ISSN 2278-3091
Volume 10, No.2, March - April 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse141022021.pdf
https://doi.org/10.30534/ijatcse/2021/141022021

An Improved Document Image Classification using Deep Transfer Learning and Feature Reduction

Aissam JADLI[1]*, Mustapha HAIN[1], Anouar HASBAOUI[2]
[1]SISM Team, I2SI2E Laboratory, ENSAM Casablanca, University Hassan II, Morocco
[2]Department of Communication and management, FST Mohammadia, University Hassan II, Morocco
* Corresponding author's Email: aissam.jadli-etu@etu.univh2c.ma

ABSTRACT

Electronic Document Management is an essential workflow within every successful ERP implementation. The integration of these documents in their respective pipelines (e.g. OCR, data extraction) inside the ERP system for processing usually requires a previous classification step to improve the success rate. Unfortunately, due to the variation in type, size, and layout of business documents (i.e. invoices, checks, delivery forms), their classification is a challenging computer task and may need an additional data for model training. This paper investigates the Transfer Learning paradigm using different pre-trained deep models to extract useful features from scanned document images. In fact, the machine learning classifiers, such as Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Gaussian Naive Bayes (GNB) process the extracted features for classification. The authors compared the constructed models performances based on various metrics. To overcome the over-fitting issue and dataset imbalance, we run a cross-validation procedure at different folds sizes (4, 6, and 8) to assess the models' generalization ability. We also inspected the effect of dimensionality reduction techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) on the overall performances and execution time. We found that the best classification rate is 97.83% achieved by combining LR, LDA, and the DenseNet121 deep model. Despite the small used dataset (546 images), this excellent performance encourages the integration of this approach in an ERP system as a separate module for document pre-processing for ERP users.

Keywords: ERP, Document image classification, Deep feature selection, Transfer learning.

1. INTRODUCTION

Organizations collect a considerable volume of documents every day concerning different workflows internally and externally; these documents can emanate from multiple organizations and have various layouts and designs[1]. These documents may exist as electronic or paper forms, and they also can be computer-generated or handwritten documents[2]. The primary used technique to extract valuable information from these documents is Optical Character Recognition (OCR). In this character recognition technique, computers differentiate between handwritten or machine-printed characters with different approaches, including deep learning-based methods[3]. The extracted information greatly accelerates the indexing, and significantly improves the search speed, which are essential features for exploring business documents[4]. In order to perform OCR in the best possible conditions, the scanned document images may need a previous enhancement step[5] by considering various parameters (e.g. brightness, chromaticity) and transformations (e.g. scaling, resize).

For this reason, the prior classification of documents can be handy to improve character recognition accuracy[6] (i.e. handwritten documents usually need a different OCR pipeline than printed ones). This classification step can be a monotonous and burdensome if done manually, especially with the presence of a significant volume of documents observed in multinational companies.

This paper explores, through a general framework, the use of transfer learning approach for corporate document classification, by combining extracting features from document images, and by using different pre-trained deep learning models with multiple machine learning classifiers for extracted features vectors.

The structure of this paper is as follows: the second section surveys the related works and similar techniques; in the third section, we report the materials and methods used for this experiment, including the dataset structure. In conclusion, we review and discuss the experiment results and present our research perspectives.

2. RELATED WORKS

The prediction accuracy of machine learning models mostly relies on the amount and the diversity of data used during training. The amount of required data is usually proportional to the number of considered parameters in the model and ultimately to the task's complexity. Dataset size and variance can be decisive factors in a classifier's performance due to the restricted availability of human-labeled training data.

In real situation, apart from data in the target domain, we can include relevant data in a separate domain to extend our prior knowledge about the target future data[7]. For instance, we adopted deep learning models pre-trained on ImageNet[8] dataset (holding 1.5M images distributed on 1000 classes) for our documents classification experience. Transfer learning addresses such cross-domain learning problems by extracting valuable features from data in a linked domain and transferring them to handle the target domain's tasks.

