

Research Article

Vikas Khullar, Isha Kansal, Jyoti Verma, Rajeev Kumar*, Karuna Salgotra, and Gurpreet Singh Saini

Deep trained features extraction and dense layer classification of sensitive and normal documents for robotic vision-based segregation

<https://doi.org/10.1515/pjbr-2022-0125>

received March 17, 2023; accepted September 16, 2023

Abstract: The digitization of important documents and their segregation can be a beneficial and time-saving activity as individuals will have greater access to important documents and will be able to use them in regular tasks as well as endeavours. In recent years, research into the application of deep networks in robot systems has increased as a direct consequence of the advancements made in classification algorithms over the past few decades. Robotic vision automation for the segregation of sensitive and non-sensitive documents is required for many security concerns. The methodology of this article is initially focused on the identification of a good computer vision-based technique for the classification of sensitive documents from non-sensitive documents. The authors first identified the standard parameters in terms of reliability, loss, precision, and recall by employing deep learning techniques, such as neural networks with convolutions and transfer learning (TL) algorithms. The extraction of features based on pre-trained deep learning models was referenced in numerous publications. Similarly, we applied most of the feature extraction

techniques to identify feature extraction from the images. Then, these features were classified by machine and ensemble learning models. However, the pre-trained models-based feature extraction along with machine learning classification resulted better in comparison to the deep learning and TL procedures. Further, the better-identified techniques were applied as the brain behind the vision of a robotic structure to automate the segregation of sensitive documents from non-sensitive documents. This proposed robotic structure could be applied when we have to find some specific and classified document from the haystack.

Keywords: deep learning, machine learning, image processing, robotics

1 Introduction

Identity documents have a specific layout and include both text and graphics [1]. Identity fraud is a significant problem with negative repercussions in today's societies. From simple fraud to organized crime and terrorist attacks, there are many different types of risks [1]. Identity theft on social media is a serious problem. Nonverbal contacts, such as user movement or activity, are more practical than verbal contact (speech or text). Identity deception categorizes into three parts: identity concealment, identity stealing, and identity fraud. It concentrates on modifying the sender's details [2]. Governments and businesses alike have been encouraged to digitize the number of processes concerning the widespread use of computers and smartphones. With the growth of online financial services over the past few years, it has become increasingly customary to onboard potential clients, pay taxes, and verify one's identity using a computer [3]. Digital technology is incorporated into every facet of business operations to raise service standards. Digital transformation projects have been carried out by numerous enterprises and governments [4,5]. The objective is to enhance digital services and eliminate manual procedures [6]. People must establish their

* **Corresponding author: Rajeev Kumar**, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India, e-mail: rajeev.kumar@chitkara.edu.in

Vikas Khullar: Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India, e-mail: vikas.khullar@chitkara.edu.in

Isha Kansal: Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India, e-mail: isha.kansal@chitkara.edu.in

Jyoti Verma: Department of Computer Science and Engineering, Punjabi University Patiala, Patiala, Punjab, India, e-mail: jyoti.snehiverma@gmail.com

Karuna Salgotra: Department of Computer Science and Engineering, CT Institute of Engineering, Management and Technology, Punjab, India, e-mail: karuna.360@gmail.com

Gurpreet Singh Saini: School of Electronics and Electrical Engineering, Lovely Professional University, Punjab, India, e-mail: gurpreetsaini149@lpu.co.in

identification, which is private and essential, to use online digital services [7].

Artificial intelligence and computer vision researchers are actively working on image classification and recognition [8,9]. Due to two crucial factors, it has made remarkable progress: the ability to take the place of human vision. In mission-critical areas like medicine or the military, such empowered machines might lessen the effort and mistakes made by people. Machines can comprehend and analyse images using image recognition to automate tasks like image segmentation, detection, and classification [6]. Researchers who study computer vision have been piqued for a very long time by the field of optical character recognition, which involves the process of removing text from photographs. However, handling both the written and visual content is yet a challenge [3]. There are four primary processes involved in the process of image classification, and they are as follows: picture pre-processing, extraction of features, classification, and results analysis. During the image pre-processing stage, raw images should be improved so that the accuracy of extracting the features and the results of image classification can be raised to a higher level. The primary purpose of the extraction of features is to gather relevant information from an image to create different classes [6]. Deep learning-based methods are, therefore, not only a good alternative but also the current trend [10]. However, new research on convolutional neural networks (CNN) has shown excellent general visual descriptions that apply to a wide range of image identification applications, including fine-grain image categorization [1].

Since, personal documents contain sensitive information, while non-sensitive images only include non-personal information. Different statistical methods are used to categorize articles as sensitive or not. Many areas, including image processing, audio classification, object recognition, and natural language processing, use deep learning techniques [7,11]. CNN is a specific variety of multilayer perceptron that is comparable to neural networks. Since their weights and biases must be learned, they are composed of neurons. Every neuron receives a certain amount of input. After that, a dot product operation is executed, which is followed by a possible nonlinearity function [7].

