Personal Documents Classification using a Hybrid Framework at a Mobile Insurance Company: A Case Study

Raissa Barcellos and Rodrigo Salvador
ADDLabs, Computing Institute - Federal Fluminense University, Brazil

Keywords: Document Classification, Convolutional Neural Networks.

Abstract: In the information age, coupled with the full range and speed of data, the ease of access to new disruptive technologies brings the relevant problem of document classification. Identifying and categorizing documents is still a very challenging initiative addressed in the literature. This paper analyzes the construction of a document classification hybrid framework in a real business context. The research is based on a case study addressing the construction of a hybrid framework that uses text and image in document classification and how this framework can be useful in an authentic context of a mobile insurance company. Excellent accuracy and precision results were found in the use of both approaches, even considering a possible fraudulent circumstance. From these results we can conclude that using the hybrid framework, using the visual approach as a filter — which is more efficient in verifying the authenticity of documents— and consolidating the results with the textual approach, is a convincing option for deployment in the company in question.

1 INTRODUCTION

Nowadays, with the era of Big Data, over-data exposes the challenging problem of recognizing and categorizing documents. In many scenarios, document classification is a sophisticated task that confronts several areas of research. This task usually consists of a feature extraction step and an automatic classification step. The primary purpose of this type of classification is to assign a document to one or more categories (Hassan et al., 2015).

Documents generally have distinct visual styles. Today, one of the challenges of document image analysis is the fact that within each type of document, there is a vast range of visual variability (Harley et al., 2015). Another critical issue is that documents of different categories regularly display considerable visual similarities. From a visual style standpoint, some erroneous recoveries under these circumstances may be justifiable, but generally, the task of document image analysis is to classify documents despite intra-class variability and class similarity (Harley et al., 2015).

Also, there are several important issues - which have serious consequences — in today's society that can be well resolved using document classification (Xiao and Cho, 2016). Such as the problem of identity fraud. These threats can be characterized as small frauds or even organized crimes. Several approaches have been proposed to classify documents - such as supervised classification, unsupervised classification, and semi-supervised classification of documents (Hassan et al., 2015). More recently, it has become more common to use neural networks, which jointly perform feature extraction and classification, for document classification. In the following subsections, we will cover these different approaches more extensively (Xiao and Cho, 2016).

The main objective of this paper is to conduct a case study, regarding the construction of a framework for personal documents classification submitted by users, in a real business context. Our case study refers to a mobile insurance company — it covers phones for loss, theft and accidental damage with mobile phone — that massively identifies the correctness of the clients' documents manually, through a call center. This particular company intends to invest in an aggressive marketing strategy, but the number of service orders — cellphone theft notification — will increase so much, that it would be necessary to double the number of resources in the call center to meet new demand. In this context, the ideal would be to invest in an automatic document identification application, so that when opening a service order through the company portal, the customer is instructed to submit their documents via the internet. This service should be able to identify personal documents with as little

Barcellos, R. and Salvador, R.
Personal Documents Classification using a Hybrid Framework at a Mobile Insurance Company: A Case Study.
DOI: 10.5220/0009340204900497
In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020) - Volume 1*, pages 490-497
ISBN: 978-989-758-423-7

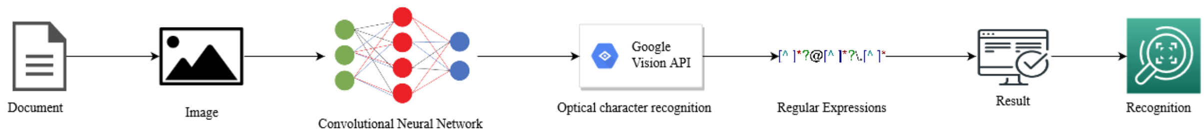

Figure 1: Developed Hybrid Framework.

human intervention as possible.

To conduct this case study, we developed a hybrid application that explores text and image to classify personal documents, as we can see in Figure 1, targeting the real business scenario. By running this case study, we hope to be able to explore the option of deploying a document classifier hybrid framework as an alternative for the company, in question to be able to invest in its new marketing strategy without having to double the call center staff. For the development and testing of the framework, we have a sample of the database provided by the company.