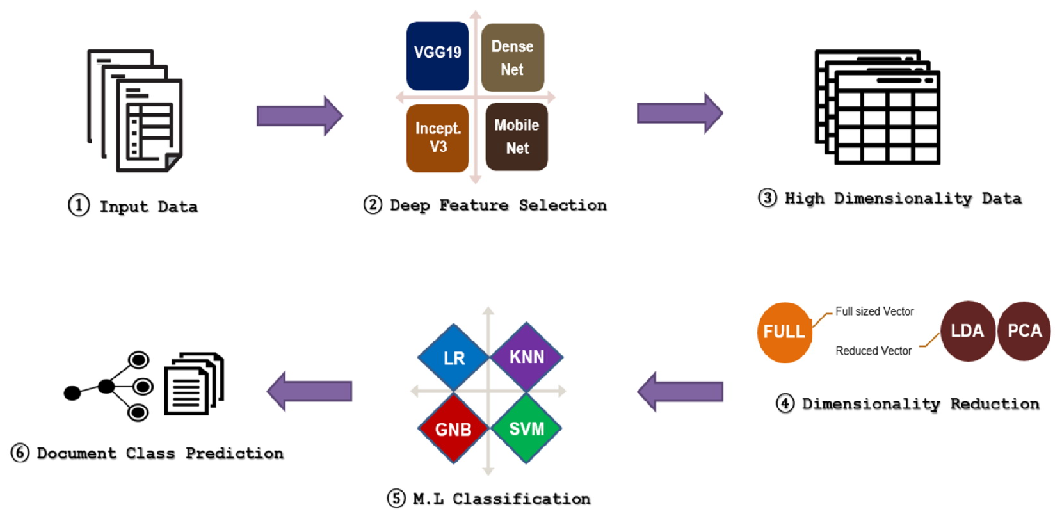

Figure 1 : Proposed Framework for Automatic Document Classification

Hubber-Fliflet et al.[9] explored using Transfer learning in an image analytics context toward legal document classification for review purposes. The proposed applications included image classification, image clustering, and object detection.

A. Kolsch et al.[10] proposed a two-stage approach for document image classification. The first stage trains a deep network that works as a feature extractor, and in the second stage, Extreme Learning Machines (ELMs) are used for classification with a final accuracy of 83.24% on the Tobacco-3482 dataset.

Liu Yang et al.[11]proposed a technique to determine whether a given source domain is useful in transferring knowledge to a target domain, they determined the amount of knowledge that should be shifted from the source domain to the target domain.

J. Zhang et al.[12] proposed a transfer learning approach that can improve classification accuracy by more than 10%, even when the connection between the auxiliary and the target tasks is not apparent.

Y. Alsabahi et al.[13], to detect lung cancer, used the Inception V3 model's weight trained on the ImageNet dataset for classification process in the Digital Radiography (DR).

C. Kang et al.[14] proposed a classification method based on multi-layer network and transfer learning that has been developed for Synthetic Aperture Radar (SAR) images.

3. MATERIALS AND METHODS

3.1 Description

This paper suggests a general framework for document image classification in which we plan to use feature selection, dimensionality reduction, and machine learning classification (Fig. 1). The proposed method for document classification is based on feature selection using pre-trained deep convolutional neural networks, dimensionality reduction using PCA and LDA, and machine learning classification using algorithms such as LR, NB, KNN, and SVM.

3.2. Dataset

The authors collected the data from a sponsoring company's archive. The dataset is composed of firm documents originated from various organizations and relating to multiple transactions. Selected samples ~~belonged~~ to different documents classes and were labelled manually into four categories:

– **Electronic Invoices (EI)**
– **Handwritten Invoices (HI)**
– **Checks (CH)**
– **Receipts (RT).**

The distribution of the dataset's samples over selected classes is summarized in Table 1.

Table 1 : Classes of documents and samples count

Document Class	Samples Count
E.I	219
HI	129
CH	80
RT	128
Total	**546**

To generalize the experiment results in different situations, the dataset images are from various angles, distances, rotations, and other lighting conditions (e.g. shadows, darkness) using a smartphone camera. Fig. 2 presents samples of different classes.

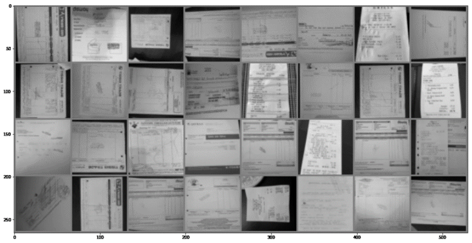

Figure 2 : Sample Documents from Dataset

This dataset can be considered as small in CNNs standards (usually, if the training process has started from scratch, CNNs may need thousands of images before producing satisfying results). However, pre-trained models reduce the need for large datasets while maintaining high performances due to transfer learning and prior knowledge exploitation.

4. PROPOSED FRAMEWORK

4.1. Deep Feature Selection techniques

Convolutional neural networks have confirmed their excellent performance compared to other classifiers in image classification tasks[15]. Even though they are not invariant to rotation and geometric distortions[16], pre-trained CNN models trained on massive image datasets (e.g. ImageNet) can select a deep feature vector invariant to rotation and form changes[17].

In this paper, the models examined were chosen because they are publicly available, free to use, and easy to modify. To find the optimal deep learning model as a feature selector for our document classification problem, we compared four deep convolutional neural networks. Table 2 provides the concrete structural parameters of each model.

Table 2. Structural parameters of deep models

Deep Model	# of Params	Depth	* FC Layer
VGG19	143,667,240	26	4096
InceptionV3	23,851,784	159	2048
DenseNetV2	20,242,984	201	1024
MobileNetV2	4,253,864	88	1280

*Fully Connected Layer

4.1.1. VGG19 Model

VGGNet[18] is a deep convolutional neural network invented by VGG (Visual Geometry Group) from the University of Oxford and was the 1st runner-up of the ILSVRC (ImageNet Large Scale Visual Recognition Competition) 2014 in the classification task and the winner of the localization task.