A subset of deep learning called transfer learning (TL) applies previously trained models to new training datasets. A learned benchmark model is integrated with additional training knowledge to increase accuracy and decrease losses. A machine learning technique called temporal learning promises to give a quicker and more accurate response while needing less effort to gather the necessary training data and construct the model. Several TL models are known, including ResNet152V2, InceptionV3, DenseNet201, and VGG16. A residual neural network, often known as a ResNet, is a type of

artificial neural network that is characterized as an open-gated or gateless deep feedforward neural network. It is hundreds of layers deep, making it significantly more advanced than earlier neural networks. The Inception-ResNet-v2 CNN was developed using data from more than a million different photos that were taken from the ImageNet collection. The completely connected layers make use of the majority of the model's 160 million total parameters. VGG is an acronym for "Very Large Genome." The structure of the network was constructed using several layers, including a convolution layer with a size of 3×3 , a max convolution with a size of 2×2 , and a fully-connected structure as the final layer. Typically, the size requirements for the supplied image are 224-by-224-by-3 pixels [12].

The development of residual networks, often known as ResNet, resulted in the creation of a family of distinct deep neural networks that share components but have varying degrees. ResNet implements a component known as a residual learning unit to prevent neural networks from degrading over time. This type of network continually generates new outputs while also taking in incoming data. The most important advantage of utilizing this component is that it leads to higher levels of classification accuracy without increasing the overall level of complexity of the model [13]. An extensive network like Google's is Inception version 3. GoogleNet claims that Inception-v3 will be an inception model that will integrate several convolutional filters of varying sizes into a single filter. This strategy lowers the total number of variables that need to be taught, which in turn lowers the amount of computing complexity [14,15]. The Inception-Resnet-V2 network architecture, which was trained on more than a million images from the ImageNet data set, combines two networks, including Residual Connections and Inception Architecture. The architecture can categorize photos into 1,000 classes and has 164 levels in total. The network can accept images up to 299-by-299 pixels [12].

The recent uptick in the significance of AI is exerting a trickle-down effect on the progression of research into mobile robotics. Because of the explosive rise of digital technologies, computer microsystems, and network technology, mobile robots have made significant progress in a short amount of time. As a direct consequence of this, a growing number of new robotics that can perform specialized tasks have been manufactured. The field of robotics draws on a diverse array of academic topics and areas of expertise, such as video processing, sensor information systems, communications technologies, the theory of servomotors, mechanical engineering, and other related areas. This means that each image has its sampling frequency, which is determined by the image data it receives, and that it provides data on brightness adjustments of a specific size. Visual data is acquired on a pixel-by-pixel rationale by

incident cameras; this implies that the data is obtained in a manner like that of digital photography. An intelligent robot control method that can detect and sense the surroundings through a range of sensors and autonomously carry out analysis, modeling, and deliberation based on the surrounding information and the condition of the surroundings. A recognition engine enhanced with AI makes it possible to automatically classify a wide range of different kinds of documents. This article's goal is to provide essential, valuable information on two technical solutions learning intelligence and robotics – and their potential uses in the categorization of sensitive and non-sensitive documents. These technological solutions include robotics and deep learning intelligence [16].

The objectives of this study as described as follows:

- The main aim of this article is to enhance the classification of sensitive and non-sensitive documents by applying pre-trained model-based feature extraction and dense neural network based classification.
- Diverse versions of pre-trained models such as DenseNet, EfficientNet *etc.*, have been applied and analysed to identify best features from experimental dataset.
- Dense Neural Network-based classification has been applied to improve overall outcome in terms of parameters defined in Section 5.1.

The article is organized into various sections where the introduction is followed by Section 2 which consists of related study, Section 3 involves methodology, Section 4 discusses model designing for the document recognition and segregation system based on robotics, and Section 5 presents Results which are followed by Conclusion in Section 6.

2 Related study

The characterization of digitally stored data is an age-old computer vision task that gained prominence with the arrival of computers in businesses and the widespread usage of digital photos and cell phones. The important phase in the digitization process is extracting text from scanned or photographed documents to facilitate text search and simple classification. Document image analysis is also among the most popular applications of contemporary deep learning techniques [3].

In a study by Xiong *et al.* [10] DP-LinkNet, a semantic segmentation network, was presented as a means of improving the accuracy of binarization applied to damaged historical document pictures. The hybrid dilated convolutional and spatially pyramid pooling block, both of which are in the middle of the processing chain between the encoder and decoder, are

principally responsible for the performance boost. To recognize and detect identity documents in images, research [17] proposed a new, original model architecture based on an artificial convolutional neural network and a semantic segmentation approach. The intersection over union (IoU) threshold value of 0.8 was reported by the study to have an accuracy above 0.75. Investigation [18] suggested a technique that mines, recognizes, and retains the patterns from the target data and can enhance itself from the constant feedback from the result. Using a binary classifier developed on top of the most recent ANN technique inside the artificial intelligence paradigm, the suggested model defines an information loss protection mechanism.

The suggested method [6] is based on two distinct approaches in which images are categorized according to their visual attributes and text features. To determine the confidence level for the feature-based classifier, a novel methodology built on statistics and regression has been developed. It has been suggested to merge the classifier results according to their confidence score using a fuzzy-mean fusion model. The final suggested model, which achieves 100% accuracy, is built on two distinct techniques. The image is identified using the SIFT (Scale-Invariant Feature Transform) feature extractor based on its visual characteristics. We investigated several variables that affect how well CNN performs on document pictures. The high-performance improvements come from using large input images, and shear transforms during training, which achieves 90.8% accuracy on RVL-CDIP, the industry standard [19]. A tabular description of the additional comparison study is provided in Table 1.

