Handwritten Text Recognition for Documentary Medieval Manuscripts

Sergio Torres Aguilar, Vincent Jolivet

Handwritten Text Recognition for Documentary Medieval Manuscripts

Sergio Torres Aguilar[1] **and Vincent Jolivet**[2]

[1]Université du Luxembourg, Institut für History, Luxembourg
[2]École nationale des chartes, Centre Jean-Mabillon, France

Corresponding author: Sergio Torres , sergio.torres@uni.lu

Abstract

The handwritten text recognition (HTR) techniques aim to effectively recognize sequence of characters in an input manuscript image by training an artificial intelligence into the historical writing features. Efficient HTR models will help to transform digitized manuscript collections into an indexed and quotable corpus able to provide meaningful research clues to historical questions. Before, several issues must be addressed such as the lack of relevant training corpora; the large number of variations proposed by each scribal hand and by each writing script or the complex page layout. This paper presents two models and one cross-model aiming to automatically transcribe Latin and French medieval documentary manuscripts, mostly charters and registers, produced between the 12th and 15th centuries and classified into two major writing scripts: Textualis (from the late-11th to 13th century) and Cursiva (from the 13th to the 15th century). The models architecture is based into a CRNN network coupled to a CTC loss. The training and evaluation, entailing 120k lines of text and almost 1M tokens, were conducted using three ready-to-use ground-truth corpora : The Alcar-HOME database, the e-NDP corpus and the Himanis project. We describe the training architecture and corpora and we discuss the main training problems, the results and the perspectives open by HTR techniques on medieval documentary manuscripts.

Keywords

medieval charters, HTR for historical documents, HTR for medieval Latin manuscripts, digital diplomatics, medieval digital studies

I INTRODUCTION

In recent years the improvement of HTR techniques and the emergence of several ground-truth corpora have made possible to develop powerful models for handwriting recognition applied to various sources of interest for historical research. The transition from the experimental phase to the production phase has been completed in the last five years thanks to the popularization of general purpose HTR toolkits and turn-key annotation interfaces. For many institutions, the automatic acquisition of the text of their collections already digitized or in the course of digitization is already a key step in their study and dissemination as structured texts allow unrestricted query, indexing and share operations. This relatively new situation has increased the demand for ground-truth close to their needs as well as for robust models previously trained in order to readapt them to their specific requirements.

To contribute to alleviate this need, in this paper we present three HTR models that can be efficiently applied on a large spectrum of manuscripts concerning charters collections, registers and serial documents, which constitute the most numerous and often under-studied sources, due to the immense documentary mass, especially from the 13th century onwards. These models, ready to be deployed and fine-tuned, focus on two script families : the so-called *Textualis*, used in manuscripts between the late-11th and 13th centuries, and the *Cursiva*, whose existence can

be traced from the mid-13th century until well into the Modern Age. The ground-truth comes from collections comprising two of the most important written sources for the study of the Middle Ages: cartularies, essentially copies of deeds with legal value concerning the transfer of rights (sales, donations, exchanges, etc.); and registers, i.e. serial documents concerning the internal management and the public activity of an institution.

This paper is organized in three parts: In the first one we present the training corpus and the architecture used for the modeling of writing; in the second one we analyze the results and the most common errors. Finally, we analyze some of the most relevant issues raised by the HTR from the perspective of medieval paleography and diplomatics.

II RELATED WORKS

The HTR techniques applied to manuscripts have evolved over the last 10 years from Markov engine-based models (Bunke et al. [1995]) to deep-learning based on neural networks applied at the character level (Graves and Schmidhuber [2008]). The emergence of CTC mechanisms and the introduction of bidirectional networks in recent years (Graves et al. [2008]) has greatly improved the accuracy obtained and significantly reduced the need for ground-truth for the develop of generalist models (de Sousa Neto et al. [2020]). In the medieval manuscript world, the Himanis (Stutzmann et al. [2017]) and Home (Stutzmann et al. [2021]) projects, supported by Transkribus, have pioneered the creation of massive aligned and documented corpora for training effective HTR and NLP models on medieval sources. While the Scripta-PSL and Cremma projects (Chagué [2021], Pinche [2022]), supported by Kraken and the Scriptorium platform, have introduced widely adopted scientific practices for the production of models applied to medieval Latin and Hebrew manuscripts.