The network is characterized by its simplicity (Fig. 3). It uses only 3×3 convolutional layers on top of each other to increase the depth, followed by a max-pooling step to reduce volume size. Two fully connected layers, each with 4,096 nodes, are then followed by a softmax-activated layer.

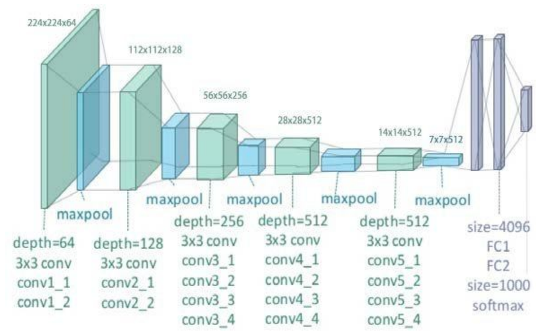

Figure 3 : VGG19 Architecture [32]

4.1.2 InceptionV3 Model

Inception V3 is a pre-trained deep learning model for image classification into 1000 classes. This model was trained on the ImageNet dataset and has achieved an error rate of 3.5%, and becomes the first Runner Up for image classification in ImageNet Large Scale Visual Recognition Competition (ILSVRC) 2015. Fig. 4 presents InceptionV3architecture.

In our case, we will stop before the last layer and recover a vector of 2048 relevant features per image, which will be processed in the next step by the ML classifiers.

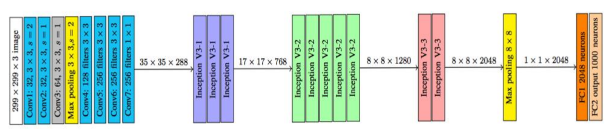

Figure 4: Outline of InceptionV3 Architecture [33]

4.1.3 DenseNet121 Model

DenseNet[19] (i.e. DenseNet-201) is a customized convolutional neural network trained on the ImageNet database. The network is 201 layers deep and can perform image classification of 1000 classes (Fig. 5)

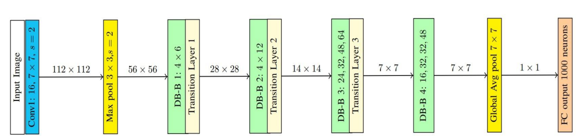

Figure 5 : DenseNet Architecture Overview [33]

The resulting model (available publicly to use) has learned rich features from database images and can be used in various other classification tasks[20].

4.1.4 MobileNetV2 Model

MobileNet-V2 is a convolutional neural network architecture that aims to run very efficiently on mobile devices. Fig. 6 presents an outline of its architecture.

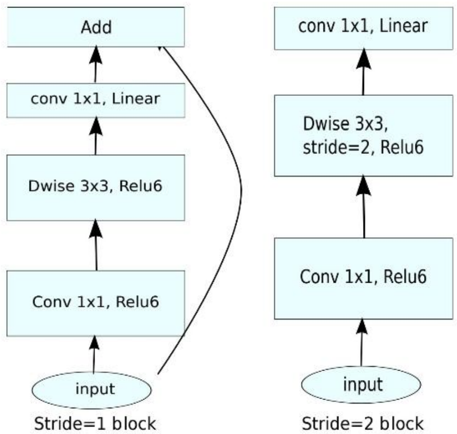

Figure 6 : MobileNetV2 Architecture [21]

It can be used either as a basic image classifier, as a feature extractor that is part of a more extensive neural network, or in combination with other classifiers.

4.2. Dimensionality Reduction

Dimensionality reduction techniques can significantly reduce the training time[22] while maintaining almost similar performances. This involves reducing or transforming the features vector into a lower-dimensional space while retaining the most relevant information/features. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are widely used for this task.

4.2.1. Principal Components Analysis

PCA is a simple data reduction technique in pattern recognition, signal processing, and bioengineering[23]. It is widely used as an unsupervised technique for dimensionality reduction[24]. It performs dimensionality reduction by embedding the data into a linear subspace of lower dimensionality. The basic idea behind PCA is to reduce the dimensionality of a dataset while retaining as much as possible the variation in the original variables[25].

4.2.2. Linear Discriminant Analysis

LDA is a pattern recognition method[26]that provides a supervised classification model based on the combination of variables.It highly predicts the category or class to which a given document belongs. The basic theory behind LDA is the classification of the dependent by dividing an n-dimensional descriptor space into two or more classes or categories, these are separated by a hyperplane defined by a linear discriminant function.