Depending on the degree of sensitivity and the nature of the information, loss, misuse, modification, or illegal access to sensitive information may harm a person's privacy or welfare, a company's trade secrets, or even the security and international relations of a country. Consequently, this research involves the creation of a machine-learning system that categorizes the information in the images as sensitive and non-sensitive. When an image is designated as sensitive, one can notify the users that it includes sensitive information. This study could be seen as the development of a novel federated learning-based method for classifying sensitive and non-sensitive material. This might potentially be used to build different frameworks for collaborative deep-learning training for the evaluation of sensitive or non-sensitive documents.

3 Methodology

This section is describing the utilized dataset, applied algorithms, and followed methodology as shown in Figure 1, to

Table 1: A comparison study of methods, dataset types, applications, and outcome of research done by researchers

Ref	Method	Dataset type	Application	Outcome
[17]	CNN	Identity Documents	Application of a complicated document localization and classification mechanism on images of identity documents	Accuracy – 75
[14]		Private Image Dataset	Classification of identifying documents as an example of an image classification challenge	Accuracy – 98
[19]		RVL-CDIP	An investigation on the use of CNN for the categorization of document images	Accuracy – 90.8
[14]		3042 real documents	Application of a complicated document localization and classification mechanism on images of identity documents	Accuracy – 96.6
[18]	ANN	Digital Documents	A technique that uses deep learning to prevent the loss of information from digital documents with many pages	Accuracy – 89.20
[6]	Machine Learning Hybrid Models	Identity Documents	An Intelligent blended Model for the Classification of Identity Documents	Accuracy – 100
[20]	Deep Learning based Method (DOC)	20 Newsgroups and 50-class reviews	Classification of textual information with a Deep open access level	F1-Score – 92.6
[3]	Multimodal Neural Network	Small Tobacco3482 and Large RVL-CDIP	Multimodal deep Networks for text and image-based document classification	F1-Score – 69.8 Accuracy – 87.8
[21]	Hidden Markov models	Document images	Classification and comparison of documents based on layout similarity	Accuracy – 87.8 —

establish conclusions on the basis of acquired results of conducted experiments which include the below steps in sequence:

1. Dataset collection and preprocessing.
2. Discussion about ML and deep learning classification in general paragraphs.
3. Pre-discussion trained deep learning models-based feature extraction. This should explain in detail how feature extraction is done using pre-trained deep-learning models.
4. Discussion about computer vision implementation over robotic structure.

The data set was acquired from the open source image dataset of sensitive and non-sensitive controls and was used in this investigation, after that pre-processing was applied to the images in the dataset, wherein pre-processing, the original images were resized to $224 \times 224 \times 3$. Later on, during the pre-processing phase, the Image pixels were transformed into grayscale images. This meant that all of the images with dimensions of $224 \times 224 \times 3$ were scaled down to dimensions of $224 \times 224 \times 1$. After that, these groups of photos were used as input for the train and test split procedures. For training purposes, this study used 300 sensitive and non-sensitive images, and for validation purposes, 50 sensitive and non-sensitive images. There were 100 training and validation epochs for the classification of the images. Further, on training images, initially, Deep Learning and TL algorithms were applied for the classification of sensitive and non-sensitive data, and through testing, image result analysis was performed. Feature Extraction Techniques are more utilizable using pre-trained models such as VGG Net *etc.* In this work, various features were extracted from different pre-trained models. Then, extracted features were fed into machine and ensemble learning algorithms for classification into sensitive/non-sensitive images. Once our finalized model achieved the highest Accuracy, Precision, and Recall, that model was chosen and embedded over the computer vision-based robotic structure. Finally, the robot is activated for the classification of sensitive/non-sensitive documents.

4 Model design of the document recognition and segregation system based on robot

The images obtained from the image sensor are processed by the graphical navigation system, the navigation lines are retrieved, navigation decisions are taken after the