Furthermore, in recent years, HTR techniques for historical manuscripts have come closer and closer to the scholarly sciences that traditionally studied the history of writing and scriptural practices such as paleography and diplomatics. This is a natural tandem, because production of ground-truth for the modeling of ancient scripts is not only expensive and time consuming, but it requires specialized knowledge for reading and decoding ancient languages and scripts and also for to propose intellectual manners to clarify misconceptions and to define annotation guidelines. Paleographers and historians try to contribute to the creation and selection of stable ground truth, as well as to the feedback analysis of the most common errors produced during the inference phases.

Unlike OCR, which is confronted with a limited set of typographic variations, HTR has to deal with a much more intense variability that forces to form specialized models not only according to the chronology of the texts, but also to the typology of the document, the scripts families, the regional practices and what is even more difficult to fit: the scribal behaviors of the scribes. All these diplomatic observations have led to the emergency of new discussions beyond the simple accuracy-based result about the nature of the ground : is it necessary to develop abbreviations; should the transcription be graphemic or diplomatic i.e. imitative (Driscoll [2006]); how to define the representativeness of a training corpus and mitigating potential bias?

Writing is a normative system and as such can be modeled. Nowadays, its considered proven that modern HTR on medieval manuscripts can produce inference rates of less than 5% CER during the evaluation phase. But it can also be state that this rate is often drastically reduced when models are confronted to documents outside the domain or belonging to another script family. In ancient scripts the reader must restore a part of the reading encoded in abbreviations, formulae and implicit knowledge that in the current state of the art cannot be fully modeled.

In previous years hand-level and author-centered modeling proposals were common, but given the great variety of writing styles over the medieval centuries, and the scarce existence of domain-specific ground truth, it is now evident to subscribe to a global level of handwriting classification, which in our case are script families. The CLAMM corpus proposes 12 book-script families (Kestemont et al. [2017]) covering the manuscript spectrum between the 9th and 15th centuries. The models we propose concentrate on the two most numerous for which sufficient ground truth exists: cursive (mid-13th-late 16th) and Textualis (late 11th - mid 13th). The division is purely consensual, since towards the end of the 14th century the number of families and sub-families scripts increases widely.

III CORPORA DESCRIPTION

3.1 The e-NDP corpus

The corpus of the registers of Notre-Dame-de-Paris and its cloister is one of the few available corpora to deal with the HTR of documentary manuscripts from the late medieval period (Claustre and Smith [2022]). The registers are the minutes of decisions of the weekly meetings that the canons of the cloister held to deal with matters concerning the management of the institution and its patrimony. One of the central objectives of the e-NDP project was to obtain the text of the more than 14k pages that make up the totality of the records in the medieval period (Archives nationales, LL105 to LL128). Later, the text inferences have been structured and transformed into a research engine able to indexing and quickly revealing clues of study on a massive source about one of the most important urban institutions of the medieval period in France. To this end, a total of 500 pages from 26 registers written between 1326 and 1504 were transcribed using Scriptorium by historians and paleographers in order to train HTR adapted models able of producing a high quality transcription of all the digitized pages.

The registers have been written using a family of cursive scripts, essentially in Latin, with some pages and formulations in medieval French. They also present an evolving page layout with certain stereotypical patterns such as lists of names, margin notes, titles and other peritextual additions on a design less careful than that of literary manuscripts. All these are typical of the so-called documentary manuscripts, i.e. documents of the management and daily life of ancient societies not always produced to last but to function as written instruments of consultation and administration.

Table 1 List of training and testing manuscripts.

Set	Manuscript	code name	script type	nº lines	nº tokens
train	Cartulary of the Notre-Dame de la Roche abbey	Roche	Cursiva antiquior	2103	19164
train	Cartulary of Saint-Denis abbey	S_Denis	Textualis	18363	132854
train	Cartulary of Charles II of Navarre	Navarre	Cursiva	6777	94262
train	Cartulary of Notre-Dame de Vauluisant abbey	Vauluisant	Textualis	12642	69364
train	Cartulary of Notre-Dame de Fervaques abbey	Fervaques	Textualis	4661	45251
train	Cartulary of Saint Nicaise of Reims	S_Nicaise	Textualis	7404	99526
train	Cartulary of Notre-Dame de Clairmarais	Clairmarais	Semi-hybrida & Cursiva	8554	77478
train	Formulary of Odart of Morchesne	Morchesne	Cursiva	10515	110033
train	Registers of the chapter of Notre-Dame de Paris	e-NDP	Cursiva	33735	202348
test	Cartulary of Nesle seigneury	Nesle	Cursiva & Textualis	3562	37756
test	Cartulary of the Pontigny abbey	Pontigny	Textualis & Cursiva antiquior	10717	78045
test	Cartulary of the Cathedral of Notre-Dame de Chartes	Chartres	Textualis	1636	14564
test	Register of the French Royal Chancery	Himanis	Cursiva	485	8441
	Total Train			104754	850280
	Total Test			16400	138806
	Train + Test			**121154**	**989086**

