Document Analysis and Recognition: A survey

Shivangi Nigam, Shekhar Verma, P. Nagabhushan

***Abstract*—The journey of research for Document Analysis and Recognition (DAR) started with the problem of automatic character recognition. Today, it has covered a vast span of recognition tasks such as text recognition, script identification, WI, word spotting, signature verification etc., in scripts like Roman, Arabic, Chinese, Japanese, Brahmi etc. Extensive advancements in deep learning techniques have achieved state-of-the-art results for various DAR tasks. In this work, we explore the challenges with different perspectives and review the DAR techniques for online/offline and printed/handwritten documents. We examine the research works with the view of script-related challenges. Various datasets for DAR are categorized into historic, printed and handwritten datasets. We present a comprehensive survey of challenges, techniques, datasets, evaluation metrics, script-related perspectives and potential future directions for DAR.**

***Index Terms*—Document analysis and recognition, text detection, text recognition, text segmentation, deep learning, neural networks**

I. INTRODUCTION

Document analysis and recognition (DAR) is the ability to localize and transform the textual content of document images into a machine-readable format. It concerns learning patterns in text images. Text localization, feature extraction and classification of document images can produce transcripted text for machine interpretation. The text in images carries semantic information which is useful for various pattern recognition and computer vision applications such as image search, intelligent inspection, industrial automation, robot navigation, Instant translation PS [1], address block location, license plate location, content-based image/video indexing, etc. The very onset of DAR used handcrafted methodologies for various tasks. These methods required application-specific pre-processing and post-processing operations. Also, the handcrafted features were limited by the pre-determined patterns and structures in the data they could process. Thus, they were incompetent in the convoluted innominate environment. The recognition depends on preliminary steps of localization or segmentation of the document into text objects like lines, words or characters. The document images can be printed or handwritten, offline or online. The perspective of Printed DAR does not tune in with Handwritten DAR. Printed documents may have a different set of characters, different fonts or different layout styles, along with graphical objects like tables, figures, charts, logos, etc. The system for printed documents must be capable of addressing these document variations. However, the same approach can be inadequate for handwritten DAR due to the added complexity of handwriting challenges. Different techniques are required to address the issues of Printed and Handwritten DAR. Online handwriting analysis and recognition is an entirely different problem than offline DAR. It is concerned with the dynamics of handwriting. It captures the sequential data, which has spatial and temporal signals of the input signal. This analysis gives us more insight into the writer's specific handwriting features. However, this advantage is limited to Roman and other western scripts. Scripts, like Chinese, have thousands of characters and no specific writing order, making online DAR for such cases challenging. Other scripts, like Asian and middle east scripts such as Arabic and Brahmi, have their respective peculiarities. This script-related nitty-gritty plays a significant role in successfully modelling techniques for various tasks of DAR.

A. Writing Systems of the World

The documentation of information in printed or digital format is influenced by the writing system as follows. We regulate various writing styles associated with one or more spoken languages through a writing system. In the current scenario, there are six writing systems used throughout the world [2], [3]: 1)Logographic system 2)Syllabic system 3)Alphabetic System (Abjads, Abugidas and Pure Alphabets) and 4)Featural system. Logographic systems like Chinese, Japanese and Korean (CJK) scripts visually represent symbols, characters and words. In a syllabic system, each syllable represents a phonetic sound. Japanese scripts follow both logographic and syllabic systems. The scripts like Roman, Greek, Cyrillic, etc., use the alphabetic system to make phonemes of a language. Roman scripts are followed by languages like English, German, French, Spanish, Portuguese etc. Russian, Ukrainian, Bulgarian etc., follow the Cyrillic script. The Roman script is the most widely used, covering North and south America except Canada; western, central and southern Africa; Australia and some European and South East Asian regions. The Abjad system has a right-to-left writing pattern followed by languages like Arabic, Farsi, Urdu, Hebrew, etc. These scripts are mostly used in the Middle east region. This system has only consonant representation and no vowels. Abugidas has both consonants and vowel representations, where consonants are primary and vowels are secondary representations. The writing systems like the Brahmi family in South Asia, Southeast Asia and Tibet with Scripts like Indic scripts (Devnagiri, Gurumukhi, Gujarati, Bengali, Manipuri, Oriya, Tamil, Telugu, Kannada, Malayalam) follow Abugidas. In featural systems, the features of phonemes are the syllables, e.g. the Korean (Hangul) script.

The above classification of scripts explains the ascendancy of a writing system on a language or a script. However, the world's writing systems are ambiguous and are derived from multiple writing systems. For example, English uses logograms such as #, $, & etc., contrasting Logographic

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Manuscript received –

TABLE I
ABBREVIATIONS AND ACRONYMS

DAR	Document Analysis and Recognition	HTAR	Handwritten text analysis and recognition	OCR	Optical Character Recognition	DLA	Document Layout Analysis
HR	Handwriting Recognition	LD	Line Detection	PS	Page Segmentation	WI	Writer Identification
LS	Line Segmentation	On	Onine	Off	Offline	TNC	Text/non-text Classification
TD	Table Detection	Binz	Binarization	IR	Image Retrieval	DR	Digit Recognition
CR	Character Recognition	Rec	Recognition	Clsf	Classification	WSeg	Word Segmentation
LLA	Logical Layout Analysis	DU	Document Understanding	CCA	Connected Component Analysis	CJK	Chinese Japanese Korean
HMM	Hidden Markov Model	GLCM	Gray Level Co-occurance Matrix	HoG	Histogram of Gradients	DWT	Discrete Wavelet transforms
LBP	Local Binary Pattern	GSC	Gradient structural concavity	k-NN	k Nearest Neighbour	RBF	Radial Bias Function
SVM	Support Vector Machine	MLP	Multi-Layer Perceptron	NN	Neural Network	ANN	Artificial Neural Network
CNN	Convolutional Neural Network	DNN	Deep Neural Networks	PCA	Principal component analysis	FCN	Fully Convolutional Network
DNN	Deep Neural Network	RNN	Recurrent Neural Network	MDRNN	Multi-dimensional RNN	MSTDNN	Multi-State Time Delay Neural Network
BRNN	Bi-directional RNN	LSTM	Long Short-Term Memory	BLSTM	Bi-directional Long Short-Term Memory	CTC	Connectionist Temporal Classification
MDCC	Multi-Dimensional Connectionist Classification	MSER	Maximally Stable Extremal Regions	SOM	Self-Organizing Maps	RLSA	Run length smearing algorithm
FOTS	Fast Oriented Text Spotting	ROI	Region-of-Interest	RPN	Region Proposal Network	SSD	Single-shot Multi-box Detector
YOLO	You Look Only Once	Seq2Seq	Sequence-to-sequence	SDSW	Statistical Dynamic Space Warping	MQDF	Modified Quadratic Discriminant Function
DTW	Dynamic Time Warping	NPC	Nearest Prototype Classifier	DFE	discriminant feature extraction	DLQDF	Discriminative Learning Quadratic Discriminant Function
PSNR	Peak Signal-to-Noise Ratio	NRM	Negative Rate Metric	DRD	Distance Reciprocal Distortion Metric	MPM	Misclassification Penalty Metric
Pr	Precision	mAP	mean Average Precision	GT	Ground Truth	B	Binary Image
TP	True Positive	TN	True Negative	FP	False Positive	FN	False Negative
EDM	Entity Detection Metric	SM	Segmentation Metric	CER	Character Error Rate	WER	Word Error Rate
PAW	Part-of-Arabic-words	DIA	Document Image Analysis	PMCOA	PubMed Central Open Access Subset	SV	Signature Verification
BLD	Baseline Detection	SI	Script Identification	SR	Script Recognition	OD	Object Detection
GC	Gender Classification						

systems as they do not match the spoken language. The script-related intricacies have a huge impact on DAR. The early phase of DAR saw script-specific systems with deeper research on Chinese, Japanese and Arabic scripts. Most of the research was on handwritten character recognition (HCR). Traditional DAR systems are designed to model factors such as document layout, line spacing, design and density of information, etc. We must develop automatic and semi-automatic generic systems that adapt to new scripts and scenarios. These systems must learn the peculiarities of various scripts to model intelligent systems. They must address the complexities of scripts, viz, 1) logographic sets of CJK scripts are very distinct from the alphabets and syllables of Alphabetic systems like Roman, Arabic and Brahmi scripts. It's difficult to model such varieties into a single generic system. 2) Vowels and diacritics of a script can introduce confusing patterns even for a model designed for a specific script. However, the non-overlapping character is a common characteristic of all scripts, an advantage for recognition systems. In contrast to the earlier approaches to recognition, recently, systems are being designed for the overall problem of DAR to address the necessity of a generic and robust DAR system.