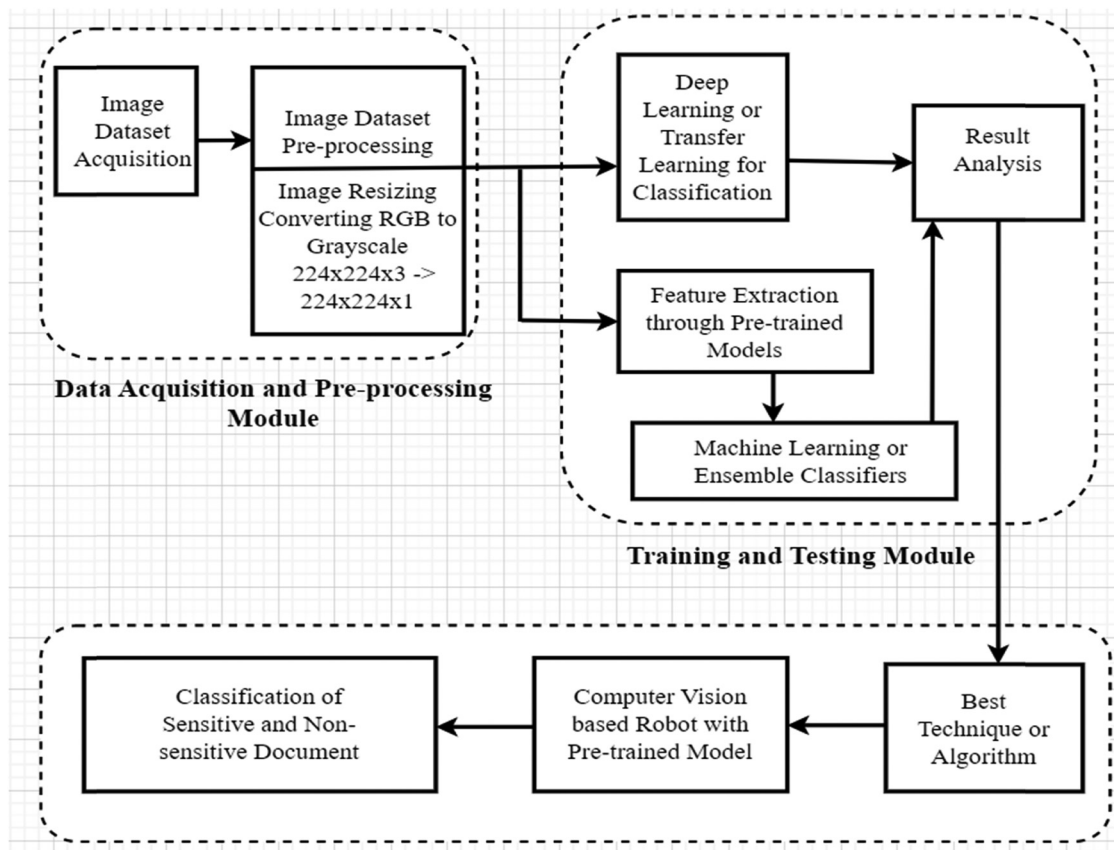


Figure 1: Methodology involving data acquisition, pre-processing module, training, and testing module.

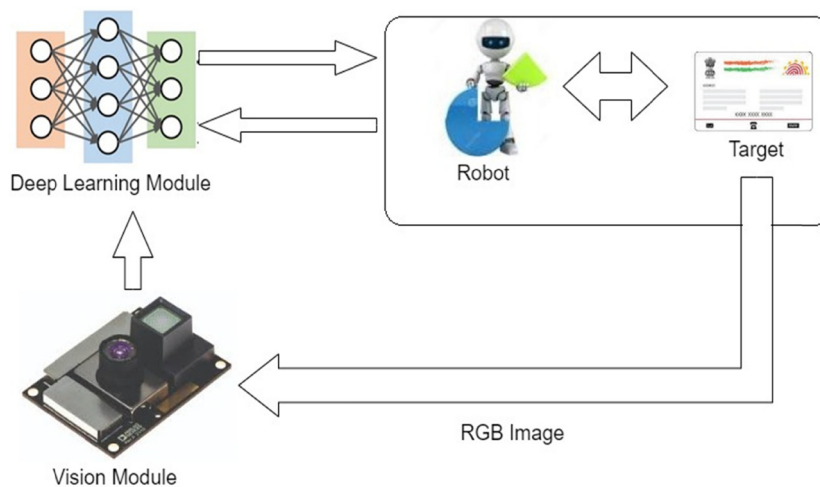


Figure 2: Model design of the document recognition and segregation system based on the robot.

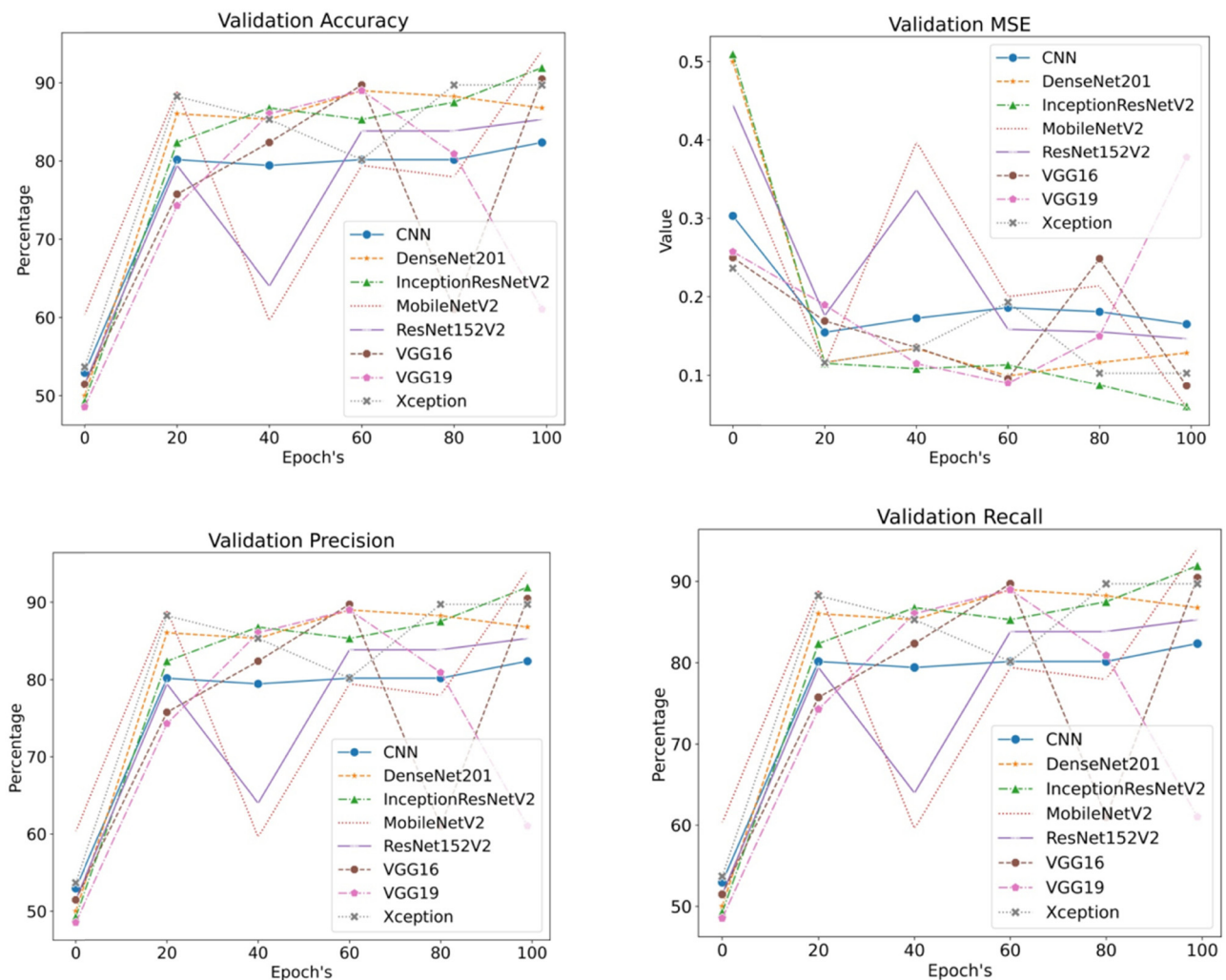
robot's navigation parameters have been calculated, and control signals are then sent to the lower computer to regulate the document segregation robot's automatic navigation. This study's visual navigation software was created using MFC and Open CV on the Microsoft Visual 2015