3.2 The HOME-Alcar corpus

The HOME-Alcar corpus published in 2021 (Stutzmann et al. [2021]) was produced as part of the European research project *HOME History of Medieval Europe*, under the coordination of Institut de Recherche et d'Histoire des Textes (IRHT-CNRS). This corpus provides the images of medieval manuscripts aligned with their scholarly editions at line-level as well as a complete annotation of named entities (persons and places), as a resource to train synchronously Handwritten Text Recognition (HTR) and NER models.

The corpus includes 17 French cartularies, i.e. volumes containing medieval copies of original documents, produced between the 12th and 14th centuries and belonging to at least 4 script families : Textualis, Cursiva, Cursiva Antiquior and Semi-Hybrida. The productions of cartularies was very common in ecclesiastical institutions since the 11th century and also in civil institutions from the 13th century onwards. These volumes are highly appreciated in medieval studies since they contain documents that have hardly been preserved in their original form such as property transfers, wills, land and debt disputes, but also rarer ones such as treaties, indemnities or successions. All of them are included in the HOME-Alcar corpus. In summary, this corpus contains 3090 acts, with 2760 in Latin and 330 in Old and Middle French, and almost 1M tokens.

3.3 The Himanis project

The Himanis project in collaboration with the READ consortium (Recognition and Enrichment of Archival Documents) developed the most efficient model to date for late medieval script recognition (Himanis Chancery M1+) based on the partial edition of the registers produced by the French Royal Chancery between 1302 and 1483 (Archives nationales, JJ35 to JJ211). This serial so-called cartulary-register contains essentially copies of charters ranging from letters of remission to mandates, amortizations, ennoblements, and property confirmations.

The Himanis model was developed aligning at line-level the digitized images with the partial critical edition of Paul Guerin (Guérin [1881]) , then encoded using the PyLaia (Puigcerver and Mocholí [2018]) architecture in Transkribus. With this model, which shows a validation CER of 0.08, more of 70k pages were transcribed in order to build a querying tool with indexing and word-spotting functionality directly on the entire French Royal Chancery corpus. In 2021 the training dataset (HIMANIS-Guérin) was freely released on Zenodo contending 1500 images and 30k text lines of ground-truth written mostly in Latin and Old French using a Cursiva script.