4.3. Machine Learning classifiers

4.3.1. Gaussian Naive Bayes

GNB is one of the simplest yet effective methods of multi-class classification based on the Bayesian Rule. Given a document instance to be classified, a vector x = $(x_1,\dots,x_n)$ represents some n features (independent variables), it assigns to this instance the probabilities $p(C_k \mid x_1,\dots, x_n)$, for each of K possible outcomes or classes C_k. Using Bayes' theorem, the conditional probability can be decomposed as Eq. (1)

$$\rho(C_k \mid x) = \frac{\rho(x \mid Ck) \times \rho(Ck)}{\rho(x)} \qquad (1)$$

The naive Bayes classifier combines the naive Bayes probability model with a decision rule. One common rule is to pick the most probable hypothesis; this is known as the Maximum a Posteriori or MAP decision rule. The corresponding classifier, a Bayes classifier, is the function that assigns a class label y = C_k for some k as follows in Eq. (2)

$$y =_{k \in \{1,\dots,K\}} \rho(C_k) \qquad (2)$$

4.3.2. Support Vector Machine

SVM is a non-probabilistic classification method introduced by V. Vapnik. It constructs a hyperplane in high-dimensional feature space by empirical risk minimization. It is a binary classification algorithm. The traditional way of performing multi-class classification with SVM is combining several binary ”one-against-all” or ”one-against-one” SVM classifiers[27]. We have k (by the number of classes) similar ”one-against-all” optimization tasks. Eq. (3)

$$\begin{gathered}\min_{w,b,\varepsilon} \frac{1}{2} w^T w + C \sum_{i=1}^{I} \varepsilon_i \\ w^T \varphi(x_i) + b \geq 1 - E_i \ \ if\ y_i = m; \\ w^T \varphi(x_i) + b \leq -1 + E_i \ \ if\ y_i \neq m; \\ \varepsilon_i \geq 0\end{gathered} \qquad (3)$$

where yi– class of xi. φ– kernel function. C– penalty parameter. Thus (4) derives the class of a document

$$c_{svm} = c\epsilon C\ ((w^c)^T\ \varphi(x) + b^c) \ (4)$$

4.3.3. K-Nearest Neighbors

KNN is a straightforward and easy-to-implement classification method. Despite its simplicity, itcontinues to reasonably perform well for large training sets. It relies on the most basic assumption underlying all predictions: observations with similar characteristics will tend to have similar results. Given a point x that we wish to classify into one of the K classes, we find the k observed data points that are nearest to x. The classification rule is to assign x to the population that has the most observed data points out of the k-nearest neighbors.

4.3.4. Logistic Regression

LR is a widely used classification algorithm in the industry. Compared with simple algorithms such as decision trees and Naive Bayes classification, logistic regression has higher accuracy[28]. It is compared with algorithms with high classification accuracy, such as support vector machine and neural network. Its training speed of is faster[29]. Because of its simplicity and efficiency, logistic regression still attracts wide attention in the scientific community. Many researchers believe that in many big data competitions some algorithms with better classification effects have the bottleneck of training speed, contrary to logistic regression that remains one of the most efficient and comprehensive evaluation algorithms. The logistic regression of multi-objective classification uses the Eq. (5) to calculate the probability of sample xi belonging to category C_i :

$$\rho(C_i | x) = \frac{e^{w_i^T x + w_{0i}}}{\sum_{j=1}^{K} e^{w_i^T x + w_{0j}}} (5)$$

The weight matrix and the bias vector are the parameters of this model. These parameters can be obtained by minimizing a loss function. This function can be defined as the Eq. (6):

$$l(\theta = w, w_0, D) = - \sum_{i=1}^{|D|} \log p\left(y^{(i)} \middle| x^{(i)}\right) (6)$$

5. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments were conducted using a Python-based tool developed especially for this study. It uses different Python scientific and open-source libraries for data processing (e.g. Pandas, NumPy), and classification algorithms (e.g. Keras, sci-kit learn). All used classifiers were tested with their default parameters (no special tuning or optimization).

5.1. Model Construction

The first experiment used all possible combinations of deep models, dimensionality reduction, and classification algorithms on our datasetto find, based on the accuracy (i.e. the percentage of instances classified correctly by the classifiers), the best possible combination for document image classification. To validate the constructed model, the dataset was split between a training set and a testing set:

- 2/3 for the training set
- 1/3 for the testing set

Table 2 shows the experiment results based on accuracy and training time.