application development. C++ and C were used to build the algorithms, respectively. Five modules comprise the software for visual segregation: the information-seeking module, the image processing subsystem, the navigation decisions module, the data transfer control unit, and the storing information

Table 2: Performance evaluation parameters

Symbol	Formula
ACC	$\frac{\text{TrPs} + \text{TrNg}}{\text{TrPs} + \text{TrNg} + \text{FlPs} + \text{FlNg}}$
L	$-\frac{1}{m} \sum_{i=1}^m y_i \cdot \log(\hat{y}_i)$
PRE	$\frac{\text{TrPs}}{\text{TrPs} + \text{FlPs}}$
REC	$\frac{\text{TrPs}}{\text{TrPs} + \text{FlNg}}$
F1S	$2 \cdot \left(\frac{\frac{\text{TrPs}}{\text{TrPs} + \text{FlPs}} \cdot \frac{\text{TrPs}}{\text{TrPs} + \text{FlNg}}}{\frac{\text{TrPs}}{\text{TrPs} + \text{FlPs}} + \frac{\text{TrPs}}{\text{TrPs} + \text{FlNg}}} \right)$
MC	$\frac{\text{TrPs} \cdot \text{TrNg} - \text{FlPs} \cdot \text{FlNg}}{((\text{TrPs} + \text{FlPs}) \cdot (\text{TrPs} + \text{FlNg}) - (\text{TrNg} + \text{FlNg}) \cdot (\text{TrNg} + \text{FlPs}))^2} \cdot 1$
KS	$\frac{\text{ACC}}{\text{ACC} + \frac{(\text{TrPs} + \text{TrNg}) \cdot (\text{FlPs} + \text{FlNg})}{2 \cdot (\text{TrPs} \cdot \text{TrNg} - \text{FlPs} \cdot \text{FlNg})}}$

module. The information-gathering module acts as both the basis and the prelude to visual navigation. After the sensor has been adjusted, the primary objective of this element is to generate a picture that is both accurate and distortion-free [20]. During the navigation process, an Inertial Measurement Unit would determine the angular position of the lens in real time to compensate for any variation in the camera's angular position caused by the robot's vibration. The module for image processing is the most essential element of the visual guidance system. Building of image regions of interest, identification based on a deep learning approach, clustering of recognition frames, image grayscale, blending filtering, corner feature point extraction in the detection frame, and navigation line fitting are the primary responsibilities of the method for processing images. The robot's movement is governed by a navigation decision

**Figure 3:** Validation MSE, Validation Precision, Validation Recall, and Validation Accuracy for CNN 1, Layer CNN 3, Layer DenseNet, Inception, ResNetV3, and ResNet VGG.

module that, based on the navigation information acquired after image processing, controls the robot's motion.

The major task involved in Figure 2 are as follows:

Document Classification – Through the use of categorization, robots in digital workflows are now able to instantaneously recognize the type of documents being processed. The ability to make informed decisions and handle documents in a manner specific to the user's needs is enabled by rapid identification. The classification process is applied not just to individual documents but also to groups of documents, sometimes known as batches or sets. Software robots can now, for instance, comprehend when a PDF file has many documents that are contained within the same file.

Document Separation – With classification, machines are now conscious of not just the document type, but also the beginning and conclusion of documents. This knowledge enables the splitting or separation of document collections into discrete elements for processing. The PDF in the preceding example can now be separated into different documents.

Document Data Extraction – The data extraction features of Smart Capture enable bots in a Principle that guides them to extract the data from all kinds of documents and generate structured data. This data can be single items, sections, or data tables. Without the requirement for layouts, robot document recognition is now more precise and can encompass a wider range of document formats.