This work is organized as follows: Section 2 presents the theoretical background, regarding to the techniques used in our work and that conceptualize the importance of this work. Section 3 presents a literature review. Section 4 presents our hybrid framework. Section 5 describes a discussion about our hybrid framework. Finally, Section 6 show our final conclusion and future work.

2 BACKGROUND

2.1 Unsupervised Document Learning

The Machine Learning community widely studies unsupervised learning (Su et al., 2019). In unsupervised learning, there is a set of N observations *(x1, x2, ..., xN)* of a random p-vector X having joint density *Pr(X)*.The goal is to directly infer the properties of this probability density without the help of a supervisor or teacher providing correct answers or degree-of-error for each observation (Friedman et al., 2001). The dimension of X is sometimes much higher than in supervised learning, and the properties of interest are often more complicated than simple location estimates (Friedman et al., 2001).

The most common unsupervised learning task is the clustering — detecting potentially useful input sample clusters. A fixed group of text is clustered into groups that have similar content. The similarity between documents is calculated with the associative coefficients. Document clustering mainly used Hierarchical clustering algorithms (Hassan et al., 2015).

2.2 Supervised Document Classification

In supervised learning, there is a set of N variables that might be denoted as inputs, which are measured or preset. These have some influence on one or more outputs. The goal is to use the inputs to predict the values of outputs. This activity is called supervised learning (Friedman et al., 2001).

In this kind of learning, approaches such as pattern recognition are used to classify a document —– examples of classifiers such as neural networks, support vector machines, and genetic programming. Multiple classifiers can be used in combination with supervised learning, but classifier accuracy can be improved using a small set of documents (Hassan et al., 2015). An example of a supervised learning technique, that is widely used today in document recognition, is the use of Convolutional Neural Networks.

2.2.1 Convolutional Neural Networks in Document Classification

Deep learning is revolutionizing the already rapidly developing field of computer vision. The convolutional neural network (CNN) is a state-of-the-art deep learning tool that learns high level features directly from a huge dataset of labeled images (Khan et al., 2018). A deep convolutional neural network consists of convolutional layers followed by fully connected layers with normalization and/or grouping performed between the layers. There are a wide variety of network architectures and layer parameters are learned from trainning data.

In traditional Artificial Neural Networks, the relationship between input and output units is determined by matrix multiplication. In Convolutional Neural Networks, convolution is used instead of general matrix multiplication, reducing the number of weights and parameters in the network (Revanasiddappa and Harish, 2019).

Besides, it minimizes network complexity by reducing memory size and improving performance. Learning algorithms bypass the resource extraction procedure due to the direct consideration of network entry. Convolution also helps to learn a multi-level representation (Revanasiddappa and Harish, 2019). Image representations are computed by taking the output of the fully connected intermediate layers or by pooling the output of the last convolutional layer. Intermediate layer extraction produces intermediate-level generic representations that can be used for various recognition and recognition tasks a wide range of data, such as document classification (Sicre et al., 2017). Convolutional Neural Networks have traditionally been implemented for image recognition, and several techniques have already been implemented to improve this architecture (Sicre et al., 2017).

Semi-supervised learning algorithms have widely been studied since the 1990s mostly thanks to Information Access and Natural Language Processing applications. In these applications unlabeled data are significantly easier to come by than labeled examples which generally require expert knowledge for correct and consistent annotation. The underlying assumption of semi-supervised learning algorithms is, if two points are close then they should be labeled similarly, resulting in that the search of a decision boundary should takeplace in low-density regions. This assumption does not imply that classes are formed from single compact clusters, only that objects from two distinct classes are not likely to be in the same cluster (Krithara et al., 2008).

2.3 Optical Character Recognition

Optical Character Recognition (OCR) is a technology that analyzes image characters and transforms them into the text format used on a computer (Lee et al., 2019). OCR is a complex problem because of the variety of languages, fonts and styles in which text can be written, and the complex rules of languages etc. Hence, techniques from different disciplines of Computer Science — as image processing, pattern classification and natural language processing — are employed to address different challenges (Islam et al., 2017).