Table 2. Performance Comparison of constructed models

Model	Accuracy (%)	Time (s)
DENSENET121-LDA-LR	97.83	13.96
DENSENET121-LDA-KNN	97.77	13.94
VGG19-LDA-KNN	97.4	6.56
VGG19-LDA-SVC	97.25	6.57
VGG19-LDA-LR	96.82	6.59
INCEPTIONV3-LDA-SVC	96.43	13.84
INCEPTIONV3-LDA-GNB	96.3	13.70
DENSENET121-FULL-SVC	96.12	13.37
MOBILENETV2-LDA-KNN	95.70	16.7
MOBILENETV2-LDA-SVC	95.45	15.71
INCEPTIONV3-LDA-KNN	95.39	13.84
DENSENET121-FULL-LR	95.27	25.52
VGG19-LDA-GNB	94.85	6.56
MOBILENETV2-LDA-GNB	94.54	16.71
DENSENET121-PCA-LR	94.41	10.62
DENSENET121-PCA-SVC	94.18	10.09
INCEPTIONV3-LDA-LR	93.70	13.86
DENSENET121-LDA-SVC	93.64	13.90
MOBILENETV2-LDA-LR	92.85	16.73
VGG19-FULL-LR	92.83	13.79
MOBILENETV2-PCA-LR	92.72	12.96
MOBILENETV2-FULL-LR	92.37	14.97
DENSENET121-LDA-GNB	92.30	13.94
VGG19-PCA-LR	92.15	5.26
INCEPTIONV3-FULL-LR	91.77	26.61
DENSENET121-FULL-KNN	91.40	3.70
INCEPTIONV3-PCA-LR	91.40	11.47
DENSENET121-PCA-KNN	91.29	9.96
MOBILENETV2-PCA-SVC	91.22	12.99
VGG19-PCA-SVC	90.62	5.15
MOBILENETV2-FULL-SVC	90.37	15.46
INCEPTIONV3-FULL-KNN	89.51	3.82
INCEPTIONV3-PCA-KNN	89.45	11.17
INCEPTIONV3-PCA-SVC	88.38	11.3
MOBILENETV2-PCA-KNN	8.15	12.86
MOBILENETV2-FULL-KNN	88.13	3.35
INCEPTIONV3-FULL-SVC	87.7	15.18
VGG19-PCA-KNN	82.94	5.02
VGG19-FULL-KNN	82.2	3.36
MOBILENETV2-FULL-GNB	75.64	2.57
VGG19-FULL-SVC	70.64	5.98
VGG19-FULL-GNB	68.39	2.98
DENSENET121-FULL-GNB	64.74	2.20
INCEPTIONV3-FULL-GNB	59.70	2.12
VGG19-PCA-GNB	50.32	5.01
MOBILENETV2-PCA-GNB	46.93	12.85
DENSENET121-PCA-GNB	43.88	9.9
INCEPTIONV3-PCA-GNB	43.96	11.16

Based on overall results, we concluded that transfer learning for document image classification is a realistic approach, because most constructed models performed excellently (40 models out of 48 have an accuracy of over 75%,and 33 models achieve an accuracy of over 90%) without customized optimization or special tuning. Furthermore, models based onVGG19 and DenseNet surpassed other deep models for this task with different comparison between the best ten models based on accuracy.

Additionally, LR and KNN were the best classifier with every feature selector and reached up to 97.8% accuracy (with DenseNetand LDA combination). GNB apparently does not fit to the task, which can be expected due to its simplicity. We also investigated the role of