5 Results

In this article, deep learning-based classification was applied to the collected data set. The mentioned algorithms were trained and tested for 100 epochs. Then, further pre-trained model-based feature extraction techniques were applied, and extracted features were fed into machine learning-based classification models. The acquired results are then compared on the basis of testing accuracy, means squared error,

Table 3: Validation results comparison of deep learning classification results

Epoch's	CNN	DN201	IRV2	MNV2	RNV2	VGG16	VGG19	Xception
Accuracy								
1	52.94	50.00	49.26	60.29	51.47	51.47	48.53	53.68
20	80.15	86.03	82.35	88.97	79.41	75.74	74.26	88.24
40	79.41	85.29	86.76	59.56	63.97	82.35	86.03	85.29
60	80.15	88.97	85.29	79.41	83.82	89.71	88.97	80.15
80	80.15	88.24	87.50	77.94	83.82	61.03	80.88	89.71
100	82.35	86.76	91.91	92.12	85.29	90.44	61.03	89.71
Loss								
1	0.30	0.50	0.51	0.39	0.44	0.25	0.26	0.24
20	0.15	0.12	0.12	0.11	0.18	0.17	0.19	0.12
40	0.17	0.13	0.11	0.40	0.34	0.14	0.12	0.13
60	0.19	0.10	0.11	0.20	0.16	0.10	0.09	0.19
80	0.18	0.12	0.09	0.21	0.16	0.25	0.15	0.10
100	0.17	0.13	0.06	0.06	0.15	0.09	0.38	0.10
Precision								
1	52.94	50.00	49.26	60.29	51.47	51.47	48.53	53.68
20	80.15	86.03	82.35	88.97	79.41	75.74	74.26	88.24
40	79.41	85.29	86.76	59.56	63.97	82.35	86.03	85.29
60	80.15	88.97	85.29	79.41	83.82	89.71	88.97	80.15
80	80.15	88.24	87.50	77.94	83.82	61.03	80.88	89.71
100	82.35	86.76	91.91	92.12	85.29	90.44	61.03	89.71
Recall								
1	52.94	50.00	49.26	60.29	51.47	51.47	48.53	53.68
20	80.15	86.03	82.35	88.97	79.41	75.74	74.26	88.24
40	79.41	85.29	86.76	59.56	63.97	82.35	86.03	85.29
60	80.15	88.97	85.29	79.41	83.82	89.71	88.97	80.15
80	80.15	88.24	87.50	77.94	83.82	61.03	80.88	89.71
100	82.35	86.76	91.91	92.12	85.29	90.44	61.03	89.71

DenseNet201-DN201, InceptionResNetV2-IRV2, MobileNetV2-MNV2, and ResNet152V2-RNV2.

Table 4: Comparison of ML classification on pre-trained deep learning models based feature extraction

Feature Extraction	Classifier	ACC	PRE	REC	F1S	ZOL	MC	KS
DenseNet121	Random Forest	93	93	92	93	0.07	0.86	0.86
DenseNet169	Histogram Gradient Boosting	92	92	91	91	0.08	0.82	0.82
DenseNet201	Random Forest	90	90	89	90	0.10	0.79	0.79
DenseNet201	Histogram Gradient Boosting	91	91	90	90	0.09	0.81	0.81
EfficientNetB0	Logistic Regression	90	89	90	90	0.10	0.80	0.80
EfficientNetB1	K Nearest Neighbors	92	93	90	91	0.08	0.83	0.82
EfficientNetB1	Multi-Layered Perceptron	92	92	91	91	0.08	0.82	0.82
EfficientNetB2	Logistic Regression	91	91	90	90	0.09	0.81	0.81
EfficientNetB2	Multi-Layered Perceptron	91	91	90	90	0.09	0.81	0.81
EfficientNetB3	Random Forest	93	92	92	92	0.07	0.84	0.84
EfficientNetB4	Histogram Gradient Boosting	94	94	93	94	0.06	0.87	0.87
EfficientNetB5	Logistic Regression	96	97	95	96	0.04	0.92	0.92
EfficientNetB6	Ada Boost	91	90	91	91	0.09	0.81	0.81
EfficientNetB6	Multi-Layered Perceptron	92	91	92	91	0.08	0.83	0.83
EfficientNetB7	Extra Trees	96	96	95	95	0.04	0.90	0.90
EfficientNetV2B0	Logistic Regression	91	91	91	91	0.09	0.81	0.81
EfficientNetV2B0	Extra Trees	92	91	91	91	0.08	0.83	0.83
EfficientNetV2B1	Histogram Gradient Boosting	92	92	90	91	0.08	0.83	0.82
EfficientNetV2B2	Extra Trees	90	90	90	90	0.10	0.79	0.79
EfficientNetV2B3	Multi-Layered Perceptron	92	91	92	91	0.08	0.83	0.83
EfficientNetV2L	Extra Trees	96	95	95	95	0.04	0.91	0.91
EfficientNetV2M	Logistic Regression	95	95	94	94	0.05	0.89	0.89
EfficientNetV2M	Random Forest	94	94	93	94	0.06	0.87	0.87
EfficientNetV2M	Extra Trees	95	95	94	94	0.05	0.89	0.89
EfficientNetV2M	Multi-Layered Perceptron	95	94	95	94	0.05	0.89	0.89
EfficientNetV2S	Random Forest	96	96	95	95	0.04	0.90	0.90
EfficientNetV2S	Extra Trees	96	96	95	95	0.04	0.90	0.90
EfficientNetV2S	Histogram Gradient Boosting	96	96	95	95	0.04	0.90	0.90
InceptionResNetV2	Multi-Layered Perceptron	91	90	91	91	0.09	0.81	0.81
InceptionV3	Logistic Regression	90	90	90	90	0.10	0.79	0.79
MobileNet	Extra Trees	90	90	89	90	0.10	0.79	0.79
MobileNetV2	Histogram Gradient Boosting	92	91	92	91	0.08	0.83	0.83
NASNetLarge	Multi-Layered Perceptron	88	86	87	87	0.13	0.74	0.73
NASNetMobile	Logistic Regression	93	93	93	93	0.07	0.86	0.86
ResNet101	Multi-Layered Perceptron	90	89	90	89	0.10	0.78	0.78
ResNet101V2	K Nearest Neighbors	86	85	84	85	0.14	0.70	0.70
ResNet101V2	Histogram Gradient Boosting	86	85	84	85	0.14	0.70	0.70
ResNet152	Logistic Regression	93	93	92	93	0.07	0.86	0.86
ResNet152V2	Histogram Gradient Boosting	91	91	90	90	0.09	0.81	0.81
ResNet50	Logistic Regression	89	88	89	88	0.11	0.77	0.77
ResNet50	Random Forest	89	88	89	88	0.11	0.77	0.77
ResNet50	Extra Trees	89	88	88	88	0.11	0.76	0.76
ResNet50	Histogram Gradient Boosting	89	88	88	88	0.11	0.76	0.76
ResNet50V2	Extra Trees	88	86	87	87	0.13	0.73	0.73
VGG16	Logistic Regression	86	85	86	85	0.14	0.70	0.70
VGG19	Extra Trees	88	87	87	87	0.12	0.75	0.75
Xception	Logistic Regression	93	92	93	92	0.07	0.84	0.84

