Global Journal of Information Technology: Emerging Technologies

Volume 13, Issue 2, (2023) 90-99

www.gjit.eu

Deep Learning for Automatic Classification of Identification Documents

Blerina Vika* Lecturer of Artificial Intelligence & Algorithm and Data Stucture, University of Tirana, Tirana (UT).
Elira Hoxha Executive Director at National Accounting Council of Albania, University of Tirana, Tirana (UT).

Suggested Citation:
Vika, B. & Hoxha, E. (2023). Deep Learning for Automatic Classification of Identification Documents. *Global Journal of Information Technology: Emerging Technologies*. *13*(2), 90-99.

Received from; July 28, 2022, revised from; August 20, 2022 and accepted from September 02.
Selection and peer review under responsibility of Assoc. Prof. Dr. Ezgi Pelin YILDIZ, Kafkas University, Turkey.

Abstract

This paper presents a general approach for identification documents classification using deep learning models. Our study gives an explanation of the main steps that need to be followed in order to implement a classification deep learning model. We have used convolution neural networks to extract features from raw image pixels on private datasets of identification documents. The implemented models use different techniques to preprocess the images in order to improve the classification performance on the test dataset and also techniques that can offer a better generalization of the models on the classification task. The experiments demonstrate that the training time-efficiency and accuracy of the models depends on the size, numbers of the pattern for each category and type of the image preprocessing. Various techniques of optimization have been applied to improve the model's performance and as a result we achieved the best classification accuracy of 90.4% on the test dataset.

Keywords: Deep Convolution neural networks, Image classification, Overfitting, Identification Documents

* ADDRESS FOR CORRESPONDENCE: **Blerina Vika*** Lecturer of Artificial Intelligence & Algorithm and Data Stucture, University of Tirana, Tirana (UT). E-mail: blerina.vika@feut.edu.al ·

1. Introduction

Deep learning, as a subset of machine learning, involves training artificial neural networks so that they can learn complex tasks such as image and speech recognition/classification (Taye, 2023) natural language processing, data visualization, etc. The image classification problem, including document classification, is one of the most popular tasks in computer vision (Mohsenzadega, et al. 2021). The term "deep" refers to the multiple layers that compose the neural network, which can range from a few to hundreds or even thousands of layers.

Nowadays, a lot of effort is put into creating systems that can automatically analyze or classify document images. These types of systems are an important tool for different organization that help them to: improve the efficiency of the classification process by reducing the time and resources needed to do it manually; enhance the accuracy and increase the reliability, by reducing the risk of human error; reduce the cost by decreasing the need for manual labor; improve the customer services.

The classification process involves grouping different images of identification documents into known categories. The learning process in this type of problem is supervised learning. In supervised learning, the model is trained on a labeled dataset using an optimization algorithm to minimize the error or loss function value between the predicted output and the actual output. A variety of deep learning architectures exists for image classification. In our work we have used convolution neural networks (CNNs). This architecture was first introduced by (LeCun et al. 1998) in a presentation on LeNET in 1998. But, regardless of the good advantages that this architecture offers, only in recent years it has become popular as the cost of multi-core computers has been significantly decreasing.

The convolutional neural networks are widely used for image classification tasks because they can effectively capture the spatial information present in images (Taye, 2023). The architecture contains one to few convolution layers. These layers are used to extract features from the input images through convolving a learnable set of kernels (also known as filters) with the input images and the result data are transformed by the activation function to produce the convolution layer's output (Mohsenzadega, et al. 2021). By applying different filters with different weights to the input data, the convolutional layer can learn to recognize various patterns and features in the input data. The combination of the patterns and features can give a more complex representation of the input data.

In this paper, we present the general framework applied to implement a system for identification documents classification. We also highlight which are the elements that need to be part of the framework that can enhance the efficiency of the classification model. The analysis used in our work for the classification model construction is a combination of literature review with the experimental results. The overall work presented in this paper can contribute to the deep learning document classification literature.