Based on the type of input, the OCR systems can be categorized as handwriting recognition and machine printed character recognition. The former is relatively simpler problem because characters are usually of uniform dimensions, and the positions of characters on the page can be predicted (Islam et al., 2017). In this work we only utilized machine printed character recognition.

Web services like Google Cloud Vision and Amazon Rekognition are OCR solutions that implement machine learning algorithms as a solution to image recognition (Pathak et al., 2019). Google Cloud Vision was launched on December 2, 2015 moreover has been growing and developing constantly. Cloud Vision is a proprietary API that can prove application development for image analysis, using as multiple REST APIs. The API has features for image recognition, including identification of landmarks, optical character recognition, face detection, and logo detection (Pathak et al., 2019).

3 LITERATURE REVIEW

Various approaches for document image classification have been proposed over the years. Generally, document image classification approaches are divided into two major groups, structure/layout based, and content based. This section provides an overview of some important works which have been reported in reference to structure or content based document classification (Afzal et al., 2015).

Khanalni *et al.* (Khanalni and Gharehchopogh, 2018) used a hybrid of the IWO algorithm — based on chaos theory — with a Naive Bayes classifier for classifying text documents. The authors used the algorithm IWO to select essential features and Naive Bayes for trainning-based document classification and tests. The results indicated that the proposed model is more accurate compared to Naive Bayes. Also, the error rate factor indicates that proposed model errors with Feature Selection are smaller in a comparison of the proposed model with other models, the results indicated that the model proposed by the authors is more accurate due to the use of Feature Selection which is capable of better exploit the resource space.

In another work, Audebert *et al.* (Audebert et al., 2019) attacked the problem of document classification based only on an image of a digitized document, and the authors performed classification using visual and textual attributes using the Tesseract OCR Engine and FastText — a library for text classification and representation learning. The authors introduced an end-to-end learned multimodal deep network that jointly learns text and image capabilities and performs the final classification based on a different representation of the document. The proposal showed consistent gains in both small and large datasets. So, there is significant interest in the hybrid image/text approach even when clear text is not available for document image classification.

Popereshnyak *et al.* (Popereshnyak et al., 2018) chosen Convolutional Neural Network to solve the problem of identifying personal documents, using the ReLU activation function. As a result, image classification performance has been tested, and an accuracy of about 85% has been achieved. It has been experimentally determined that a neural network can recognize multiple classes at once in one image.

This option allows more improvement of the neural network, increasing the number of classes, and increasing recognition accuracy. Kolsch *et al.* (Kölsch et al., 2017) addressed the problem of real-time training for document image classification. The authors present a document classification approach that trains one millisecond per image, ie, in real-time. The approach is divided into two stages — the first stage uses resource extraction from deep neural networks, and the second stage uses Extreme Learning Machines (ELMs) for classification.

According to Tensmeyer *et al.* (Tensmeyer and Martinez, 2017), convolutional neural networks are very efficient models for document image classification. However, many of these approaches are based on architectures designed to classify natural images, which differ from document images. In this paper, the authors question whether this custom is appropriate and conduct an empirical study to find out which aspects of convolutional neural networks most affect document imaging performance. In general, the application of shear transformations during trainning and the use of large input images lead to the most significant gains in performance. trainning and testing at various scales also improve the specifically for smaller trainning sets. Also, Batch Normalization is a useful alternative to Dropout in datasets with great visual variety. A trained convolutional neural network is also examined, and the authors report evidence that it is learning characteristics layout intermediaries. Neurons fire based on the type of layout component (graphic, text, handwriting, noise) and tend to shoot at specific places in the image.

The contribution of this present work is to explore the development and implementation of a hybrid document classifier framework, in a real scenario of a mobile insurance company, using a not artificial sample of documents. We also performed a framework implementation evaluation, using a company dataset containing actual data of varying quality.