This paper attempts to present a comprehensive survey of challenges, tools, techniques, and datasets for the DAR problem concerning major scripts used worldwide. The key contributions of this research are enlisted as follows.

- Elaborate analysis of the problem of DAR from various perspectives.
- Comprehensive review of techniques for basic tasks of DAR, viz, Pre-processing, Segmentation and Recognition.
- Comparative analysis of state-of-the-art techniques through tables and figures.
- Detailed categorization of Datasets with categorization into Historic, Printed, Handwritten(offline and online)
- Exposition of different variety of evaluation metrics proposed for various tasks
- A critique of script-related challenges for DAR.
- Insights for potential future research directions for a generic DAR model

The paper is organized as follows. Section II presents an introduction to the problem. Section III discusses the complexities around DAR systems. Section IV reviews various text analysis and recognition techniques. The section is categorized based on basic steps towards the problem of text analysis and recognition. This section also refers to the challenges posed in section III and examines the techniques in detail to resolve the challenges. Section V discusses the datasets which support various tasks of DAR. Section VI covers various evaluation matrices for different DAR tasks. A detailed discussion of script-related complexities and their effect on the DAR is presented in section VII. Finally, section VIII provides insights into the scope of future research directions; thus, we conclude the survey in section IX. The table I has the acronyms and abbreviations used throughout the paper.

II. DOCUMENT ANALYSIS AND RECOGNITION

DAR was originally conceived as an automatic character recognition problem focused on printed text with a specific font. Automatic collection of data, identifying objects and recognizing optically processed characters without human intervention was termed OCR [4], [5]. A mechanical device which could recognize a single character at a time was used for OCR. A photodetector matched the input with the stored templates, thus recognizing only a few small sets of text fonts or handwritten text documents. Due to storage constraints, better structure analyzing algorithms were being proposed for OCR. But these approaches were limited by their processing power and image acquisition options. With growing information technologies, more accurate acquisition devices such as scanners, mobile phones and modern feature extraction and recognition methods, the capability of Modern OCRs on scanned documents has reached beyond 99 per cent. OCRs can convert PDF files, scanned papers or camera-captured document images into machine-readable textual formats. The techniques of modern OCR systems are derived from pattern recognition, machine learning, deep learning, and others [6]. The results generated by OCR support searching and editing within the text. Dedicated and hardwired OCRs are used for specific recognition problems like check reading, bar-code scanning, etc. The software for OCR systems is much cheaper but has lesser throughput rates than hardwired systems. A major limitation of OCR systems is that they are programmed for specific fonts and styles, and any major variation can cause them to fail. Commercial off-the-shelf OCR engines like Tesseract [7], EasyOCR, Ocropus, Ocular, Attention OCR, Doctr, etc., offer exceptional performance. All state-of-the-art systems assume good-quality input documents and are designed for generic use-case applications, making them unsuitable for heterogeneous, unconstrained recognition problems (business-specific documents, handwritten and historical documents). DAR techniques for unconstrained environments are superior to OCR systems, though the latter has advanced in recognizing printed and handwritten (offline) recognition problems. Marrying OCRs with modern machine learning and deep learning techniques may result in a recognition system with immense potential. The past decade has seen research evolve from character recognition to full-page DIA, multi-lingual, multi-font, handwritten text recognition. This study discusses the advancements in DAR. The developments of OCR are out of the scope of this work.

In this work, we have categorized the approaches to DAR into Step-wise and Integrated methodologies. These are discussed in detail in the section IV. Some DAR methods perform step-wise processing and require the pre-segmentation of documents into lines, words or characters for corresponding recognition. It involves breaking down the document into various primitives and consequently re-connect these primitives into desired segments. The inception of the segmentation problem was with the detection and recognition of characters. Character recognition systems' development escalated the character segmentation problem around the 1970s. It raised the need for page segmentation (PS) for extracting lines from pages and, consequently, characters from those lines. With the development of sophisticated document structures, this problem has outgrown recognizing lines and words where the document layouts can be diverse (scientific documents, business letters, official forms, etc.). Also, the documents can have a complex layout (printed and handwritten text). The segmentation task depends on the layouts and types of documents it will be processing. The challenges concerning the text content include noisy documents, skewed text, overlapping/touching content, and curved or broken text. These challenges have been discussed in the section III. Among the segmentation, problems are Page Segmentation (PS), line segmentation (LS), word segmentation (WS) and character segmentation (CS). The problem of LS and WS is relatively easier than the other two.

For online documents, segmentation results can be words, characters or strokes. In this case, the segmentation is internal if it requires recognition or external if it is done without target class information [8]. Spatial (space between words/characters) and temporal (time-outs) clues are used for word segmentation of online documents [9]. The Asian scripts, like Chinese, Devanagari etc., do not require character segmentation as the scripts have well-separated syllables.

Integrated segmentation and recognition methodologies have paved the way for the automatic recognition of text documents. The techniques of end-to-end recognition combine segmentation, feature extraction and classification into a single pipeline and take full-page or paragraphs as input. Recently, the focus has been on training a single end-to-end system to preserve the information throughout to optimize the various intermediate tasks. However, joint optimization of the overall process of recognition is challenging. Deep learning recurrent structures like LSTM, BLSTM, attention-based schemes, encoder-decoder architectures, object detection (OD) and segmentation-free methods have been proposed to handle the issues mentioned above. These methods are discussed in section IV-C2. Earlier efforts were made to make models sequential by joining the feature extraction process with the subsequent models like sequence models: HMM and the representations model: CTC. However, these were trained independently and dealt with already segmented inputs(lines or words).

III. Challenges

With the evolution of processing and storage techniques, more diverse problems of DAR have been addressed. Different documents have different complexities of their own. However, as all the documents to be processed are images, image acquisition and related complications add to the problem of DAR. The techniques have specialized in specific tasks and have achieved state-of-the-art results. However, it becomes challenging when the problem of DAR is considered a whole. There are many obstacles [8] on the road to the evolution of DAR. In this work, we review DAR challenges from three perspectives. 1) Challenges with DAR: covers various complications in recognizing document images. These can be due to text in various languages, fonts and styles, the complexity of the recognition process and the lack of standard formats.2) Challenges with Online recognition: reviews the difficulties triggered due to electronic devices and the underlying technologies like machine translation, OCR, etc. 3) Challenges with text in Document images: covers image acquisition and storage issues with various Online or Offline, Handwritten or printed text. Figure 1 presents some examples of challenges categorized into content variation in documents, image acquisition, and environment-related challenges. These challenges are discussed as follows.