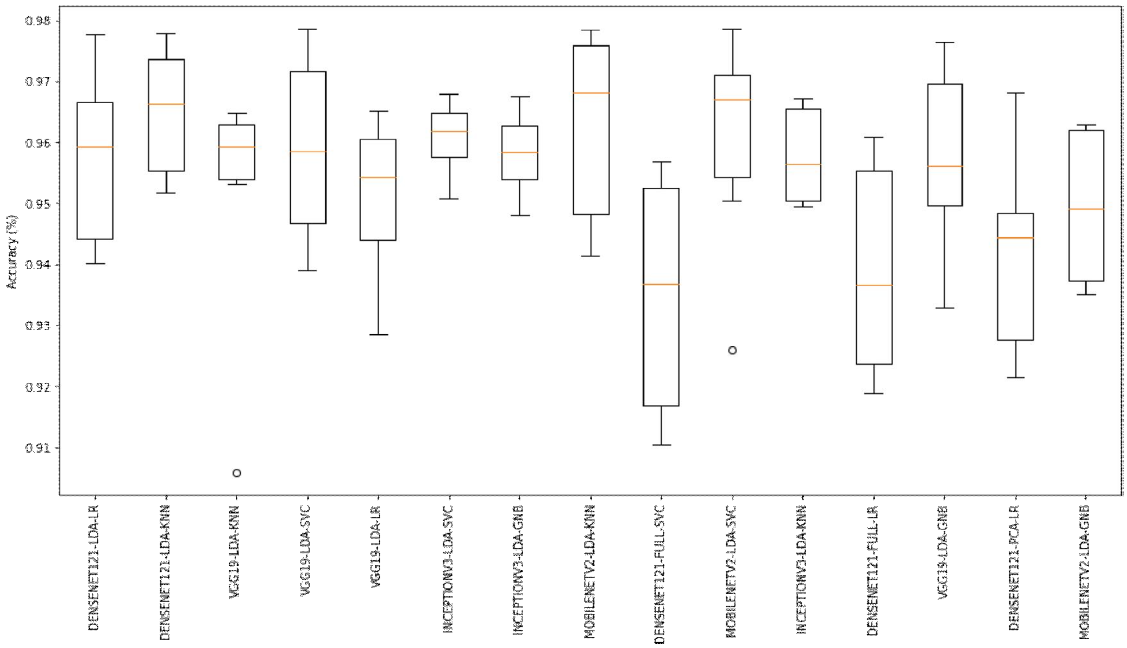

Figure .7 Performance Comparison between best 10-models

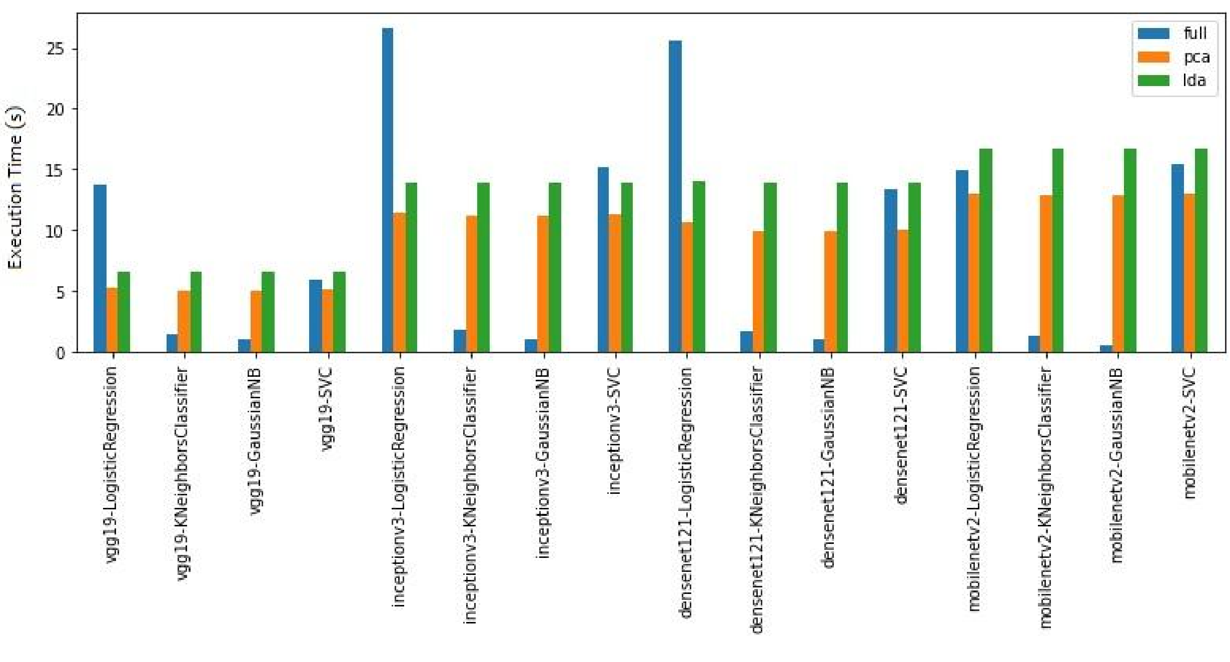

Figure .8 Average Execution time per deep Learning model

classifiers in the last step. Fig. 7 shows a performance dimensionality reduction techniques(i.e., PCA and LDA) to lower the processed features passed to the classification layer. This approach can allow for faster processing (if needed), and reduce classifier training time for almost the same performance levels. Fig.8 shows a comparison of training time between constructed models using different dimensionality reduction techniques.

The performances demonstrated by models trained on reduced features using PCA or LDA were significantly comparable to full-sized ones. For example, models constructed with the LR model on the classification layer took substantially more time for training than almost any other combinations. However, the implementation of PCA and LDA in the same situation has helped to reduce the needed time, for training the same model while keeping similar performance levels. Such techniques can be interesting in environments where computational power is limited and faster processing is required.

5.2. Cross-validation

The second experiment was about executing a cross-validation procedure for every constructed model. The whole dataset is partitioned into k-folds subsets of equal size. Many cross-validation rounds can be performed using many different partitions to reduce variability, and then an average of the results is taken. Cross-validation is an essential technique in estimating the model performance and especially assessing the model's generalization ability;that isits capacity to correctly classify an unseen document image. By varying the number of folds (k), we can overcome the unbalancing dataset's classes and test our conclusions on every part of the dataset.Fig. 9 shows the results of cross-validation tests with no feature reduction.

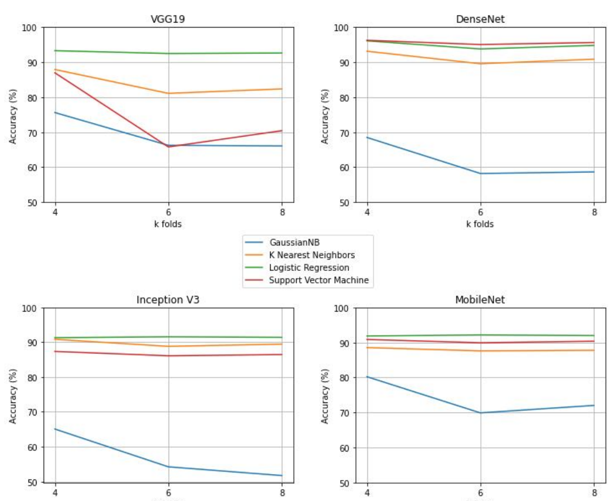

Figure. 9 Cross-Validation results

The results confirm the superior performance of Logistic Regression with all the deep models (over 90%), followed by KNN and SVC at comparable levels. At the same time, GNB apparently does not fit to the task, with accuracy levels at 80% at best.