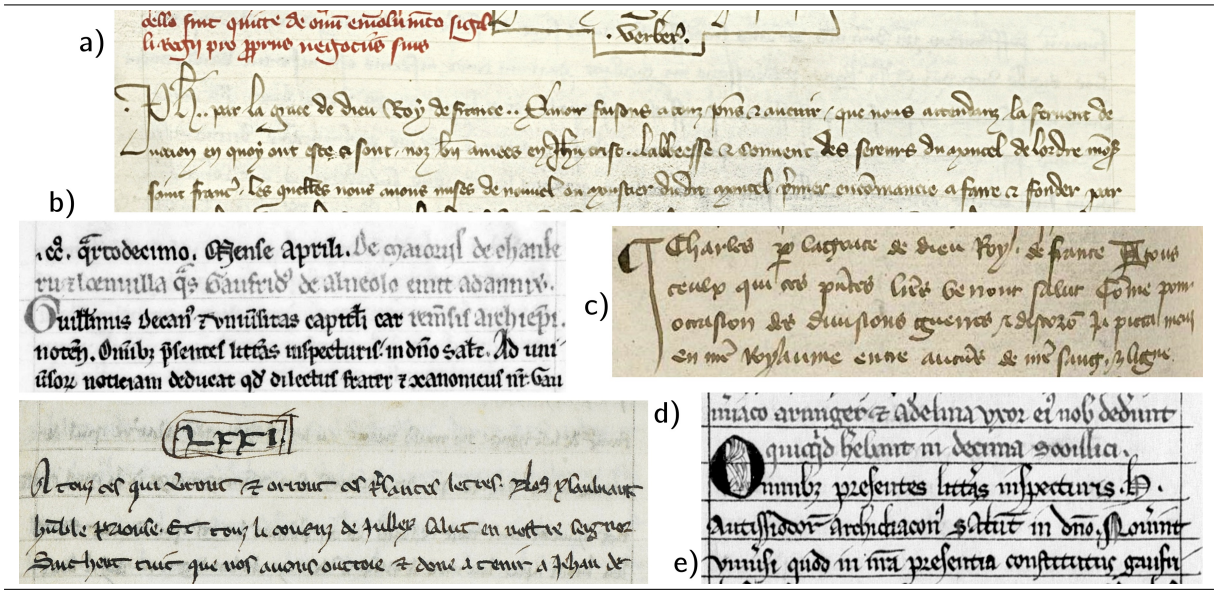
3.4 Datasets configuration

During the training phase, the manuscripts were distributed into two groups (see Table 1) according to the predominant type of writing:

- Cursiva (**G1**) contains 5 elements: The e-NDP registers; and the cartularies of Navarre, Clairmarais, Notre dame de la Roche and the formulary of Odart de Morchesne for a total of 61684 lines.
- Textualis (**G2**) contains 4 elements: The cartularies of Saint Denis, Fervaques, Saint Nicaise and Valuisant for a total of 43070 lines.

Four manuscripts not used during training were reserved for the test phase: (1) the Nesle cartulary, essentially written in Cursive but with some pages in Textualis; (2) the Chartres cartulary, written entirely in Textualis; (3) the Pontigny cartulary which presents interspersed parts in Cursive Antiquior and Textualis and finally (4) a Cursiva random set of 30 pages of Himanis taken from all the registers volumes.

Figure 1 Five examples of act protocols from Himanis (a), Chartres (b), e-NDP (c), Nesle (d) and Pontigny (e). In both Latin and French, all five follow a similar act opening containing an intitulatio, a general address (”to all who will read this letter”) and a salutation.



Both the training and test corpus are bilingual in an overall ratio of 4:1 (Latin / French) for the training set and 5:2 for the testing set. In addition, as in most editions, text abbreviations have been developed and punctuation normalized.

Besides, the commas (,) of the ground truth were modified in all documents in order to better reflect their use in the manuscript: In the cases where the modern editor introduces the comma (or the semicolon) to mark a pause in the sentence although this punctuation mark does not exist in the manuscript the comma was eliminated; in other cases the comma was replaced by the period (.) since in most manuscripts it is the period that exercises the three functions (soft, middle and final) of separation of the sentences or clauses.

IV HTR ARCHITECTURE

The HTR architecture applied in this paper can be defined as a classical CRNN pattern recognition approach using 5.9M of trainable parameters. The operations follow three steps: (i) gray-scaled images are the inputs of four CNN layers to extract and encode features; (ii) three RNN layers propagate the information from CNN and map the features in a bidirectional mode ; finally (iii) a Connectionist Temporal Classification (CTC) algorithm calculates the model loss value and decodes the text inference.

The convolutional block is composed by layers with variable kernels size (4x16, 3x8) and variable 16n filters by layer (32, 32, 64, 64). After each convolution the neurons pass thought a MaxPooling layer with a 2x2 kernel. Furthermore, we apply a 2D dropout (0.1 probability) to each layer and we use Rectifier Linear Units (ReLU) as activator.

Before apply the recurrent block a reshape layer collapse the non-1 height dimensions to a single value as the height of the images was fixed to 128 pixels while a variable width was allow in order to better leading with a variable size of manuscripts lines.

The recurrent block, ideal to process sequences of variable length, is activate afterwards. It is composed by three BiLSTM networks using a 1D dropout (0.3 probability). The number of

hidden units in all LSTMs is set to 256. Finally, the output goes through a dense layer (softmax activation) with a size equal to the charset size + 1 (CTC blank symbol). A 1D dropout is also applied before this last layer (probability 0.3).

In Kraken which use the VGSL network specification, this training architecture can be fully replicated by using : `-s '[1,128,0,1 Cr4,16,32 Do0.1,2 Mp2,2 Cr4,16,32 Do0.1,2 Mp2,2 Cr3,8,64 Do0.1,2 Mp2,2 Cr3,8,64 Do0.1,2 S1(1x0)1,3 Lbx256 Do0.3,2 Lbx256 Do0.3,2 Lbx256 Do0.3]'`

4.1 Hyperparameters

During the training we use a batch size of $\{1\}$; a learning rate of $\{10^{-4}\}$; a pad size of $\{24\}$ pixels (useful to give more space to the kernel in order to cover the image); and we ran a $\{$ReduceOnPlateau$\}$ optimizer with patience $\{3\}$. Furthermore, we enable the usual data augmentation techniques (distorsion, blur and rotation) to the images.