Challenges with DAR

1) Multi-lingual environments: Text can be written in various scripts and languages. English was the first and most successful language o be chosen for OCRs. Languages such as CJK, Arabic and Hindi pose greater challenges to OCR systems [10]. There are thousands of character shapes for Chinese and Japanese. Arabic is written using connected components where the shapes change in accordance with the positioning of the character. The ligatures formed in these languages are vaguely similar to the original ones. Most research focuses on a language-specific model instead of developing a multi-lingual solution. Thus, creating a generic OCR that works well for all can be challenging.

2) Font variations: There are large varieties of fonts in text, each representing a unique way of representing the same character. It is difficult for the OCR to recognize every variation of fonts precisely when many characters are in a class. Some scripts, as in Roman, have overlapping characters, which makes the segmentation task more difficult. [11], [12].

3) Handwritten text [13]: every individual has a unique pattern and style in their handwriting. Also, individual hand-

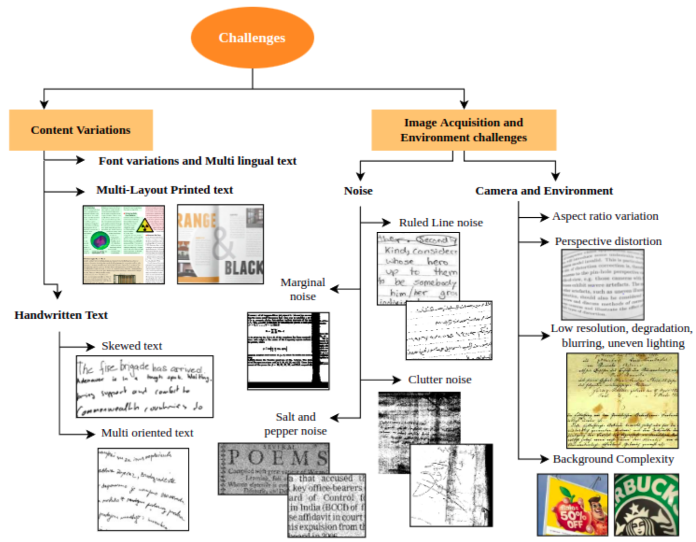

Fig. 1. Challenges With Document Analysis and Recognition 1) Content variations 2) Challenges with respect to text in document images

writing varies from time to time. It isn't easy to design a technique to detect and recognize the text for many reasons. The baseline of the written text is skewed, which requires baseline detection (BLD) and skew normalization. The slant text again increases the complexity of the text detection and recognition system. Erratic hand movements while writing introduce more variations in an individual's handwriting.

4) Dependency on Intermediate steps: Pre-recognition steps like pre-processing, segmentation and classification affect the whole recognition process.

4) Lack of linguistic tools: The research in DAR is very script-specific, and only certain scripts, like Roman, Chinese, and South Asian scripts, have seen major developments. There is a need for language models and linguistic tools in various scripts so that a single generic model can cater to more scripts.

5) Challenging automatic systems: Generic user-friendly automatic systems are difficult. With powerful tools and mechanisms, generic systems for trivial DAR tasks can be expected to achieve soon. However, the systems cannot be fully automatic, and they would require some post-processing to produce expected targets.

6) Need for standardization: Lack of standard formats of printed/handwritten/online documents makes the process of DAR difficult. There are innumerable complex layouts of documents being created, and that too with diverse semantics. The systems cannot be trained to handle such huge variations.

Challenges with Online Recognition [14]

1) Handwriting peculiarities: Handwriting is an individual's characteristic; every individual has unique handwriting traits. This uniqueness causes large variations which need to be modelled by the recognition system. *2) Behavioural and personal factors*: An individual's behaviour also affects writing patterns. For instance, stress, excitement, distracted, laziness, etc., result in significant changes in the positioning and strokes of the writing. *3) Writer-dependent and independent systems*: The scarcity of input data results in systems that have learned some specific data styles and thus become writer-dependent. A writer-independent system requires to have large amounts of data from various writes. This is a major setback in the development of online recognition systems. *4) Idiosyncrasies of various scripts*: The characteristic peculiarities of the scripts are an add-on challenge to the recognition task. For instance, characters like *d* and *l* can confuse the recognizer as they have similar stroke patterns. *5) Learning difficulties for scripts with large character sets*: The scripts like Chinese, Japanese, and Korean, the systems are difficult to train and achieve higher accuracies.

Challenges with Text in Document images

1) Noise: Noise in text images is the degradation of content which can occur either through physical degradation or degradation caused by digitization. Scanning devices also introduce Noise in the image [15]. Numerous research has classified Noise into various types [16] as in 1) Ruled line noise, 2)Marginal Noise, 3) Clutter Noise, 4) Stroke-like pattern Noise 5) Salt and pepper Noise. The ruled line noise arises due to the ruling lines present in the documents. It can be thick/thin, broken, and merged with the characters like *L*, *Z*. The marginal Noise results from document scanning and occurs on the sides of document images. The unwanted content in the foreground of the images is clutter noise, usually caused due to punch holes, pepper noise etc. It makes text segmentation difficult, making text connectivity poor or overlapping the text. In 2011, the work in [17] introduced Stroke-like pattern noise and gave a solution. Salt and pepper noise is caused due to dirt on the conversion of documents.

2) Camera and environment challenges: As images are captured through various sources such as digital cameras, PC-cams, cellphones, scanners etc., they suffer from low resolution, blur, perspective distortion, background complexity, uneven lighting, blurring and degradation, and aspect ratios. For instance, the camera-based images give better recognition than scanner-based input [18]. In contrast, digital cameras are very convenient for image acquisition, but they cause other challenges of geometric distortion, focus loss and uneven document lighting. The low resolution, blur and uneven lighting make even simple text detection difficult. Perspective distortion occurs when the image is not captured at a plane parallel to the image. The resulting images have distorted characters. Images captured from a non-planer surface cause curved lines. Compressed images are challenging for text recognition as they tend to lose the sharpness required.

A generic text analysis and recognition system could solve the above challenges. Early research was usually using manual methods of feature extraction. These methods could only cover part of the scope of the problem. The focus is on these challenges individually. However, recent times have experienced rapid developments in pattern recognition and image processing. The motivation behind this development is multi-fold. 1) High-performance computing systems 2)Increased applications 3) Large-scale recognition networks.

IV. Tasks of Document Analysis and recognition

DAR has been addressed in literature by various methods and systems. However, all methods can be categorized into