4 DEVELOPED HYBRID FRAMEWORK

In order to build a document classification hybrid framework, in a real business scenario, we used some technologies, and we combined two approaches. The visual approach uses Convolutional Neural Networks to identify the documents to be classified automatically, through the image only. This step is essential to identify the document class, since this will provide relevant information about the layout of data to be extracted, and about the security measures present on that document that will allow detecting document forgery. In the textual approach, we used the Google Vision API to extract text from images, along with the use of regular expressions, identifying common words present in documents to be classified.

4.1 Dataset

The document dataset, for trainning and testing, contains images of scanned documents, collected from the mobile insurance company's private database. In total, the database has over 30.587 documents, hand-labeled with tags. The three categories are "identity document/driver's license", "invoice" and "occurrence report".

4.2 Convolutional Neural Networks

The convolutional neural networks we used in this experiment are models that map input images $x \in \mathbb{R}^{H \times W \times D}$ into the probability vectors $y \in \mathbb{R}^{C}$, where D is the input image depth, W is a filter which is applied to a window of H words to produce a new feature, and C is the number of classes. Each layer performs a transformation with learnable parameters followed by non-linear operation(s): $x_l = g_l(W_l \star x_{l-1} + b_l)$ where $1 \leq l \leq L$ is the layer index, x_0 is the input image, W_l, b_l are learnable parameters, $\star$ is either matrix multiplication for fully connected layers or 2D convolution for convolution layers, and g_l is a layerspecific non-linearity, constituted by Rectified Linear Units— $ReLU(x) = max(0,x)$, and optionally max-pooling, batch normalization, or dropout. Deep convolutional neural networks with ReLUs train much faster than their counterparts with Hyperbolic Tangent. The output of the last layer is a input to a sigmoid function — the softmax function is a more generalized logistic activation function which is used for multiclass classification. Similar approaches were used in (Tensmeyer and Martinez, 2017) (Krizhevsky et al., 2012). For each type of document, we trained a different convolutional neural network, using Keras[1] — a high-level neural networks API, written in Python and capable of running on top of TensorFlow.

4.2.1 Identity Document

1. **Trainning Details**
   For this trainning, we utilized 3 datasets: a trainning dataset, a testing dataset and a validation dataset. We use trainning data to train the algorithm and then create the predictive model. Only

[1]https://keras.io/

IDs is in the training data. We used validation data to evaluate the model during trainning. We used the testing data to validate the performance of the already trained model, ie, we presented the model with data that he did not see during training to ensure that he can make predictions. The first dataset is composed of 10.595 images, splited into two paths 5.257 (IDs) and 5.339 (others). The second dataset is composed of 607 images. The third dataset is composed of 139 images.

2. **Trainning Hyperparameters**
   We used 32 features for a 2D array and defined our array as 3x3 format. So, we converted all our 256x256 pixel images into a 3D array. We applied the max-pooling layer to reduce the size of the feature map, added four convolution layers, applying max-pooling layers between them. We used Data Augmentation technique to generate samples by transforming trainning data, with the target of improving the accuracy and robustness of the model (Fawzi et al., 2016). We applied Flatten to convert the 2D data structure to a 1D structure, ie, an array. The rectifier activation function (relu) is used, and then a sigmoid activation function to obtain the odds of each image containing an identification document or not. To compile the network, we used the Adam optimizer — first-order algorithm for gradient-based optimization of objective functions based on an adapted estimate of low order moments. We used a log loss function with binary cross-entropy because it works well with sigmoid functions. We used 5000 steps in our trainning set for 4 epochs. We chose 2000 validation steps for validation images.

3. **Accuracy**
   We achieved an accuracy of 97% for the trainning set and 91% for the test set.

4.2.2 Occurrence Report

1. **Trainning Details**
   Like for Identity Document, we utilized 3 datasets: a trainning dataset, a testing dataset, and a validation dataset. The first dataset is composed of 7.373 images, splited into two paths 3.447 (IDs) and 3.926 (others). The second dataset is composed of 938 images. The third dataset is composed of 638 images. Only Occurrence Reports is in the training data.