All the training cycles were realized using a RTX 3090 (24GB) coupled to a 64GB of RAM for about 6-40 hours depending of the size of the corpus.

V EXPERIMENTS

Our experiments were developed based on three objectives: (i) To create generic models for each of the script families; (ii) To investigate the relationship between the performance accuracy and the amount and variety of ground truth; (iii) To determine the disadvantages and benefits of creating a composite model combining the resources available from both script families. For this purpose four experiments were performed:

1. A step-wise training using batches of documents for each script family. (Tables 2 and 3)
2. A cross-script and cross-lingual training on quartiles (Q1 is composed of 400 pages : 200 random pages from each script family; Q2 contains 400 of each one, etc.) using the entire ground-truth (G1 + G2)
3. A fine-tuning exercise (retraining on initialized weights) for the models developed in points 1 and 2.
4. A Testing exercise on four out-of-domain multilingual and mixed-families manuscripts unseed during training in order to examine the generalization capacity and the retraining sensitivity of the models.

For each of the models the metrics will be expressed in the widely known accuracy (errors / characters), CER (Character error ratio) and WER (word error ratio) without applying normalization to the results. For the fine-tuning exercises 10 reserved pages from each manuscript were used.

VI RESULTS

The results on Tables 2 and 3 show that the proto-Gothic and early Gothic scripts (Textualis) converge much faster to a result of over 0.9 accuracy than Late-Gothic scripts (Cursiva) which needs in general a significantly higher number of pages to converge to the same result. On the other hand, the fine-tuning on the Textualis manuscripts (Chartres) does not provide a large percentage improvement to the generalist model (+1.2%). In this case, the Cursiva manuscripts (Nesle and Himanis) are more sensitive to the fine-tuning as the accuracy is highly improved (+6% and +8%) on three of the tests experiments performed. (See Table 4). The fitting on Pontigny, mixing Textualis and Cursiva antiquior, is setting naturally in the middle: +4.5%.

Table 2 Evaluation results for G1 models. *val_acc*: validation accuracy during training. *test_acc*: testing accuracy on the cartulary of Nesle

model_name	Content	pages	val_acc	test_acc	CER	WER
G1_test_1	Clairmarais, Roche	121 + 66	0.935	0.696	0.296	0.629
G1_test_2	Clairmarais, Roche, +Navarre	187 + 208	0.939	0.798	0.192	0.518
G1_test_3	Clairmarais, Roche, Navarre, +Morchesne	395 + 176	0.941	0.797	0.194	0.514
G1_test_4	Clairmarais, Roche, Navarre, Morchesne, +e-NDP/5	571 + 100	0.939	0.807	0.183	0.501
G1_test_5	Clairmarais, Roche, Navarre, Morchesne, +e-NDP full	671 + 400	0.936	0.840	0.156	0.448
G1_FineTuning	+10 pages Nesle	1071 + 10	**0.942**	**0.901**	**0.095**	**0.277**

Table 3 Evaluation results for G2 models. *test_acc*: testing accuracy on the cartulary of Chartres.

model_name	Content	pages	val_acc	test_acc	CER	WER
G2_test_1	S_Denis, Fervaques	199 + 90	0.929	0.885	0.097	0.278
G2_test_2	S_Denis, Fervaques, +S_Nicaise	289 + 109	0.932	0.906	0.078	0.225
G2_test_3	S_Denis, Fervaques, S_Nicaise, +Vauluisant	398 + 106	0.935	0.914	0.071	0.213
G2_FineTuning	+10 pages Chartres	504 + 10	**0.938**	**0.926**	**0.060**	**0.163**

Table 4 Evaluation results for cross-scripts models by quartile. Each quartile (Q*n*) contain 25% of the entire training corpus (G1 + G2 : 1575 pages). The Fine-tuning experiments (Q*n* FT) were performed using 10 reserved pages from each manuscript. For Nesle and Chartres these pages and the test-set are the same as in the previous experiments.