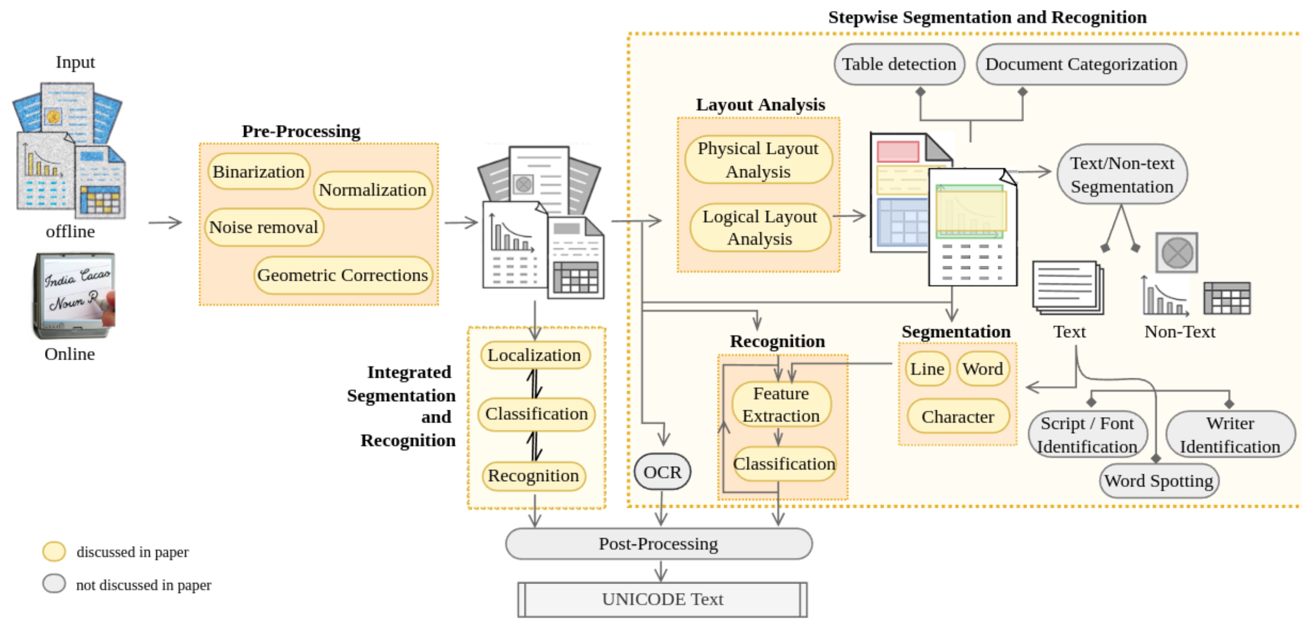

Fig. 2. Tasks of Document Analysis and Recognition with major categories as Step-wise and integrated Segmentation and Recognition

three classes of tasks: 1) Pre-processing (Section IV-A) step eliminates the unwanted elements and improve the quality of images, 2) Segmentation (Section IV-B) steps deals with pre-processed documents and extracts various textnon-text elements and lastly 3) Text Recognition (Section IV-C) performs feature extraction (Section IV-C1) and classification (Section IV-C2). The output of recognition requires post-processing for producing machine-readable UNICODE text. The methods of post-processing are out-of-scope for this study. Figure 2 presents various tasks of the DAR process and their interrelations. Figure 2 presents a flowchart to represent the flow of various tasks of DAR and their inter-relations. It represents all basic tasks of DAR, from acquiring inputs (online or offline) to producing machine-interpretable UNICODE text. This work covers only basic tasks of DAR like pre-processing, layout analysis, segmentation, and end-to-end recognition. Tasks like OCR [4], [6], [7], [19], Text/non-Text segmentation [20], [21], Table Detection (TD) [22]–[24], Word Spotting (WS) [25]–[27], Document Categorization (DC), Script Identification (SI) [2], [3], [28], [29], Writer Identification (WI) [30]–[33] and Post-processing have not been considered. There are some specific topics which have been researched with respect to the DAR, such as, Forgery detection [34], Image quality assessment [35], Strike-off removal in handwritten scripts [36], [37] etc.

A. *Pre-processing*

The first step towards a text analysis and recognition system is pre-processing of images [38]. The pre-processor techniques eliminate the unwanted noise and other issues in the image before they are fed into the analysis and recognition system [38]. It enhances the images and converts the image into suitable format [5]. It handles the image acquisition challenges, as discussed in section 1. The basic purpose of a pre-processor is to produce an image ready for feature extraction such that it improves the recognition capability of the text recognition system. Images need correction of the artefacts which deviate from the standards and enhancement of the image according to the feature extraction and recognition algorithms used in the system . Corrections fix the issues, and enhancement methods aim to optimize the image for a specific architecture. Various researchers have categorized the sub-problems under pre-processing differently. Pre-processing online handwritten documents is concerned with removing noise due to the sensor or resolution of digital device, abrupt strokes [9], and random dots. This paper discusses pre-processing techniques like Image correction, enhancement, and compression.

1) Image correction

a) Noise reduction: The noise reduction methods [39] can be broadly viewed in the following points. 1) Filters are designed for thresholding, smoothing, sharpening, removing background textures and contrast adjustment purposes. 2) Morphological operations create data-specific features similar to noise, translated over the image to locate the ruled line or salt and pepper noise. Due to the dependency on the data, these cannot handle variations. 3) Hough transforms, unlike morphological operations, need not require apriori knowledge of the data. It transforms points of the image space into a line or curve in hough space with the help of a mapping function. The properties of hough space help to determine similar pixels of the line or curve (noise) [40]. 4)Projection profiling creates projections (sum of pixels) across the axis for 2-dimensional images and horizontal and vertical profiles. These methods are sensitive to rotations, and characters with horizontal strokes like L and Z lose their information. 5)The method to reduce marginal noise involves detecting and recognizing the noisy component and then identifying the boundary of the concerned text. [41].

b) Normalization: Normalization refers to correcting text to remove all variations in the text, done mostly for handwritten text and scanned text. Normalization techniques are used for Skew correction, slant normalization, contour smoothing and size normalization. The projection profiles, nearest neighbour clustering, cross-correlation method between lines, and Hough transforms are used for baseline extraction. After the baseline is detected, some rotations are applied. This solution is applicable in different sequences for skew, slant and size normalization. Counter-smoothing reduces the number of sample points needed to represent data.

c) Lighting correction: The problem of shadows, unstable and uneven lighting distribution and uneven colour intensities fall under the category of environment-related challenges. Line-based Empirical Mode Decomposition [42], nonlocal variational models [43], and other methods using filtering are used for Lighting correction [44].

d) Geometric corrections: Image auto-rotation, perspective correction, and non-linear image transformations(curved text) involve BLD methods to extract the baseline and then apply rotations etc., for further corrections [38].

2) Image Enhancement

a) Enhancing illumination, blur and focus: Images can be enhanced using Global illumination methods like simple LUT remapping, pixel point operations, histogram equalization, and pixel remapping or Local illumination methods like gradient filters, local histogram equalization, and rank filters. Sharpen filters can enhance the edge features before gradient computations. For blurred images, the scale-space pyramid-building process is used. It involves applying a Gaussian blur filter to the sub-sampled images to remove the jagged artefacts.

b) Background estimation: The background estimation is mostly accompanied by binarization of images, for instance, median filter with large window size [45], Morphological grayscale closing [46], heuristic estimation [47]. Classic techniques of binarization like Niblack [48], and Sauvola [49] used high recall initial binarization followed by inpainting of foreground regions.

c) Grayscale conversion: A grayscale image is generally a prerequisite of most binarization methods. A most basic conversion of RGB to Grayscale is computing luminosity by $g = 0.21r + 0.72g + 0.07b$. However, this is not fruitful when multiple colours are in the foreground. Principal component technique [50] is also applied over RGB pixels. The first component of PCA is mostly the background pixels, which result in a high-intensity background and low-intensity foreground.

3) Image Compression

a) Compression techniques: refer to transforming an image from space domain to space domain to retain the shape information. The compressed image has reduced size helps to reduce the computational complexity of the feature extraction and recognition network Thresholding and to thin are the compression methods. With thresholding, [51], [52], values below a threshold are discarded. It can be local when the threshold is computed for a local region and globally, where the region considers the whole image. Thinning is used to extract the shape information of the characters. It is based on pixel-based and non-pixel-based methods. The former iterates on a local region to obtain skeleton and non-pixel methods, later using global methods.

The pre-processing mostly implies image processing techniques for various corrections and enhancement operations. These image techniques depend largely on the feature extraction methods and the application of images, as every extraction model has a different objective based on its application. Some application-based pre-processing methods are: Local binary feature (pixel intensity comparison) descriptors. Spectral descriptors, e.g. SIFT/SURF, for methods which use image pyramids. Basis space descriptors, e.g. Fourier Methods, wavelets, Slant transform, Walsh Hadamard, KLT. These methods transform the data into another domain for analysis. Polygon Shape descriptors are used to extract shapes from images. However, these are of no use for text document images.