2. **Trainning Hyperparameters**
   We used 32 features for a 2D array and defined our array as 3x3 format. So, we converted all our 256x256 pixel images into a 3D array. We applied the max-pooling layer to reduce the size of the feature map, added four convolution layers, applying max-pooling layers between them. Like for Identity Document, we used Data Augmentation technique. We also applied a technique called Batch Normalization to increase trainning speed. Batch Normalization works by first linearly scaling and shifting each neuron's activations to have zero mean and unit variance (Tensmeyer and Martinez, 2017). We inserted Batch Normalization after each convolution layer. We applied Flatten to convert the 2D data structure to a 1D structure. Like for Identity Document, we used the rectifier activation function (relu), and then a sigmoid activation function. To compile the network, we used the Adam optimizer and a log loss function with binary cross-entropy. We used 3000 steps in our trainning set for 4 epochs. We chose 2000 validation steps for validation images.

3. **Accuracy**
   We achieved an accuracy of 96% for the trainning set and 89% for the test set.

4.2.3 Invoice

1. **Trainning Details**
   We also utilized 3 datasets: a trainning dataset, a testing dataset, and a validation dataset. The first dataset is composed of 9648 images, splited into two paths 5459 and 4189 (others). The second dataset is composed of 3369 images. The third dataset is composed of 5537 images. Only Invoices in is trainning data.

2. **Trainning Hyperparameters**
   We also used 32 features for a 2D array and defined our array as 3x3 format, and we converted all our 384x384 pixel images into a 3D array. Like for Occurrence Report, we applied the max-pooling layer to reduce the size of the feature map, added four convolution layers, applying max-pooling layers between them. We also used Data Augmentation technique and we applied Batch Normalization to increase trainning speed. We applied Flatten and we used the rectifier activation function (relu), and the sigmoid activation function. To compile the network, we also used the Adam optimizer and a log loss function with binary cross-entropy. We used 3000 steps in our trainning set for 4 epochs. We chose 2000 validation steps for validation images.

3. **Accuracy**
   We achieved an accuracy of 94% for the trainning set and 82% for the test set.

Table 1: Documents type, keywords and regular expressions used.

Document Type	Keywords	Regular expressions
Occurrence report	"police", "report", "record"	re.search("((police)+(.+))+((report | occurrence | record)+(.))+(.+)?", line)
Invoice	"danfe", "invoice", "tax coupon","nf", "taxes", "tax", "tributes", "nfce"	re.search("((danfe | invoice | tax coupon | nf)+(.+)) +(tributes | tax | taxes | nfce)+(.+)?", line)
Identity Document/ Driver's license	"secretary of", "safety", "public", "identity", "doc source", "id card", "director", "national traffic department", "permission"	re.search("(((((secretary of)+ (safety)+(public)+(.+)) | ((identity)+(.+)))+ ((doc source | doc . source | doc. source) +(.+))+((id card | director)+(.+))+(.+)?)) | (((national traffic department) +(.+))+ (permission)+ (.)+(cat)+(.+)?)", line)

4.3 Google Vision API and Regular Expressions

We used the Google Vision API text extraction feature to implement regular expressions, refining the text extracted from document images. For the construction of regular expressions, we listed the most common keywords in all selected document types, according to the Table 1. The classification of the documents happens through the results of the regular expressions.

4.3.1 Accuracy

We performed a test with a total of 198 real documents, where: (i) 75 documents corresponded to the occurence report type, (ii) 51 documents corresponded to the identification document or driver's license type and (iii) 72 documents corresponded to the invoice type. We measured the accuracy of the correct classification for each of these types of documents. Table 2 presents the accuracy measurements obtained.

Table 2: Accuracy measurements obtained.

Document Type	Accuracy
Occurrence report	97,5%
Invoice	94,8%
Identity Document/ Driver's license	87,5%

5 HYBRID FRAMEWORK EVALUATION

As a hybrid framework evaluation methodology, we compute the precision metric for the same test dataset in both approaches with 198 documents, as we can see in Table 3. The visual approach — using Convolutional Neural Networks — works learning the procedures that it needs to follow through images, any kind of anomaly that comes through is going to be detected and can be classified as a potencial for fraud that need to be checked out. So, the textual approach — using Google Vision API and Regular Expressions — after 198 test cases, we do not compute precision errors, there is no false-positive occurrence information for this approach, so the precision metric is considerated 100%. However, the textual approach works with text extraction and regular expressions, does not consider fraudulent situations, considering the image format and characteristics.