The rest of the paper is organized as follows: in the next section we describe some related work concerning the areas of knowledge that are relevant in identification document categorization. Then we explain the important steps of a general classification framework that need to be accomplished in order to build a robust system. Then the following section represents the description and the outcomes of the experiments performed in a private dataset with different types of identification documents categories. Finally, based on the framework and related to the experiment results, we present some conclusions and future work that will increase and complete the identification documents classification system.

Related Work

In recent years a lot of effort has been put into creating efficient image classification methods. There are different methods that can be used for image classification problems. Two are the main categories of methods: traditional or classical methods and deep learning methods.

Classic techniques use local descriptors to try to find similarities between images (Rana, 2021). Some of the traditional methods include: support vector machine (SVM) (Wang, et al., 2021), (Foody, Mathur, 2004). text & color recognition (Stoecker etc., 2011). Deep learning can be seen as one of the methods that have dramatically improved the state-of-the-art in speech/images recognition, visual object recognition, object detection and many other domains (LeCun, et al. 2015). Thus, the focus of our work lies in the CNN based identification documents classification.

Kölsch et al. in their work (Kölsch, et al. 2017) have used a two-stage approach which uses feature extraction from deep neural networks and efficient training using an Extreme Learning Machine (ELM) for a document classification approach that trains in real-time, where ELM is a single-layer feed-forward neural network. Also, other authors that have worked in deep learning for document classification are Mohsenzadegan et al. that have developed, validated, and benchmarked a deep CNN model with a selection of the most relevant recent document classification models. The results show that injecting different artifacts during the training process significantly improves the model's sensitivity.

Clearly, we can denote that the document image classification is not a simple task, most of the time it needs to build complex deep learning models that can better understand the different aspects of the classification process, starting from features extraction to the selection of the correct category label.

The Identification Documents Classification Framework

The framework used for identification documents (ID) classification (Figure. 1) follows the basic steps, which are: data collection, pre-processing the raw input images, defining the network architecture (CNN model), the model implementation, testing the trained model and checking the performance of the classification model.

Figure 1. The Phases Framework for Identification Documents Classification System

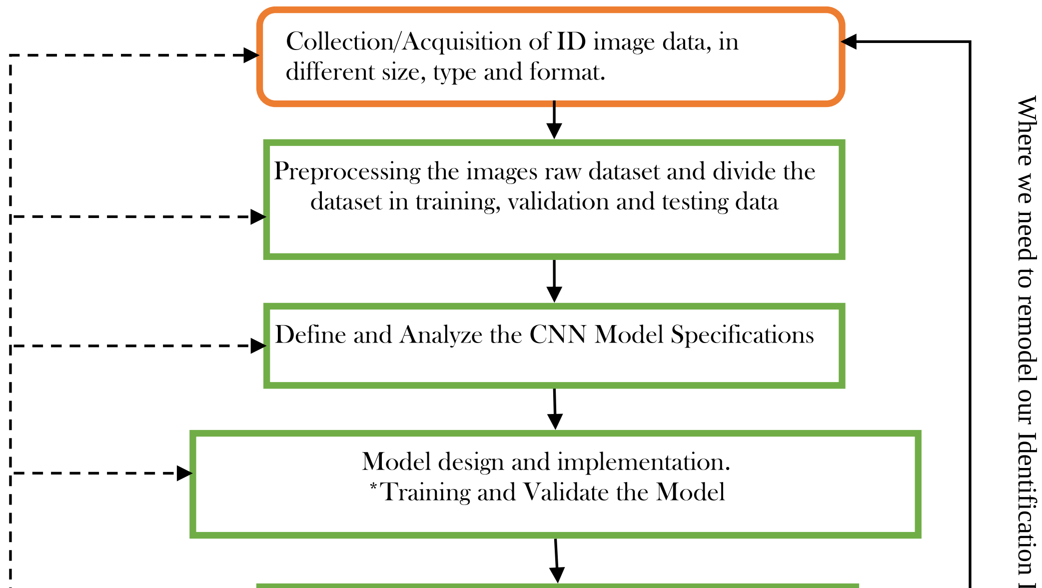


Table 1. gives a brief explanation for each of the phases. It illustrates the general framework applied for the implementation of ID images classifications system.