4) Binarization

Document binarization involves dividing a document image into two groups: foreground pixels and background pixels. The foreground pixels are turned black, and the rest are white. The binarization process makes document analysis easier by shifting focus on significant data portions. This step is crucial for severely degraded documents such as historic documents. The degradation can be due to complex backgrounds, multiple font colours and sizes, stains and creases, etc. The binarization can be permed globally or locally. A single value is computed for global thresholding, while local methods use adaptive thresholding. One of the most basic algorithms is given by Otsu [53]. It derives the threshold T_{Otsu} by computing a grey-scaled image histogram ($L = 256$ bins for 8-bit). The research works [52], [85]–[87] have presented elaborate discussions on techniques of binarization in literature. Table II summarizes techniques of binarization over the years. It discusses various categories of binarization with their pros and cons. Most of the early methods of binarization were generic and did not consider the document-specific nitty gritty. Global thresholding methods don't perform well in these cases. For a while, image processing models were widely used. With the advances in deep learning, the techniques with CNN and related architectures outperformed the former ones. The downside of these models is the requirement for better training hardware and large data. The trend in literature reveals that parameter tuning is also significantly efficient in aiding many algorithms. Also, pre-processing and post-processing steps improve the performance of the overall systems. Table II presents a brief evolution of techniques from binarization with Otsu [54] to the modern methods like machine learning (SVM [67], [68]) and deep learning methodologies (FCN [69], GAN [75], [76], LSTMS [77], [78], U-Net [74]). However, there is no single method which works perfectly for every document. Different techniques are used for different specific cases. Thus these methods have to be tried and tested for better performance.

B. Page Segmentation

The segmentation of a document results in physical and logical components. A crucial problem related to PS is Document Layout Analysis *DLA*. Before text segmentation, we need to understand the document's layout and extract the textual component for further segmentation [8]. There is no general

TABLE II
BINARIZATION METHODS AND RELATED RESEARCH WORKS FOR PRINTED AND HANDWRITTEN DOCUMENT IMAGES

Category	Method	Description
Global threshold	Otsu: [53] $k = \frac{\sigma^2 B}{\sigma^2 G}$; $\sigma^2 B$=global variance, $\sigma^2 G$= between class variance	$+$ Maximize between-cluster variance of pixel intensity; $-$ It produces images with some background pixels being darker than foreground pixels; imperfect results
Local threshold	Niblack [48]: $T(x, y) = m(x, y) + k\delta(x, y)$; std. dev. δ, mean m, k=-0.2 ;Local threshold based on window mean & std. Dev.	$+$ adaptive threshold can handle intensity distribution overlap of the foreground and background pixels $-$ when window covers only background pixels, darkest background pixels are set to foreground #: Sauvola [49], AdOtsu [54]
Edge Detection	Su [55]: Edge detection with Canny edge detector; Local thresholding near edges: $C(i, j) = \alpha \frac{M(i,j) - m(i,j)}{M(i,j) + m(i,j) + \epsilon} + (1 - \alpha)(M(i, j) - m(i, j))$; contrast: $C(i, j)$, image intensities(max: $M(i, j)$, min: $m(i, j)$)	$+$ Filters canny edges using local contrast, Local thresholding near edges after background removal, $-$ detected edges do not completely surround the foreground regions, #: Jia [56], Valizadeh [57]
Image transforms	Sehad [58]:Performs background removal using Fourier Transform	$+$ image transforms based local features: Gabor filter banks, Fourier Transform, Contourlet Transform, # Nafchi [59]
Mixture models	FAIR [60]: Ensemble of MoG with post-filtering; prob. dist.: $p(x) = \sum_{k=1}^{K} \pi N(\mu_k, \Sigma_k^2)$; k= mixture components3	$+$ MoG can converge to any arbitrary distribution, $-$ log-normal distributions provide a better fit and performance compared to MoG,#: Mishra [61], Mitainoudis [50]
Conditional Random Fields	Howe [62]:Local energy functions score how well a given binarization agrees with input; $E(B, I) = \sum_{i=1}^{N} E_i(B_i, I) + \sum_{i=1}^{N} \sum_{j=1}^{i} (B_i, B_j, I)$; I:input(N pixels), B:discrete segmentation of I, Ei, Eij: decomposed unary & pairwise energy functions	$+$ CRFs are excellent for including spatial dependencies among pixels into the binarization process; $-$ CRFs with handcrafted energy functions do not learn parameters, #: Ayyalasomayajula [63], Ahmadi [64]
Game theory	GiB [65]: Extracts features for clustering using game theory	
Shallow ML	Kasmin [66]: Ensemble of 8 SVMs	#: Hamza [67], Rabelo [68], Pastor [69]
Deep Learning	*FCNs*: Pastor [70], Calvo-Zaragoza [71],PDNet [72], *Morphological networks with grayscale dilation and erosion operations*: Mondal [73], *U-Net architecture with attention layers:* Kang [74]; *GANs:* Tensmeyer [75], Zhao [76]; *LSTMs*: Afzal [77], Westphal [78] LSTMs	$+$ Deep learning methods outperform a non-convolutional deep MLP (including histogram and median filter features; $-$ Less availability of labelled data for training; Recurrent networks (LSTM, BI-LSTM) not as effective as FCN; # Bhunia [79], Krantz [80]
Parameter tuning	*Supervised tuning*: Xiong [81], Messaoud [82]; *Unsupervised tuning*: Ntirogiannis [83], Liang [84],	$+$ Finds best parameter settings for algorithms, $-$ Output of algorithm dependant on these parameter settings

$+$ Pros, $-$ Cons, # Related research

layout analysis procedure, as the document types are varied. With this information, there are two levels of segmentation. First, a document is segmented into text and non-text regions (graphics, tables, etc.). Further, the text region is segmented into paragraphs, lines, words, and characters.

1) DLA: An assumption of content in the document is necessary to proceed with DLA. A document has pages with components like text blocks, graphics, tables, etc. The document layouts can be categorized based on the organization of the text block. Figure 3 shows various categories of documents like Regular, Manhattan, non-Manhattan, Complex, and Overlapping [8], [88]. The black bounding boxes in the figure represent the expected segmentation of different layouts. The layout with single, large rectangular zones of text falls into Regular, while the one with multiple rectangular zones is the Manhattan category. Irregular-shaped text portions are the non-Mannattan Layout category. Some documents have overlapping texts or show-through effects. These are categorized as Overwritten Layouts. Complex category documents can have arbitrary text shapes; thus, it's the most challenging category. The segmentation problem on Rectangular and Manhattan layout documents has attained remarkable results. The non-Manhattan and overlapping layout styles, like handwritten documents [89], noisy document images, historical documents, etc., are challenging to match the state-of-the-art results of the Manhattan style. Some other factors affecting layout analysis can be summed into 1) Clean vs Degraded: Any document can have degradations like noise, blur, ink spots etc. These degradations can be in printed documents due to scanning and long storage durations or in handwritten documents like historical documents. 2) Text-line variation: The orientation of text lines plays a significant role in layout analysis. The lines can be parallel (vertical, horizontal or multiple angles) or arbitrary with freedom of shape.