Manuscript /	Nesle (G1)			Chartres (G2)			Pontigny (G1 + G2)			Himanis (G1)		
metric	test_acc	CER	WER	test_acc	CER	WER	test_acc	CER	WER	test_acc	CER	WER
Q1 (25%)	0.841	0.153	0.430	0.911	0.073	0.209	0.821	0.166	0.454	0.789	0.222	0.580
Q1 FT	0.890	0.106	0.292	0.917	0.068	0.183	0.879	0.110	0.301	0.863	0.136	0.416
Q2 (50%)	0.862	0.133	0.386	0.919	0.065	0.179	0.843	0.146	0.397	0.803	0.202	0.544
Q2 FT	0.904	0.090	0.256	0.926	0.060	0.163	0.888	0.101	0.275	0.884	0.115	0.361
Q3 (100%)	0.868	0.127	0.363	0.923	0.062	0.170	0.848	0.142	0.373	0.832	0.171	0.468
Q3 FT	0.910	0.084	0.234	0.928	0.057	0.158	0.892	0.097	0.264	0.893	0.104	0.324

The HTR modelizations are usually conducted on an entire corpus of manuscripts, but as we can see in the cross-family models (Table 4), using batches with a smaller number of pages but of more diversified origin can be sufficient to achieve good results. The cross-family models with 200 pages of each script (Q1, 25%) from 9 manuscripts identify faster the dominant trends and achieve very similar results to the best Textualis (0.914 vs 0.917) and Cursiva (0.840 vs 0.841) models (Tables 2 and 3) containing more than two times the amount of ground truth. The improvements are much more costly from the inflection point of around 85% accuracy in Cursiva and 90% in Textualis, and the increase in quantity importing scarce new information brings discrete gains and lead the accuracy model to plateau. As shown by the Q2 and Q3 models (table 4), the improvement by 1 to 2 points of accuracy on the four test manuscripts were made at the cost of introducing more than a thousand new pages of ground truth.

The tests conducted to verify the impact of a multilingual and multifamily corpus show that a cross-training does not entail any loss in accuracy, but on the contrary shows a slight improvement in Chartres (0.914 vs 0.923) and a significant improvement in Nesle (0.840 vs 0.868), the latter mainly written in Cursiva with some pages in Textualis. As for the bilingualism, as in the case of Nesle, Himanis and Pontigny, the existence of a double linguistic record does not seem to dramatically reduce the quality of the inference, although it could be a significant source of errors, since the Latin abbreviation system is not the same as the one used in French. This type of multilingual and multifamily models should therefore be practiced not only because is a easy manner to reduce bias an increase variance but because they respond better to medieval

documentary textuality : in the one hand, ancient documentary compilations containing documents written by diverse hands and family scripts, sometimes within the same page; on the other hand, handwriting is evolutionary and transitional scripts (such as Cursiva antiquior) are widely practiced. Besides, multilingual capacities of HTR models for late-medieval sources are a prerequisite as from the mid-13th century the acts in vernacular and Latin coexist within the same manuscript or collection and the Latin formulation in legal value documents is widely used in late-medieval periods (Glessgen [2004]).

Table 5 The ten characters accumulating the most errors (substitutions + deletions + insertions) according to the hypotheses of the Q3 and Q3 FT models on the test datasets. The updating of the models on new data allows for significant improvements in the reading of the most problematic characters.

model /	Nesle (205064 chars)			Chartres (97129 chars)			Himanis (52110 chars)			Pontigny (464660 chars)		
char	Q3	Q3 FT	+diff (%)	Q3	Q3 FT	+diff (%)	Q3	Q3 FT	+diff (%)	Q3	Q3 FT	+diff (%)
space	2706	2057	24	1066	1022	4	536	347	35	5721	4822	16
s	2383	863	64	435	392	10	707	335	53	5987	3876	35
i	2211	1734	35	675	613	9	958	598	38	7870	5303	33
e	2307	1492	22	643	598	7	1098	546	50	6370	4891	23
n	2045	1143	44	379	346	9	731	434	41	5365	3457	36
r	1775	1190	33	301	308	-2	823	401	51	4461	2925	34
u	1426	930	35	392	369	6	620	455	27	4473	2911	35
t	1390	795	43	359	327	9	605	326	46	2852	2036	29
a	1087	784	28	328	300	9	587	248	58	3981	2232	44
m	979	738	25	294	265	10	449	252	44	3949	2945	25
c	866	684	21	424	321	24	437	279	36	2614	1448	45
Total	19175	12410	35	5296	4861	8	7551	4221	44	53643	36846	31
impr acc	205064 / 6765		**+3.30**	97129 / 435		**+0.45**	52110 / 4221		**+6.39**	464660 / 16797		**+3.61**