precision, recall, *etc.* The results section is further elaborated for deep learning-based classification and a hybrid approach of pre-trained model-based feature extraction and machine learning-based classification.

5.1 Evaluation parameters

There are six evaluation parameters used to compare various Learning Model's Performance including Accuracy (ACC), Loss (L) (m training samples, y denotes Actual value for i th training sample), Precision (PRE), Recall (REC), F1-Score (F1S), Zero One Loss (ZOL), Mathew's Correlation Coefficient (MC), and Kappa Statistics (KS) as described in Table 2, where the terms True Positive (TrPs), True Negative (TrNg), False Positive (FPs), and False Negative (FNg) have been used [22].

5.2 Results of deep learning classification on collected dataset

In this section, deep learning-based classification using CNN, DenseNet201, InceptionResNetV2, MobileNetV2, ResNet152V2, VGG16, VGG19, and Xception was applied. As shown in Figure 3 and Table 3, MobileNetV2 and InceptionResNetV2 have reported the highest results in terms of accuracy (91%), precision (91%), and recall (91%) with the least mean squared error (0.06). However, other algorithms CNN, VGG16, VGG19, DenseNet201, and ResNet152V2 have resulted in low in comparison to the MobileNetV2 and InceptionResNetV2.

5.3 Results of machine learning classification on pre-trained deep learning models based feature extraction

In this section, hybrid approach including feature extraction through pre-trained deep learning models and classification through machine learning models was applied and analysed. As highlighted in Table 4, different variants of EfficientNet have resulted highest in terms of accuracy (95–96%), precision (94–97%), recall (94–95%), F1-Score (94–96%) in average with least Zero One Loss (value of 0.04–0.05), highest Mathew's Coefficient (value of 0.89–0.92), and Kappa Statistics (value of 0.89–0.92). The mentioned results are acquired with the classification of different machine learning algorithms.

Results discussed in this section have shown significant improvement in classification by extracting features from given dataset images using pretrained models. Using standard deep learning algorithms, the maximum achieved accuracy was 90% whereas by applying proposed feature extraction approach accuracy raised up to 95%. Similarly, other parameters including precision, recall, F1 score, *etc.*, also improved significantly.

6 Conclusion

The study initially concentrated on locating a computer vision-based method for classifying sensitive documents from non-sensitive documents. To begin, various deep learning approaches, such as CNN and TL algorithms were implemented, to locate the benchmark parameters. To identify feature extraction from the photos, the authors have applied the majority of the available feature extraction techniques. Machine learning and ensemble learning models were used to categorize the features. In contrast to deep learning and TL, the results of model-based feature extraction and machine learning classification obtained by using pre-trained models were superior. As shown in Table 4, different variants of EfficientNet have resulted highest in terms of Accuracy, ranging from 95 to 96%; precision, 94–97%; recall, 94–95%; F1-Score, 94–96%; Least Zero One Loss ranging from, 0.04–0.05; Mathew's Coefficient, 0.89–0.92; and Kappa Statistics value of 0.89–0.92. The above-selected models proved to be the most effective and were implemented as the “brain behind vision” of a robotic structure that was designed to automate the process of separating classified material from non-sensitive documents. This proposed robotics structure could be utilized in situations in which we are tasked with locating a specific and confidential article amidst a pile of hay.

Funding information: The authors state no funding is involved.

Author contributions: IK designed the experiments and VK carried them out. RP and RK developed the model code and performed the simulations. JV prepared the manuscript with contributions from all coauthors.

Conflict of interest: The authors state no conflict of interest.

Informed consent: Informed consent was obtained from all individuals included in this study.

Ethical approval: The conducted research is not related to either human or animals use.