On the other hand, all the deep models showed comparable performances with all classifiers except for VGG19, which accuracy declined while increasing the folds count (k). On the other hand, DenseNet121 performed very well regardless of the classifier used (over 90% in all combinations except GNB).

6. CONCLUSION AND PERSPECTIVES

This paper presents a comprehensive framework for automatic document classification using Deep Transfer learning feature selection with different award-winning pre-trained deep learning models. Machine learning algorithms classify relevant extracted features for classification purposes into four different document categories. We also explored the dimensionality reduction effect on the model's performance using two different techniques (PCA and LDA). The experimental results reveal that despite a small dataset size (546 document images), transfer learning is a helpful approach in a document image classification context (over 70% of constructed models reached over 90% of accuracy).

We also compared our results with available results in other studies in related fields; in [30] an accuracy score of 98.4 %, by combining the use of KNN and AlexNet on invoice classification (on a dataset of 1380 invoices), and in [31]the authors achieved an accuracy score of 96.6% on Identity document classification based on deep learning and document modeling.

Seeing that the combination of VGG19 model with Logistic Regression provides an excellent performance (97.8%) with default parameters, we can recommend that this approach can be used in practice for document image classification in an ERP ecosystem (e.g. Odoo) as a document management module, or as a pre-processing step for an OCR procedure in order to improve data extraction systems. In our future work, We will explore new feature selection techniques and more specialized machine learning algorithms, we will bigger datasets for training and validation, and will investigate other improvement rooms in document classification, as data augmentation and feature engineering.

Author Contributions

Conceptualization, methodology, investigation and implementation, Aissam JADLI; validation and formal analysis, Mustapha HAIN; writing-review, edition and project administration, Anouar HASBAOUI.

REFERENCES

[1] H. Bast and C. Korzen, "A Benchmark and Evaluation for Text Extraction from PDF," Jul. 2017, doi: 10.1109/JCDL.2017.7991564.

[2] B. Vani, M. S. Beaulah, and R. Deepalakshmi, "High accuracy optical character recognition algorithms using learning array of ANN," in *2014 International Conference on Circuits, Power and Computing Technologies, ICCPCT 2014*, Mar. 2014, pp. 1474–1479, doi: 10.1109/ICCPCT.2014.7054772.

[3] V. Kotu and B. Deshpande, “Lecture 04 : Training Neural Networks,” in *Data Science*, Elsevier, 2019, pp. 307–342.

[4] A. Hamdi, A. Jean-Caurant, N. Sidere, M. Coustaty, and A. Doucet, “An analysis of the performance of named entity recognition over OCRed documents,” in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, Jun. 2019, vol. 2019-June, pp. 333–334, doi: 10.1109/JCDL.2019.00057.

[5] M. Shen and H. Lei, “Improving OCR performance with background image elimination,” in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015*, Jan. 2016, pp. 1566–1570, doi: 10.1109/FSKD.2015.7382178.

[6] W. Zhu, Y. Liu, and L. Hao, “A novel OCR approach based on document layout analysis and text block classification,” in *Proceedings - 12th International Conference on Computational Intelligence and Security, CIS 2016*, Dec. 2017, pp. 91–94, doi: 10.1109/CIS.2016.28.

[7] L. Shao, F. Zhu, and X. Li, “Transfer learning for visual categorization: A survey,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 5, pp. 1019–1034, May 2015, doi: 10.1109/TNNLS.2014.2330900.

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” pp. 248–255, 2010, doi: 10.1109/cvpr.2009.5206848.

[9] N. Huber-Fliflet, F. Wei, H. Zhao, H. Qin, S. Ye, and A. Tsang, “Image Analytics for Legal Document Review : A Transfer Learning Approach,” in *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, Dec. 2019, pp. 4325–4328, doi: 10.1109/BigData47090.2019.9006551.

[10] A. Kolsch, M. Z. Afzal, M. Ebbecke, and M. Liwicki, “Real-Time Document Image Classification Using Deep CNN and Extreme Learning Machines,” in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, Jan. 2018, vol. 1, pp. 1318–1323, doi: 10.1109/ICDAR.2017.217.

[11] L. Yang, L. Jing, J. Yu, and M. K. Ng, “Learning Transferred Weights from Co-Occurrence Data for Heterogeneous Transfer Learning,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 27, no. 11, pp. 2187–2200, Nov. 2016, doi: 10.1109/TNNLS.2015.2472457.

[12] J. Zhang, “Deep transfer learning via restricted Boltzmann machine for document classification,” in *Proceedings - 10th International Conference on Machine Learning and Applications, ICMLA 2011*, 2011, vol. 1, pp. 323–326, doi: 10.1109/ICMLA.2011.51.