As concerns the analysis of the errors, as we can see in Table 5 in all cases a similar pattern is followed: the most numerous errors correspond to insertions on a more or less restricted list of characters: *i*, *e*, *s*, *r*, *n*, *m*, *t*, *u*, *a*, *white-space*. On closer inspection, these errors often coincide with a misrecognition of some typical phenomena of medieval handwriting : indistinctness in the characters composed of succesive *minims* (single strokes) (case of *n*, *m*, *i*, *u*, e.g : *in*d*iu*i*due*, *manda*u*imu*s); misrecognized ligatures (case of *st*, *ct*); undeveloped or incompletely developed abbreviations by suspension, easy to fit (ex. no[-bis], franc[-orum]) or by contraction, harder to fit (ex. m[a]g[is]t[er], d[o]m[i]n[us]); final form variants (case of *s*, *r* and *a*); accented spellings (*é*, *à*) or unresolved diphthongs (case of *ae* and *oe*). The case of the *white-space* is special because it responds to several phenomena: one is editorial, since modern editions, from which most of the transcriptions come, introduce punctuation marks and spaces whose existence is often not reflected in the manuscript; other are intrinsic to the writting such as the fake ligatures and the long endstrokes, specially after prepositions or articles (see figure 2 and Q3 transcription) which hinder the automatic isolation of words; the *scripta continua* (more common in manuscripts prior to the 12th century) or the dissimilar use of the blank space and the period practiced by scribes to separate words and blocks of sentences in documentary manuscripts.

Finally, it is pertinent to ask what are the specific features learned by the model during the fine-tuning process. It is clear that some optimizations are the result of a synchronic capture of new data points. The fine-tuned model allows for a more accurate learning of the cursive flow of a hand and of the abbreviating behavior of a scribe. This leads to a more efficient segmentation of words (white-space errors) and to a more complete development of abbreviations (word-ending and middle-word errors) thus avoiding literal transcription (e.g. *Guille* instead of *Guillaume*). This is the case of Himanis and Pontigny who progress in a synchronous manner on the *i*, *e*, *m*, *u*, *r*, *b* and the *blank* (see figure 2). In other, more rare cases, it may be the learning of a

style feature specific to the scribe. This is the case of Nesle (see figure 3), since one of the scribes uses an atypical letter " s " whose form is very close to the " l " used by the ground truth manuscripts. The Q3 model transcribes an " l ", while fine-tuning corrects this defect. This simple change increases the global accuracy by almost 1% since it occurs on two of the most used characters in the manuscript.

Figure 2 Transcripted line in Himanis (AN, JJ073, 44v, l.22). GT: ground-truth; Q3: prediction of the Q3 model; FT: Prediction of the Q3 fine-tuned model.

GT : royal de grace especial et certaine science au dit Guillaume à ses hoirs et à ceulx qui auront cause de
Q3 : royal de quice especial et cavine science ausdit Guille a ses hoirs et aceube qui auront cai se de
FT : royal de grace especial et certauine science au dit Guillame a ses hoirs et a ceulx qui auront cause de

Figure 3 Transcripted line in Nesle (f. 106r, l.16).

GT : estoient en sa justise et en sa seignerie et des quex il estoit saisiz quant nos
Q3 : estoient en la justile et en la leignerie et desquet il estoit lailiz qoant nes
FT : estoient en sa justise et en sa seignerie et des quer il estoit saisiz quant nos

VII DISCUSSION

These results have several explanations from the point of view of paleography and medieval diplomatics. On the one hand, the Textualis model has been trained exclusively on cartularies which closely follow the bookhand script. Monks writing books and copying cartularies take great care in the layout of the letters and in the absence of fancies to facilitate reading. Except for occasional ornamentations and stylistic preferences introduced by some professional pen-writers, the scribal scripts and their abbreviative systems during the 12th century were similar from one place to another (Hasenohr [1998]). In this universe where individual variability is limited by the type of script, the unseen test data agree close with training data, which produces models with high generalization capacity, although still carrying a significant bias level. This may largely explain the fact that updating weights on a robust generalist Textualis model provides for only modest improvements in handwritting recognition. On this assumption we can presume that this model should work well on book manuscripts, but not as well on original charters collections which display a richer graphical apparatus.