Table 1. Description of the Framework Phases

Phases	Description	Part of
Data Collection	The raw images input data used in our model are grouped in different categories. For each category we have a suitable number of identification documents that can be used for training and testing the model. The input data are in different formats (such as pdf, jpeg, and png), different types (with colors or in grayscale) and different sizes (in pixels).	Data Preparation
Data Preprocessing	This phase is considered an important step in preparing data for image classification problems. It can help to improve the accuracy and performance of the deep learning model by making the input raw data more consistent and easier to work with. Some of the preprocessing techniques that can be apply to the images data are: *Resizing* (all the images are turned in a common size to standardize the input images); *Normalization* (scaling the pixel values in the images to a common range); *Data Augmentation* (is the process of flipping, rotating, zooming, cropping or adding noise to the images by creating a more diverse dataset.)	
Dataset Partitioning	In order to train and evaluate the model, in our work the input data is divided in three subsets: training, validation and testing sets. The training set is used to train the classification model in training phase. We use the validation set to prevent the model overfitting by utilizing the early stopping method [10]. Then, the testing set is used to evaluate the performance of the learned CNN model.	
Model Specification	The model used for the identification documents classification is the Convolution Neural Network. The basic elements that need to be specify for the CNN model are the number of convolution layers and pooling layers, the activation function, dropout layers, the connection between the neurons in different layers of the neural network, loss function and the optimizer (the algorithm used to update the weights of the network during training to minimize the loss function).	Network Architecture
Model training & validation	This is the step in which the classification model learns from the training data in supervised learning. All the input data in the training set are passed different times (also called epochs). Using the validation set during the training phase we tend to avoid model overfitting (early stopping procedure).	
Model Testing	After training the model, it is needed to test how the model has been learned.	

Evaluate the performance	The results of the classification model in the testing phase are analyzed and checked with the goal performance. In case that the model accuracy is admissible, then the trained and tested classification model is used to categorize new identification document images that come in the system. Otherwise, it is required to return back in the framework trying to change the model specification and type of preprocessing.	Model Selection and Re-training
Retraining the Classification model	Retraining the identification documents classification model can help to improve its accuracy and performance for a specific task or dataset and can also allow it to adapt to changing data distributions or recognize new classes of images.	

Identification Documents Classification System

Datasets

In this paper is used a unique dataset with 3886 images which are distributed irregularly across 8 classes (categories) of different identification documents. The number of images for classes varies from a minimum of around 180 images, to a maximum of 870 images per class. The input data are in different formats (such as pdf, jpeg, png, etc.), different types (with colors, or in grayscale), and different sizes (in pixels). Also, it should be mentioned that the collected images have different qualities, they are taken by the individuals themselves.

In the pre-processing phase, all the input data are transformed in the same format. Different datasets are created in order to evaluate the learning performance of the model and time required to train the model. The datasets differ by the size in pixels (128x128; 256x256; 512x512), the type of images (RGB, Grayscale) and the color of the pixels (white or black) used to extend the images to fit the required size.

It is known that the learning performance of the model depends on the size of the datasets. The deep learning models have a better classification performance when they are trained with large datasets and small datasets may trigger over-fitting (Althnian et al. 2021). The number of identification documents images per category, in quantities, are not large enough. A good approach that the literature review suggests being applied in this case is data augmentation.

Data augmentation (DA) is another way we can reduce overfitting on models, where we increase the amount of training data using information only in our training data (Wang & Perez, 2017), (Pednekar & Slater, 2019). Clearly, the application of data augmentation plays a significant role while training the deep learning model (Shorten & Khoshgoftaar, 2019). The types of augmentation that suit our problem are rotation and zooming. We have applied these two techniques for each dataset. We have selected the *online* data augmentation way to apply the DA techniques. In this form the rotation and zooming are randomly applied to the training data in execution time before fitting the model and it is not needed to save any data on the memory.