[13] Y. A. L. Alsabahi, L. Fan, and X. Feng, “Image classification method in DR image based on transfer learning,” Jan. 2019, doi: 10.1109/IPTA.2018.8608157.

[14] C. Kang and C. He, “SAR image classification based on the multi-layer network and transfer learning of mid-level representations,” in *International Geoscience and Remote Sensing Symposium (IGARSS)*, Nov. 2016, vol. 2016-Novem, pp. 1146–1149, doi: 10.1109/IGARSS.2016.7729290.

[15] J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, and W. Xu, “CNN-RNN: A Unified Framework for Multi-label Image Classification,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-Decem, pp. 2285–2294, doi: 10.1109/CVPR.2016.251.

[16] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553. pp. 436–444, 2015, doi: 10.1038/nature14539.

[17] A. S. Tarawneh, C. Celik, A. B. Hassanat, and D. Chetverikov, “Detailed Investigation of Deep Features with Sparse Representation and Dimensionality Reduction in CBIR: A Comparative Study,” Nov. 2018, doi: 10.3233/IDA-180895.

[18] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Sep. 2015, Accessed: Sep. 19, 2019. [Online]. Available: http://arxiv.org/abs/1409.1556.

[19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Nov. 2017, vol. 2017-Janua, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[20] Y. Liu *et al.*, “Automatic delineation of ribs and clavicles in chest radiographs using fully convolutional DenseNets,” *Comput. Methods Programs Biomed.*, vol. 180, p. 105014, Aug. 2019, doi: 10.1016/j.cmpb.2019.105014.

[21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, Jan. 2018, doi: 10.1109/CVPR.2018.00474.

[22] A. S. Tarawneh, D. Chetverikov, C. Verma, and A. B. Hassanat, “Stability and reduction of statistical features for image classification and retrieval: Preliminary results,” in *2018 9th International Conference on Information and Communication Systems, ICICS 2018*, May 2018, vol. 2018-Janua, pp. 117–121, doi: 10.1109/IACS.2018.8355452.

[23] A. Meyer-Baese and V. Schmid, "Feature Selection and Extraction," in *Pattern Recognition and Signal Analysis in Medical Imaging*, Elsevier, 2014, pp. 21–69.

[24] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010, doi: 10.1002/wics.101.

[25] H. Moutachaouik and I. El Moudden, "Mining Prostate Cancer Behavior Using Parsimonious Factors and Shrinkage Methods," *SSRN Electron. J.*, Jun. 2018, doi: 10.2139/ssrn.3180967.

[26] Y. Ren, H. Liu, C. Xue, X. Yao, M. Liu, and B. Fan, "Classification study of skin sensitizers based on support vector machine and linear discriminant analysis," *Anal. Chim. Acta*, vol. 572, no. 2, pp. 272–282, Jul. 2006, doi: 10.1016/j.aca.2006.05.027.

[27] X. Sun, L. Liu, H. Wang, W. Song, and J. Lu, "Image classification via support vector machine," in *Proceedings of 2015 4th International Conference on Computer Science and Network Technology, ICCSNT 2015*, Jun. 2016, pp. 485–489, doi: 10.1109/ICCSNT.2015.7490795.

[28] K. Kang, F. Gao, and J. Feng, "A new multi-layer classification method based on logistic regression," in *13th International Conference on Computer Science and Education, ICCSE 2018*, Jan. 2018, pp. 52–55, doi: 10.1109/ICCSE.2018.8468725.

[29] D. A. Salazar, J. C. Salazar, and J. I. Vélez, "Comparison between SVM and logistic regression: Which one is better to discriminate?," *Rev. Colomb. Estad.*, vol. 35, no. 2 SPEC. ISSUE, pp. 223–237, 2012.

[30] A. S. Tarawneh, A. B. Hassanat, D. Chetverikov, I. Lendak, and C. Verma, "Invoice Classification Using Deep Features and Machine Learning Techniques," in *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology, JEEIT 2019 - Proceedings*, Apr. 2019, pp. 855–859, doi: 10.1109/JEEIT.2019.8717504.

[31] A. M. Awal, N. Ghanmi, R. Sicre, and T. Furon, "Complex Document Classification and Localization Application on Identity Document Images," in *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, Jan. 2017, vol. 1, pp. 426–431, doi: 10.1109/ICDAR.2017.77.

[32] Y. Zheng, C. Yang, and A. Merkulov, "Breast cancer screening using convolutional neural network and follow-up digital mammography," May 2018, p. 4, doi: 10.1117/12.2304564.

[33] M. A. Ponti, L. S. F. Ribeiro, T. S. Nazare, T. Bui, and J. Collomosse, "Everything You Wanted to Know about Deep Learning for Computer Vision but Were Afraid to Ask," in *Proceedings - 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials SIBGRAPI-T 2017*, Jul. 2017, vol. 2018-Janua, pp. 17–41, doi: 10.1109/SIBGRAPI-T.2017.12.