The gotization of writing from the 13th century onwards introduces changes that had to do with extreme angularity, pen tools and writing speed. Shortly after, the Cursive variant, made by a hand that does not rise between each character, introduces deeper changes, moving the axis of the writing from vertical to horizontal, simplifying the ductus and considerably developing the ligatures (Poulle [1966], Guyotjeannin et al. [1993]). The cursive style makes the modifications in the shape and size of the letters more dependent on the hand of the scribe, thus introducing a greater degree of variability. Besides, when the laity, and no longer only the monastic orders begin to produce acts, we see the emergence of new written instruments, many of them multilingual, such as accounts, fiefs and registers, open documents whose handwriting and page setup is more variable. In the long term, this will mean that any HTR work on the Cursiva must be done on a much greater multiplicity of written records, hands, languages and documentary typologies. which increases the number and complexity of the features and trends to be learned

by the model. This can undoubtedly result in models with underfitting problems that make it imperative to collect and provide new specific information through the fine-tuning.

VIII CONCLUSION

In this paper we have presented specialized models in handwriting recognition for documentary and serial manuscripts produced between the 12th and 15th centuries. Our robustness experiments show that a single model is able to recognize with high accuracy (≥ 0.85 points without fine-tuning, ≥ 0.90 with fine-tuning) several handwriting families and to work on various linguistic and documentary registers. Training experiments show that small batches of ground-truth from diverse sources generate results similar to those of training on large uniform corpora. Finally, that fine-tuning shows positive results when the manuscript involves a high number of variations that elude the regularity of the script, which is more frequent in late-Gothic scripts or in late-medieval documentary manuscripts.

IX MODEL REPOSITORIES

A Zenodo repository contending the training logs and models supporting this work is available at:

https://zenodo.org/record/7401833.

References

Horst Bunke, Markus Roth, and Ernst Günter Schukat-Talamazzini. Off-line cursive handwriting recognition using hidden markov models. *Pattern recognition*, 28(9):1399–1413, 1995.

Alix Chagué. Cremma: Une infrastructure mutualisée pour la reconnaissance d'écritures manuscrites et la patrimonialisation numérique. In *Sciences du patrimoine-sciences du texte. Confrontation des méthodes*, 2021.

Julie Claustre and Darwin Smith. e-ndp notre-dame de paris et son cloitre (1326-1504). *Revue Mabillon*, 2022.

Arthur Flor de Sousa Neto, Byron Leite Dantas Bezerra, Alejandro Héctor Toselli, and Estanislau Baptista Lima. Htr-flor: A deep learning system for offline handwritten text recognition. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 54–61. IEEE, 2020.

Matthew Driscoll. Levels of transcription. In *Electronic textual editing*, pages 254–261. Modern Language Association of America, 2006.

Martin-D Glessgen. L'écrit documentaire dans l'histoire linguistique de la france. *La langue des actes. Actes du XIe Congrès international de diplomatique*, 2004.

Alex Graves and Jürgen Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. *Advances in neural information processing systems*, 21, 2008.

Alex Graves, Marcus Liwicki, Santiago Fernández, Roman Bertolami, Horst Bunke, and Jürgen Schmidhuber. A novel connectionist system for unconstrained handwriting recognition. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):855–868, 2008.

Paul Guérin. *Recueil de documents concernant le Poitou contenus dans les registres de la Chancellerie de France*, volume 11. Société des archives historiques de Poitou, 1881.

Olivier Guyotjeannin, Jacques Pycke, and Benoît-Michel Tock. *Diplomatique médiévale*. Brepols, 1993.

Geneviève Hasenohr. Abréviations et frontières de mots. *Langue française*, pages 24–29, 1998.

Mike Kestemont, Vincent Christlein, and Dominique Stutzmann. Artificial paleography: computational approaches to identifying script types in medieval manuscripts. *Speculum*, 92(S1):S86–S109, 2017.

Ariane Pinche. Htr model cremma medieval, June 2022. URL https://doi.org/10.5281/zenodo.6669508.

Emmanuel Poulle. *Paléographie des Écritures Cursives en France du XVe au XVIIe*. Librairie Droz, 1966.

Joan Puigcerver and Carlos Mocholí. Pylaia. https://github.com/jpuigcerver/PyLaia, 2018.

Dominique Stutzmann, Jean-François Moufflet, and Sébastien Hamel. La recherche en plein texte dans les sources manuscrites médiévales: enjeux et perspectives du projet himanis pour l'édition électronique. *Médiévales. Langues, Textes, Histoire*, 73(73):67–96, 2017.

Dominique Stutzmann, Sergio Torres Aguilar, and Paul Chaffenet. HOME-Alcar: Aligned and Annotated Cartularies, 2021. URL https://zenodo.org/record/5600884. Zenodo: https://doi.org/10.5281/zenodo.5600884.