Netskope SkopeAI DLP license and the Netskope Advanced DLP license, and is included in the Netskope Enterprise packages of the Netskope products.

We are continuously expanding our AI and ML technologies in data protection to meet our customers' needs, and we welcome the opportunity to demonstrate our capabilities to your organization whether you are an existing Netskope customer or interested in learning more about our products. Please contact us for more information.

To learn more, visit:
https://www.netskope.com/data-loss-prevention

Interested in learning more?

Request a demo

Netskope, a global SASE leader, is redefining cloud, data, and network security to help organizations apply zero trust principles to protect data. Fast and easy to use, the Netskope platform provides optimized access and real-time security for people, devices, and data anywhere they go. Netskope helps customers reduce risk, accelerate performance, and get unrivalled visibility into any cloud, web, and private application activity. Thousands of customers, including more than 25 of the Fortune 100, trust Netskope and its powerful NewEdge network to address evolving threats, new risks, technology shifts, organizational and network changes, and new regulatory requirements.

Learn how Netskope helps customers be ready for anything on their SASE journey, visit netskope.com.

©2023 Netskope, Inc. All rights reserved. Netskope is a registered trademark and Netskope Active, Netskope Cloud XD, Netskope Discovery, Cloud Confidence Index, and SkopeSights are trademarks of Netskope, Inc. All other trademarks are trademarks of their respective owners. 10/23 WP-429-3