To begin with DLA, an assumption of some background information is essential. This information can be classified into five categories as: 1)Analysis Document: *Printed* or *Handwritten*, 2)Analysis Object: *Foreground* or *Background*, 3)Analysis Strategy: *Bottom-up* or *Top-down*, 4)Analysis Layout: *Overlapping* or *Non-Overlapping*, 5)Analysis primitives: *pixels*, *CC* etc. Figure 3 shows the categorization of various physical and logical layout analysis approaches as discussed above. It also extends the physical layout analysis to the LLA, which concerns object classification and relationship detection, which are discussed further. Segmentation aims to detect the text components with a precise location to group them into candidate text regions with as minimum background as possible. The early approaches towards it were: CCA and the sliding window approach. CCA helps to extract CC from the text. It works on binarized images to create a graph by classifying every pixel on the image based on colour and edge features. These CCs are further classified for text/nontext. [90], [91]. Some techniques for CCA were MSER and MSER++. [92], [93]. MSER are proficient in detecting images with noise, low resolution and low contrast. The problem with MSERs was that they resulted in repetitive components being removed before further steps. Sliding window/Region-based methods work by analyzing the whole image using some small window which slides over the input [94]. The results of region-based methods are dependent on the initial segmentation result. If the initial segmentation is faulty (e.g. may have the text and non-text region), the further classifier will identify it as text or non-text. The Connected component method works to overcome this erroneous classification for an entire region. Another similar method is a Voronoi diagram [95]. It uses connected components to compute Voronoi points and edges leading to a neighbourhood graph. It is a time-taking solution; therefore, it is not used for high-resolution documents. Also, these methods give better results when text appears with non-text, e.g. images. These methods mostly suffer from the issue of touching components. The approach of starting from the primitive components of a page, like pixels or CCs, and then merging these primitives until we reach page level is called the Bottom-up strategy. Other solutions to segmentation with a bottom-up strategy include Texture analysis, Machine learning and deep learning solutions. In texture analysis, the primitives are mostly homogeneous regions of pixels which makes it computationally expensive. ML methods like SOM [96] and SVM [97] have been used for DLA with the most focus on ANN [98]. The approach to segmentation, from the page level to lower levels like paragraphs, figures, tables etc., is called Top-down. Some methods with this approach are: Projection-based methods, White-space analysis [99], RLSA [100], etc. The projection-based methods create vertical or horizontal projection profiles of a page, e.g. Recursive XY cut algorithm [101]. Some White space analysis methods use the

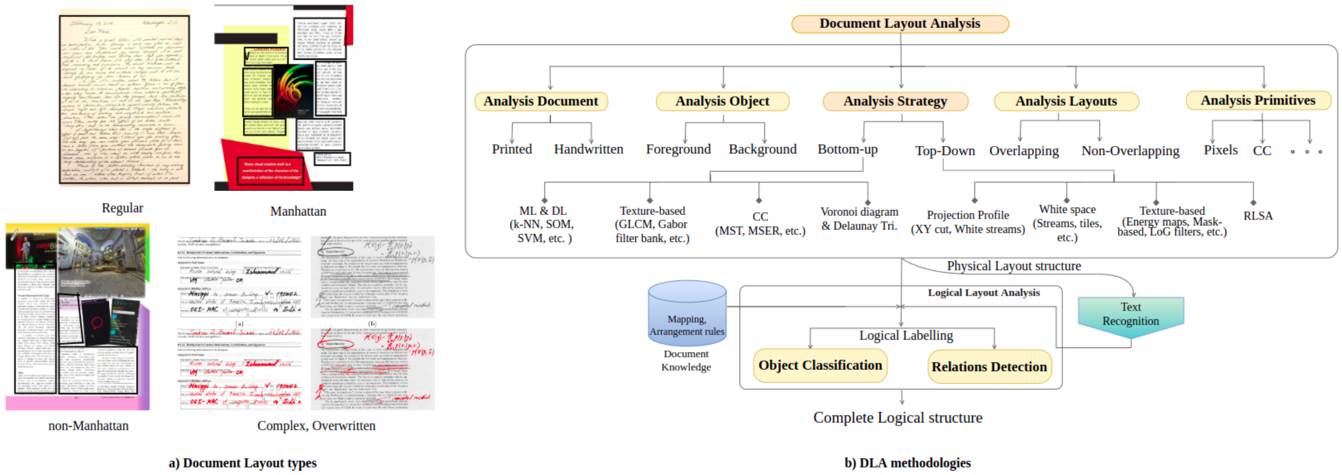

Fig. 3. Document Layout Analysis a)Document Layout types b) DLA methodologies: Physical Layout Analysis techniques followed by Logical Layout Analysis techniques

Recursive XY cut algorithm to find isolated white spaces. A distinct method to analyze a page is smearing-based methods like RLSA [100]. It smoothens the pixels which are below a defined threshold. This method cannot be applied to cases with high variation among pixel intensities, like a handwritten document. A class of documents with overlapping layout styles uses a different segmentation strategy. Mostly top-down strategy with texture-based analysis is the focus here. Feature extraction is used to determine the feature of each pixel and classified into text and non-text. A work in [21] uses Gabor filters to extract features. This method can process documents with other layouts and skewed text as well. Wavelet-based methods are also useful for extracting features for text/Non-text classification (TNC).

With all the various segmentation methods, no generic method is applicable in every case. Before choosing a strategy, an assumption of content is required for better results. For example, documents with a simple rectangular layout can be segmented with good accuracy by algorithms like Voronoi diagram [95], Docstrum [102] etc. Computationally expensive algorithms like smearing-based [100] and projection-based can be applied to a down-sampled version for high-resolution documents. Computationally cheaper methods, such as CC based: MST [103], [104] and Voronoi [95], are better when there are larger primitives in the document.

DU, also termed LLA, is a semantic analysis of documents [8]. It extracts the logical components of a document or information relating to some context, such as a name, dates, etc. It's crucial to understand the contents of the documents to enable the systems for automated processing of documents. The logical structure is significant for many applications like the digitization of books, requiring navigation and search facilities [105]. A DU system is modelled as a physical and logical part. The physical layout of the document is extracted and used to assign logical labels with the help of document knowledge. Logical labelling depends a lot on the content of the documents, e.g., for developing digital libraries, labelling includes identifying headings, titles, footnotes etc. The information regarding the content is provided by *Document Categorization* [106]. The techniques of logical labelling can be summed as Rule-based approaches, Syntactic Methods, Perception-based methods, Learning-based methods, and Knowledge-based systems [8]. The state-of-the-art methods have achieved impressive performance on many Datasets available for the purpose. However, research is still limited to limited domains. There is ample scope to develop generic systems and improve accuracy.

2) Line, word and Character Segmentation: Text segmentation involves segmentation into lines, words or characters. It can be performed directly on the pre-processed document images or the text regions identified by layout analysis. In both cases, layout analysis techniques are also useful for line, word or character segmentation. LS involves the localization of text lines in a document image. It can be used further for word segmentation, spotting, recognition, etc. The lines in a document may be overlapping/touching/broken lines, lack of proper baseline, curvilinear lines, etc. Also, the number of lines is not known at the start. The task becomes more complicated with difficult layouts like complex or overwritten, handwritten and historic documents. The early segmentation methods used basic techniques like CC analysis, MSER, Voronoi diagram [95], [107]; projection analysis [108]–[111]; energy functions [112]–[114] etc., to determine the text groups. Figure 4 shows how the projection profile method can be applied for various scripts to segment the lines. These techniques have shown good results but are specific to their application domain. Consequently, various solutions like pixel-level segmentation and OD have used deep learning methods. Both categories use CNN architectures to extract features from the input document image. The pixel-level segmentation uses the sliding window method to classify each pixel into two categories: foreground (text) and background, by combining similar pixels. This approach has some drawbacks. All the lines detected are represented in a 2D bounding box of the same orientation. It produces faulty predictions when the lines need to be well-spaced (overlapping/touching) or are in a non-linear orientation. Also, this approach's cost of detecting objects(lines) is high, as every possibility is checked. On the other hand, OD approaches have successfully predicted the curvilinear bounding box for curved text lines [115]. The OD model

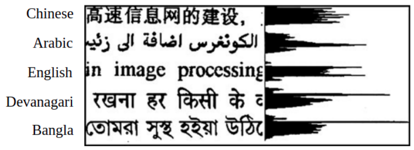

Fig. 4. Horizontal projection profiles of different printed scripts [8]

uses a pre-trained CNN for feature extraction and an RPN to identify the probable objects of interest at a much lower cost. Extensive research is done in this regard for a variety of research problems.