Given that accuracy indicates the overall performance of the approach, that is, of all ratings, how many have the approach correctly rated, by using our hybrid framework, the company will be able to automate much of the manual document classification process today. Given our results, in visual approach, we were able to accurately exclude between 82% and 92% handwritten documents considered fraudulent when attempting to reproduce either type of document. By aggregating the textual approach, we are able to guarantee between 87.5% and 97.5% accuracy in the textual approach and arguably satisfactory accuracy for scanned documents. Given that precision indicates, among all the positive class ratings the ap-

Table 3: Precision measurements obtained for document type/approach.

Document Type	First step precision (CNN)	Second step precision (OCR+ Regex)	Number of approved documents	Number of inconclusive documents
Occurrence report	97%	100%	192	6
Invoice	100%	100%	198	0
Identity Document/ Driver's license	99%	100%	196	2

proach has taken, how many are correct, by using our hybrid framework, in visual approach, we were able to precisely classify between 97% and 100% documents. By aggregating the textual approach, we are able to precisely classify 100% documents.

Therefore, with the result of this evaluation, we can conclude that the high classification accuracy, provided by the hybrid framework, gives remarkable confidence to the model's ability to classify documents correctly. As for accuracy, we can conclude that by evaluating the error in the classes equally — true positives and negatives — we have a high accuracy of document classification, this factor is also a good general indication of the excellent performance of the hybrid framework. For our scenario, a high measure of precision will be more beneficial than a high measure of accuracy. For, considering that one purpose of building the hybrid framework would be to reduce as much human interference as possible in document identification, the precision measure reports whether the framework accuses the document of a particular type, but is not. With a very high level of precision, as presented, the framework will be able to quickly meet the demand of the mobile insurance company by efficiently automating the work previously done by the call center industry manually.

We compared our accuracy results with some related works, given the union between image and text to classify documents. In the study (Audebert et al., 2019), the authors obtained about 90.6% accuracy in the RVL-CDIP dataset and between 68% and 98% accuracy in the Tobacco3482 dataset, both contained in documents such as emails, letters, questionnaires, and presentations. That is, the datasets did not contain personal documents. In the work (Popereshnyak et al., 2018), the authors conduct training with personal documents such as passports and driver's licenses, reaching an accuracy of around 85%, through only one CNN to classify all types of documents together.

6 CONCLUSIONS

The problem of document classification still consists of several types of research in the academic field. The search for an efficient and effective approach, which can identify various types of documents with the best yet, is extensive, although it is addressed by many areas today. In this paper, we conduct a case study, in a real business scenario, where a mobile insurance company needs a solution that automatically classifies documents, with the goal of leverage the investments in other areas of the business, such as marketing, without any increase in call center resources —– industry that manually classifies documents. For this case study, we developed a hybrid document classifier framework that explores text and image documents using technologies such as Optical character recognition and Convolutional Neural Networks. We use a real database, used today in production, for the construction and testing of the framework. This framework has two approaches: (i) the visual approach explores the document format and the image itself through supervised machine learning, and (ii) the textual approach explores only the text itself after its extraction — it is a fact that the textual approach does not consider handwritten/digitized texts.

For the visual approach, we built Convolutional Neural Networks to classify each type of document. In this approach, we train hyperparameters by applying various techniques such as Data Augmentation and Batch Normalization. Already for the textual approach, we use an Optical Character Recognition solution for text extraction and build regular expressions through the most recurring terms between documents.

After building the framework, we could observe the results of some metrics like accuracy and precision. Given the results, the conclusions we have obtained is that the best way to use our hybrid framework is to ensure that the visual approach works as a filter, because the visual approach exploits the image, avoiding fraudulent attempts — fundamental in our scenario.