Data availability statement: The data sets generated during and/or analysed during the current study are available online (<https://github.com/DhilipSanjay/Detection-of-Sensitive-Data-Exposure-in-Images/tree/main/dataset>).

References

- [1] R. Sicre, A. M. Awal, and T. Furon, "Identity documents classification as an image classification problem," *Image Anal. Process. - ICIAP*, vol. 2017, pp. 602–613, 2017.
- [2] R. Kumari and S. K. Srivastava, "Machine learning: A review on binary classification," *Int. J. Comput. Appl.*, vol. 160, no. 7, pp. 11–15, 2017.
- [3] H. Blockeel, K. Kersting, S. Nijssen, and F. Železný, *Machine learning and knowledge discovery in databases*, Springer Berlin, Heidelberg, 2013.
- [4] I. Kansal and S. S. Kasana, "Minimum preserving subsampling-based fast image de-fogging," *J. Mod. Opt.*, vol. 65, no. 18, pp. 2103–2123, 2018.
- [5] J. Snehi, M. Snehi, D. Prasad, S. Simaiya, I. Kansal, and V. Baggan, "SDN-based cloud combining edge computing for IoT infrastructure," In: *Software Defined Networks: Architecture and Applications*, Wiley, Texas, 2022, pp. 497–540.
- [6] N. Khandan, "An intelligent hybrid model for identity document classification," arXiv preprint arXiv:2106.0434, 2021, doi: 10.48550/arXiv.2106.04345.
- [7] J. Snehi, A. Bhandari, M. Snehi, U. Tandon, and V. Baggan, "Global intrusion detection environments and platform for anomaly-based intrusion detection systems," in *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*, 2021, pp. 817–831.
- [8] H. Kaur, D. Koundal, and V. Kadyan, "Image fusion techniques: A survey," *Arch. Comput. Methods Eng.*, vol. 28, no. 2, pp. 4425–4447, 2021.
- [9] M. Mukhtar, M. Bilal, A. Rahdar, M. Barani, R. Arshad, T. Behl, et al., "Nanomaterials for diagnosis and treatment of brain cancer: Recent updates," *Chemosensors*, vol. 8, no. 2, p. 117, 2020.
- [10] W. Xiong, X. Jia, D. Yang, M. Ali, L. Li, and S. Wang, "DP-LinkNet: A convolutional network for historical document image binarization," *KSII Trans. Internet Inf. Syst. (TIIS)*, vol. 15, no. 1, pp. 1778–1797, 2021.
- [11] J. Snehi, A. Bhandari, M. Snehi, V. Baggan, and H. Kaur, "AIDAAS: Incident handling and remediation anomaly-based IDaaS for cloud service providers," *10th International Conference on System Modeling & Advancement in Research Trends*, vol. 1, no. 4, 2021, pp. 356–360.
- [12] M. Phankokkruad, "COVID-19 pneumonia detection in chest X-ray images using transfer learning of convolutional neural networks," *Proceedings of the 3rd International Conference on Data Science and Information Technology*, vol. 2, no. 2, 2020, pp. 147–152.
- [13] H. Li, X. Dou, C. Tao, Z. Wu, J. Chen, J. Peng, et al. Rsi-cb: A large-scale remote sensing image classification benchmark using crowdsourced data. *Sensors*, vol. 20, no. 6, pp. 1594, 2020.
- [14] A. M. Awal, N. Ghanmi, R. Sicre, and T. Furon, "Complex document classification and localization application on identity document images," *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1, no. 2, 2017, pp. 426–431.
- [15] J. Ahamed, M. Ahmed, N. Afreen, M. Ahmed, and M. Sameer, "An inception V3 approach for malware classification using machine learning and transfer learning," *SSRN Electron. J.*, vol. 4, no. 4, pp. 11–18, 2022.
- [16] N. Dominic, D. Daniel, T. W. Cenggoro, A. Budiarto, and B. Pardamean, "Transfer learning using inception-ResNet-v2 model to the augmented neuroimages data for autism spectrum disorder classification," *Commun. Math. Biol. Neurosci.*, vol. 39, pp. 1–21, 2021.
- [17] M. Kozlenko, V. Sendetskyi, O. Simkiv, N. Savchenko, and A. Bosyi, "Identity documents recognition and detection using semantic segmentation with convolutional neural network," *CEUR Workshop Proceedings*, vol. 2923, no. 1, 2021, pp. 234–242. doi: 10.5281/zenodo.5758182.
- [18] A. Guha, D. Samanta, A. Banerjee, and D. Agarwal, "A deep learning model for information loss prevention from multi-page digital documents," *IEEE Access*, vol. 9, pp. 80451–80465, 2021.
- [19] C. Tensmeyer and T. Martinez, "Analysis of convolutional neural networks for document image classification," arXiv preprint arXiv:1708.03273, 2017, doi: 10.48550/arXiv.1708.03273.
- [20] L. Shu, H. Xu, and B. Liu, "DOC: Deep open classification of text documents," *Conference on Empirical Methods in Natural Language Processing*, vol. 1, no. 2, 2017, pp. 2911–2916.
- [21] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," *J. Ind. Inf. Integr.*, vol. 6, no. 1, pp. 1–10, 2017.
- [22] I. M. De Diego, A. R. Redondo, R. R. Fernández, J. Navarro, and J. M. Moguerza, "General performance score for classification problems," *Appl. Intell.*, vol. 52, no. 2, pp. 12049–12063, 2022.