Table III discusses various literature works for line, word and character segmentation and their applications. The segmentation of text into lines is a subjective problem as the documents with printed text are quite easy to segment with basic techniques like projection profiles as shown in figure 4. It works for most of the scripts in a similar manner. However, the segmentation of handwritten documents, particularly historical documents, is difficult to handle. Several approaches have been proposed for such complicated cases like Region growth techniques for Historical documents [116]; Probability Density for Printed [117] and handwritten texts; Smearing based methods for printed [118], historical [119] and handwritten documents [120]; Dynamic programming for Online documents [121]; Grouping-based methods for printed, handwritten [122]; Hough transform for printed [121], handwritten documents [123], [124]; Graph and MST for printed [125] and handwritten documents [126]; Energy maps for handwritten [113] and historical documents [112], [114]; Fringe maps for printed and handwritten [127]; Watershed & Flooding for printed, handwritten documents [128].

Word segmentation is a script-specific problem. In some scripts like CJK, the separation of words is less distinctive than in other scripts like Roman, Brahmi, and Arabic. Therefore CJK scripts mostly consider only character segmentation. The approach to recognition is also different based on the technique of recognition. The word can be further segmented for character recognition, or it can be considered as a single entity for holistic recognition. Both ways have their advantages and disadvantages. Holistic recognition can be computationally useful for removing the intermediate step and also avoiding errors of the same. However, it requires a large dictionary for recognition. Different distance metrics to compute the distance between CCs (Euclidean distance, Convex hull, bounding box, average run length) have been proposed for word segmentation in handwritten [129]–[131] documents. Recognition-based approaches have been proposed for printed [132], handwritten documents [130], [133].

In resolution to the problem of character recognition in the early phase of DAR, research progressed towards character segmentation after extracting text lines. Since then, the problem has been addressed in traditional and modern ways. The research gaps that persist are [8] 1) Touching characters in printed, handwritten documents, 2) Broken characters in printed documents, 3) Lack of baseline in handwritten documents, 4) Variety of fonts and styles in printed documents, 5) Different orientations of text, etc. These gaps have been addressed by following techniques. Probabilistic algorithms like HMM find non-linear paths between overlapping text lines and handwritten documents in [134]. Level set methods are proposed to separate the printed, or handwritten text regions from the background using partial derivatives and an external vector field in an iterative process in [135], [136]. Segmentation-based (Explicit) solutions have solved the touched characters problems [137], [138] and Recognition-based (implicit) techniques [139], [140].

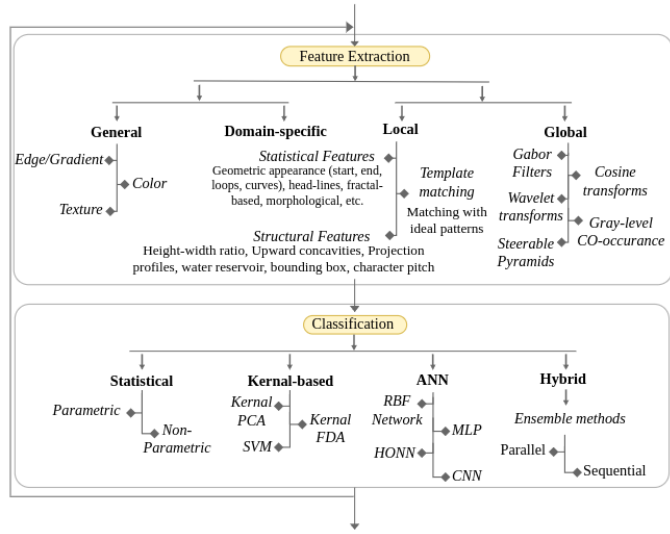

Fig. 5. Text Recognition classified into Feature extraction and Classification approaches

C. Text Recognition

The recognition of text from document images corresponds to observing characteristic patterns of text elements and utilizing the features of these elements for recognition purposes. The classifiers use these features as feature vectors, graphs, strings of codes or a sequence of symbols. A recognition system's performance depends on extracted features' quality, which is further based on the pre-processing and segmentation steps. A character recognition system classifies character segments into an associated class followed by word decoding models. Additional knowledge like Language Models and a recognition system for meaningful recognition results may be useful. Segmentation is not necessarily approached by every recognition method. Segmentation-free recognition methods consider text elements (lines/words) as a whole entity. Most systems convert the text segments like (lines/words) into smaller components such as words and sub-words, as its infeasible to categorize a large segment into classes. The smaller components undergo classification, and the consequent results are concatenated for predicting lines/words. The similarity scores, like probability, etc., are compared with lexicon entries and further used to test the predicted patterns with respect to the ground truth labels. There are different recognition strategies based on the recognition scripts' vocabulary sizes. Applications with a large vocabulary provide better recognition of the sub-word model. Dynamic programming approaches have been used to optimize the problem of finding the best match between the observed scores and the word models.

1) Feature Extraction: The pattern features have been widely studied to solve the problem of DAR. Feature extraction involves transforming input data into informative features

TABLE III
SEGMENTATION METHODS FOR TEXT ELEMENTS LIKE LINES, WORDS AND CHARACTERS AND RELATED RESEARCH WORKS

Method	Description	Method	Description
Line Segmentation			
Horizontal Projection	Histogram of foreground pixels for horizontal lines: Pri [110], hw [109], [111]	Region-growth	Sub-groups of similar neighbouring pixels; bottom-up buildup procedure; computationally expensive: Hist [116]
Probability Density	Discrete estimate of probability density; computationally expensive: Pri [117]	Smearing based	smearing of consecutive black pixels along the horizontal direction and for a distance less than the threshold, white spaces are filled with black ones: Pri [118], Hist [119], handwritten [120]
Dynamic programming	global minimization of segmentation cost function: On [121]	Grouping	aggregating primitives in bottom-up approach: Pri, Hw [122]
Hough transform	helps to find the skew angle: Pri [121], Hw [123], [124]	Graph-based	MST, other graph-based solutions: Pri [125], handwritten [126]
Energy maps	Energy map of pixels helps to determine the text and non-text seams: HW [113], Hist [112], [114]	Fringe maps	Fringe numbers are assigned based on the distance to the nearest black pixel: Pri/handwritten [127]
Watershed	Flooding the morphological surface and segmenting the input into catchment and basin: Pri, handwritten [128]	Hybrid	Combination of above approaches: Hist [141]
Script-specific: Hist [112], [114], [115], [119], [129], [141], [142], Urdu [143], Persian [109], Chinese [124], [126], English [144], [145] Brahmi: Hw Devanagari [142], Pri Devanagari [146], Bangla [147], Gurmukhi [111]			
Word Segmentation			
Distance metric	Different metrics to compute the distance between CCs (Euclidean distance, Convex hull, bounding box, average run length): Hw [129]–[131]	Recognition	Feedback from recognition system; Finds word boundaries based on classification algorithms (Scale space, k-means clustering, Hough transform): Pri [132], Hw [130], [133]
Script-specific: Off [129]–[133], Brahmi: offline [147], [148]; Chinese: [149], Arabic [150], [151], Pri [152], On [153]			
Character Segmentation			
Stochastic	Probabilistic algorithms (HMM) used to find non-linear paths between overlapping text lines: HW [134]	Level set method	separate the text regions from the background using partial derivatives and an external vector field in an iterative process: Pri [135], Hw, both [136]
Touched characters	Segmentation-based (Explicit) [137], [138]	Touched characters	Recognition-based (implicit) [139], [140]
Script-specific: Hist [154], CJK [155]–[157], Arabic: [150], [151], Brahmi [147], [154]			

that can be processed by machine learning or deep learning algorithms for solving classification problems. It preserves the information in the original input dataset and maps those features on a lower dimension. The features can be general, such as colour, shape, edge and texture based, or domain-specific, as shown in figure 5. Domain-specific features are application-oriented peculiar features. However, both can be further categorized into Local and Global features [29]. The local features draw intrinsic features from local structures like characters or words and can be Statistical, Structural, or Template matching [3]. PCA has been used to extract feature vectors [158]. However, the input should be pre-processed to get better features. DNNs are very successful in DAR for feature extraction, and classification [159].