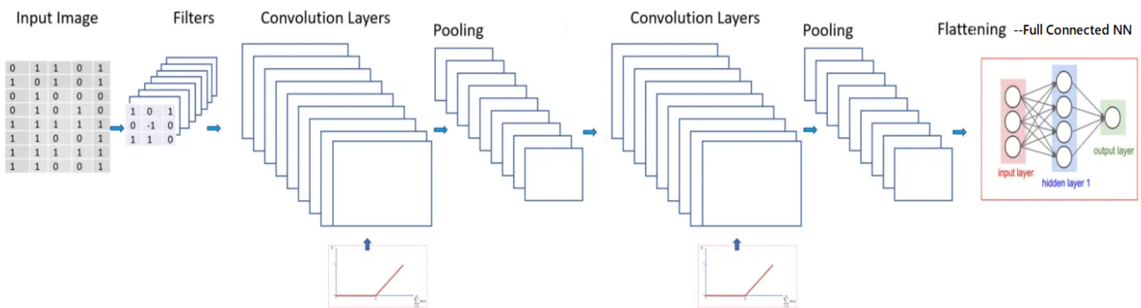
The data partitioning utilized is the random division of the dataset in three different sets: training, validation, and testing. The partitioning applied is respectively in the ratio of 80:10:10 percent. The validation set is used to assess the performance of the model during training, by evaluating the model's ability to generalize to new and unseen images data.

System Architecture

The architecture presented in this paper (Figure 2.) for identification documents is not based on document features that require a high resolution, such as optical character recognition. Instead, the classification uniquely depends on the structure and layout of the input documents images to categorize them.

The two-dimensional CNN is the model used as a basic classification tool for the proposed architecture. The CNN model consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers. In our model, we have used two convolutional layers, two pooling layers, and a fully connected feed-forward neural network with a hidden layer.

Figure 2. The CNN architecture for IDs Classification System



Both the convolution layers use as activation function the ReLU (Rectified LinearUnit) function. This non-linear activation function introduces non-linearity into the model, which is important for capturing complex patterns in the data. It allows the network to learn complex representations of the input data, while also reducing the risk of the vanishing gradient problem (Rasamoelina, et al. 2015). Each of the convolution layers is followed by a max pooling layer that operates by partitioning the data of the previous layer into a set of non-overlapping rectangular regions called pooling regions. The output of the max pooling layer for each pooling region is the maximum value found within that region.

The fully connected network used in this architecture is a feed forward neural network (FFN) with a hidden layer. The hidden layer consists of 128 neurons fully connected, and 8 neurons in the output layer, which correspond with the number of identification documents categories.

Prevent Model Overfitting

Creating a classification model that has more than 30 million parameters and at the same time having a dataset not large enough with low quality images, requires the usage of techniques that can reduce overfitting. Below are described the elements included in our model, in order to avoid overfitting:

- *Data Augmentation*

  Data Augmentation is one of the effective tools used to train a deep learning model and reduce the overfitting risk (Wang & Perez, 2017), (Shorten & Khoshgoftaar, 2019). For the dataset we apply two types of DA, the rotation and zooming transformations of the images. The transformation types are done in an online data augmentation way. In this form the rotation and zooming are randomly applied to the training data in execution time before fitting the model and do not require to save any data on the memory.

- *Dropout layer*

  Another easy way used to prevent overfitting is adding dropout layers inside the model architecture (Srivastava et al. 2014), (Lim, 2021). A dropout layer randomly drops out a specified fraction of the neurons in the previous layer during each training iteration. The purpose of dropout is to force the network to learn more robust and generalizable features by preventing it from relying too heavily on any particular set of neurons. By randomly dropping out neurons during training, the network is forced to distribute the workload more evenly across all the neurons, and this helps to prevent overfitting.