Since the visual approach already maintains a high level of accuracy and precision, not reaching 100%, but working with visual and not semantic characteristics, the textual approach acts as a consolidator, addressing textual characteristics and ensuring 100% precision in the classification of the documents. In this study case, a high measure of precision will be more beneficial than a high measure of accuracy. This factor demonstrates the efficient contribution of this paper because the mobile insurance company in question, the focus of our case study, will be able to invest in an aggressive marketing strategy without having to double the number of call center resources to meet the new needs. Adding the visual approach prevents fraudulent attempts — fundamental in our mobile insurance company scenario.

As a limitation of this work, we point out that other technologies could be implemented to increase the accuracy of the textual approach, such as Natural Language Processing. We will consider this limitation as future work. Another future work will be collecting actual fraudulent data to experiment using our hybrid framework.

REFERENCES

Afzal, M. Z., Capobianco, S., Malik, M. I., Marinai, S., Breuel, T. M., Dengel, A., and Liwicki, M. (2015). Deepdocclassifier: Document classification with deep convolutional neural network. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1111–1115. IEEE.

Audebert, N., Herold, C., Slimani, K., and Vidal, C. (2019). Multimodal deep networks for text and image-based document classification. *arXiv preprint arXiv:1907.06370*.

Fawzi, A., Samulowitz, H., Turaga, D., and Frossard, P. (2016). Adaptive data augmentation for image classification. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 3688–3692. Ieee.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Harley, A. W., Ufkes, A., and Derpanis, K. G. (2015). Evaluation of deep convolutional nets for document image classification and retrieval. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 991–995. IEEE.

Hassan, H., YehiaDahab, M., Bahnassy, K., and Idrees, A. M. (2015). Arabic documents classification method a step towards efficient documents summarization. *International Journal on Recent and Innovation Trends in Computing and Communication*, 3(1):351–359.

Islam, N., Islam, Z., and Noor, N. (2017). A survey on optical character recognition system. *arXiv preprint arXiv:1710.05703*.

Khan, M. J., Yousaf, A., Abbas, A., and Khurshid, K. (2018). Deep learning for automated forgery detection in hyperspectral document images. *Journal of Electronic Imaging*, 27(5):053001.

Khanalni, S. and Gharehchopogh, F. S. (2018). A new approach for text documents classification with invasive weed optimization and naive bayes classifier. *Journal of Advances in Computer Engineering and Technology*, 4(3):31–40.

Kölsch, A., Afzal, M. Z., Ebbecke, M., and Liwicki, M. (2017). Real-time document image classification using deep cnn and extreme learning machines. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, volume 1, pages 1318–1323. IEEE.

Krithara, A., Amini, M. R., Renders, J.-M., and Goutte, C. (2008). Semi-supervised document classification with a mislabeling error model. In *European Conference on Information Retrieval*, pages 370–381. Springer.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Lee, Y., Song, J., and Won, Y. (2019). Improving personal information detection using ocr feature recognition rate. *The Journal of Supercomputing*, 75(4):1941–1952.

Pathak, A., Ruhela, A., Saroha, A. K., and Bhardwaj, A. (2019). Examining robustness of google vision api based on the performance on noisy images.

Popereshnyak, S., Suprun, O., Suprun, O., and Wieckowski, T. (2018). Personal documents identification system development using neural network. In *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, volume 1, pages 129–134. IEEE.

Revanasiddappa, M. and Harish, B. (2019). A novel text representation model to categorize text documents using convolution neural network.

Sicre, R., Awal, A. M., and Furon, T. (2017). Identity documents classification as an image classification problem. In *International Conference on Image Analysis and Processing*, pages 602–613. Springer.

Su, Y., Li, W., Nie, W., Song, D., and Liu, A.-A. (2019). Unsupervised feature learning with graph embedding for view-based 3d model retrieval. *IEEE Access*, 7:95285–95296.

Tensmeyer, C. and Martinez, T. (2017). Analysis of convolutional neural networks for document image classification. In *14th IAPR International Conference on Document Analysis and Recognition, ICDAR 2017, Kyoto, Japan, November 9-15, 2017*, pages 388–393.

Xiao, Y. and Cho, K. (2016). Efficient character-level document classification by combining convolution and recurrent layers. *arXiv preprint arXiv:1602.00367*.