  The proposed architecture has two dropout layers, the first one was set after the convolution layers with a dropping fraction at 20 percent and the second one is used after the hidden layer of the FFN with a 30 percent dropping ratio.

- *Validation Set*

  The validation set is used in the “early stopping” strategy by stopping training before the performance stops optimization (Rice et al. 2020). The role of the validation set in machine learning is to provide a way to evaluate the model's performance on new, unseen data and to prevent overfitting during the training process.

Experiments

Experiments were conducted using a dataset with 3886 images which are distributed irregularly across 8 classes (categories) of different identification documents. Three different input images sizes were used: 128x128, 256x256, 512x512. Each of them is in RGB or Grayscale type. The test machine used for this paper is a PC with the following settings: Intel(R) Core (TM) i7-5600U CPU @ 2.60GHz 2.59 GHz with 16 GB RAM. The code is written on Python 3.9 with Tensorflow package.

We used an equal learning rate for all layers, and it was initialized at 0.01. The number of epochs changes in different tested models in order to take into consideration how vary the model classification performance. An epoch is defined as one complete pass through the entire training dataset during the model training phase. The optimizer algorithm used for all the tested models is Adam (Adaptive Moment Estimation). It is an important tool for optimizing the weights of a neural network during training and improving the accuracy of the model in image classification tasks. A lot of studies have demonstrated the superiority of this algorithm as optimizer for deep CNN models (Bera & Shrivastava, 2020), (Kumar, et al. 2022).

Results Obtained and Discussion

To find which is the best ID classification model to use for further classification of new identification documents images we have trained different configurations of our proposed architecture. The evaluation of the best model is based on the accuracy that is achieved in the testing set.

In Table 2. we have listed some of our relevant results. Based on this outcome we can denote that the model that uses a dataset in color type (RGB) independently on the images size have a better accuracy on out-of-sample classification.

Table 2. The Experimental Results of Different Training ID Classification Models

Datasets Details		No. Epochs	Training Acc (in %)	Testing Acc (in %)	Training Time
128x128	RGB	30	89.9	89.7	01:29:21
128x128	Grayscale	30	85.7	79.9	01:17:10
128X128	RGB	40	93.2	90.4	02:13:29
128X128	RGB	50	95.6	90.2	02:47:20
256x256	RGB	30	86.8	83.5	03:50:22
256x256	Grayscale	30	79.8	77.9	04:59:14
512x512	RGB	30	82.9	80.2	04:58:20

As expected, the increase of the image size is reflected in increased training time. But we cannot say the same thing for testing accuracy which for the smallest size is better than that of bigger size. The grayscale color type gives the worst classification performance of the ID classification system.

The best trained model with the high testing accuracy is the one that uses the dataset with the images size 128x128 on RGB color type and is trained with a total of 40 epochs. Increasing the number of training epochs can raise the training accuracy but the performance on out-of-sample is reduced, such as the model trained with the dataset parameters 128x128, RGB in 50 epochs. In this case the model achieves the best training accuracy, 95.6, but testing accuracy is lower compared to the testing accuracy for the best model.

Conclusion and Future Work

We have proposed a new deep learning model for identification documents classification. First, in our paper, we give a general framework with eight steps that need to be followed for constructing a robust ID classification model. Secondly, we have described the implementation of the proposed framework and highlighted the integration of some tools and techniques that improve the performance of the ID classification model.

Different models were progressively trained and tested until we reached and created our best model to be used. The experiments demonstrate that the training time-performance and accuracy of the models depends on the size, numbers of the pattern for each category and type of the image preprocessing. Various techniques of optimization have been applied to improve the model performance and as a result the model achieved the best classification accuracy of 90.4% on the test dataset. The overall work presented in this paper can contribute to the deep learning document classification literature.

An interesting future dimension for our ID classification model is the integration of incremental learning approaches when retrain is needed. Another future work for the model upgrade is to use multi thread GPU processing in order to reduce the configuration time of the ID classification model.

References

[1] M. Taye, "Theoretical Understanding of Convolutional Neural Network: Concepts, Architectures, Applications, Future Directions," *Computation,* vol. 11, p. 52, 2023.

[2] Mohsenzadega K. et al., "A Deep-Learning Based Visual Sensing Concept for a Robust Classification of Document Images under Real-World Hard Conditions," *Sensors,* vol. 21, no. 20, p. 6763, 2021.

[3] Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P., "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 11, no. 86, p. 2278–2324, 1998.

[4] Òscar Lorente and Ian Riera and Aditya Rana, "Image Classification with Classic and Deep Learning Techniques," *Computer Vision and Pattern Recognition,* 2021.

[5] Pin Wang, En Fan, Peng Wang,, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognition Letters,* vol. 141, pp. 61-67, 2021.

[6] Giles M. Foody, Ajay Mathur,, "Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification," *Remote Sensing of Environment,* vol. 93, no. 1-2, pp. 107-117, 2004.

[7] W. V. Stoecker, M. Wronkiewiecz, R. Chowdhury et al, "Detection of granularity in dermoscopy images of malignant," *Computerized,* vol. 35, no. 2, p. 144–147, 2011.

[8] LeCun, Y., Bengio, Y. & Hinton, G., "Deep learning.," *Nature,* vol. 521, p. 436–444, 2015.

[9] A. Kölsch, M. Z. Afzal, M. Ebbecke and M. Liwicki, "Real-Time Document Image Classification Using Deep CNN and Extreme Learning Machines,," in *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan,, 2017.

[10] J. D. Domingo, R. M. Aparicio and L. M. G. Rodrigo, "Cross Validation Voting for Improving CNN Classification in Grocery Products," *IEEE Access,* vol. 10, pp. 20913-20925, 2022.

[11] Althnian, A.; AlSaeed, D.;Al-Baity, H.; Samha, A.; Dris, A.B.;Alzakari, N.; Abou Elwafa, A.;Kurdi, H, "Impact of Dataset Size on Classification Performance: An Empirical Evaluation in the Medical Domain," *Applied Sciences,* vol. 11, no. 796, 2021.

[12] Wang J.; Perez L.;, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," *Computer Vision and Pattern Recognition,* 2017.

[13] Gu SH. Pednekar M.; & Slater R.;, ""Improve Image Classification Using Data Augmentation and Neural Networks," *SMU Data Science Review:,* vol. 2, no. 2, 2019.

Vika, B. & Hoxha, E. (2023). Deep Learning for Automatic Classification of Identification Documents. *Global Journal of Information Technology: Emerging Technologies*. *13*(2), 90-99.

[14] Shorten, C., Khoshgoftaar, T.M., "A survey on Image Data Augmentation for Deep Learning," *Big Data,* vol. 6, p. 60, 2019.

[15] A. D. Rasamoelina, F. Adjailia and P. Sinčák, "A Review of Activation Function for Artificial Neural Network," *IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI),* pp. 281-286, 2020.

[16] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R., "Dropout: a simple way to prevent neural networks from overfitting.," *The journal of machine learning research,* vol. 15, no. 1, pp. 1929-1958, 2014.

[17] H. I. Lim, "A study on dropout techniques to reduce overfitting in deep neural networks.," *In Advanced Multimedia and Ubiquitous Engineering: MUE-FutureTech 2020,* pp. 133-139, 2021.

[18] Y. X., "An Overview of Overfitting and its Solutions," *Journal of Physics: Conference Series,* vol. 1168, no. 2.

[19] Rice L.; Wong E.: Kolter Z.;, "Overfitting in adversarially robust deep learning," in *37th International Conference on Machine Learning*, 2020.

[20] Bera S. and Shrivastava V. K., "Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification," *International Journal of Remote Sensing,* vol. 41, no. 7, pp. 2664-2683, 2020.

[21] A. Kumar, V. Bali and S. Pandey, "Improving the efficiency of Plant-Leaf Disease detection using Convolutional Neural Network optimizer-Adam Algorithm,," in *12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2022.