GENERAL FEATURES

a) Colour-based features: These are mostly useful for applications such as Advertisements, book covers, magazines, posters, handwritten text images etc. These are sensitive to image acquisition challenges such as illumination, brightness etc. [160]. The work in [161] has used the colour clustering method to the image in the pre-processing step, then applied a connected component approach for text extraction. Another work [162] Uses colour separation methods using colour maps to segment the text touching the graphic component of the image. A 3D colour histogram approach is presented by [163], which is used then to cluster the edge map samples of document images. The mean shift-based procedure obtains the reduced colours.

b) Edge(Gradient)-based: The approach based on edge/gradient works around the text gradient against its background. The abrupt pattern changes cause significant gradient change, which is used as the feature to extract these regions.Some work in this direction are as [164], [165] [166]. The work in [165] assumes the text is in horizontal lines only and applies the straightness property to identify edges. [164] computes an edge histogram of the images and then uses binarization and minimum distance classifier to classify text characters using a shape-based feature. The HOG [167] method computes gradients over many non-overlapping sub-regions of the image. The histogram of these gradients is drawn after some normalization. It was proposed first for human detection. HOG is robust to photo-metric challenges such as illumination and other geometric issues. Thus its widely used for Text extraction purposes. [168], [169]. The handwritten digit recognition is presented by [168]. Handwritten Devanagari character recognition is proposed by [169].

c) Texture-based method: These methods are applied mostly for dense documents, mostly coupled with region-based classification methods, e.g. Statistical features (mean, median, modal values) for different characteristics of an image. GLCM is a matrix representing the count of pixel intensities occurrence in an image. This is then used to analyze textures in an image. The initial feature extraction in cites sethy2019gray uses GCLM as other texture feature descriptors such as Wavelet transforms. Wavelet transforms like DWT is used to transform from the image domain to the frequency domain. It results in the multi-resolution representation of the input image. The work proposed by [170] uses DWT in multiple levels, first for detecting the edges and further for differentiating between text/non-text pixels. Various methods have used Wavelet transforms as in [171]. Some other texture features which have been utilized in the texture-based analysis of text images are as run length-based features [172], white tile-based texture feature [173], LBP-based features [174], transitional features [160] and spatial features [175]

DOMAIN-SPECIFIC FEATURES: Different scripts possess unique characteristic features that must be addressed to develop an efficient recognition system. These are domain-specific features and require some special analysis and extraction techniques. The elements of the handwriting, such as strokes (horizontal, vertical or diagonal), loops, ascenders and descenders, inclination etc., are descriptive features of scripts. Such cases require apriori knowledge of the unique features to address them explicitly. Component-based feature analysis is done to extract script-specific components. For instance, the Devanagari script has a specific feature: *shirorekha*. Arabic scripts have *dots* and *diactrics*. These components are extracted before the recognition process and are added in the post-processing step to the final result for better recognition results.

LOCAL FEATURES: Local features are concerned with character-specific nitty gritty like statistical, structural, morphological, and contour-based features. The statistical features examine the mathematical elements like the circularity of the component, contour area, standard deviation, aspect ratio etc. The techniques used are like Upward concavities [176], Projection profiles [108]–[111], Bounding box, character pitch, etc. Histogram features have been used in various forms like GSC features, Directional element features [177], Percentile features [178], etc. A weighted average or 'moment' of character pixels is also useful for describing local features like scale and rotation invariant. Another technique for determining local features is Zoning. Various features are proposed in the literature for extracting information from the zones of an image. The structural features of text, like start, end, loops, and curves, are distinctive characteristics that, when combined, represent a character. HMMs have also been used to explore these structural elements' variations [179]. Vectors representing shape-based features are another way of structural analysis. Various vector representations like run-length information, black-white transitions, partial derivatives, HoG, Graph edit distance, etc., have been proposed. The template matching approach computes features of a local component (character or its sub-parts) as a whole. The features are compared with a pre-defined template; therefore, this approach can be used only for printed or a limited domain of handwritten documents.

GLOBAL FEATURES: The global feature extraction methods are DWT, Gabor features, and feature pyramids. These are needed mostly for OD in contrast to the local feature, which is useful for recognition. Projection histograms were the earliest solution for feature extraction and were mostly used for segmenting printed text. Figure 4 shows horizontal profiles of various languages. Gabor filter is a linear transformation of input into arbitrary orientations. These are useful for highly degraded document images [180]. Other methods of global feature extraction are 2D DWT with morphological variations, Image Moment variations, Steerable Pyramids, Cosine transforms, etc. These methods are more robust to noise and are faster than local methods.

2) Classification: Feature extraction and classification go hand-in-hand with each other. The classification of extracted features into pre-defined classes was the recognition system of early DAR. A character recognition problem for the English language has ten digits classes, 52 upper/lower character classes and 62 alphanumerical classes. The classification methods can be supervised or unsupervised depending on the objective of the problem at hand. The similarity of features of some classes, e.g. 'o', "0", "O" is challenging for classifiers. The choice of feature extraction technique is crucial for classifier performance. Also, the feature extracted can be variable or fixed-length, requiring different classifiers. Some popular classification models include *k-NN*, *CNN*, *SVM*, *GMM*, *HMM based* [181], *NN based* (MLP, PNN), *Tree-based classifiers*. All these methods can be categorized into four classes of classification as shown in figure 5. The statistical methods use a Bayesian classifier-based measurement procedure to determine some metrics used to define the object's class [182]. There can be parametric such as HMM and non-parametric methods such as SM, NN, Nearest neighbour, Bayesian classifiers, etc. The parametric methods have been mostly used in earlier work with limited data requirements. However, the approaches had restricted decision boundaries which tend to fail for complex pattern classification. In these cases, non-parametric methods were useful as they were more flexible with the data requirements and decision boundaries.

Techniques of Text Recognition

OFFLINE RECOGNITION

The task of text recognition has been approached using different techniques. These can be categorized on how they model the dependencies (visual or sequence). With this perspective, we have grouped the techniques into three categories:

1) Modelling Visual Dependencies (Images)
2) Modelling Sequence-to-Sequence Dependencies (Sequences of feature maps)
3) End-to-end modelling

1) Modelling Visual Dependencies (Images)

We aim to extract the content(text) from the document images in DAR. It includes extracting specific features corresponding to target labels. Various techniques have been proposed to model the recognition of the visual patterns in the data. Some of these techniques are discussed below:

a) k-NN: k-NN is a non-parametric technique which categorizes objects based on a distance metric. An unseen object is labelled to the category of its closest neighbour. k is the number of features to be considered, which decides the number of neighbour categories. k-NN of the connected components can be used to analyze the document structure analysis. This technique of classification makes local predictions which can result in noisy outcomes. Hence, global information is accompanied by this method for more reliable results. k-NN combined with proper feature extraction methods like GSC, etc., have succeeded in character recognition [169]. Due to a large number of training samples, it is inefficient in languages with very large character sets (e.g. CJK).

b) SVM: SVMs have been widely used for classification problems like text categorization, character recognition, etc. The objective of SVM is to find an optimal hyperplane or a set of hyperplanes (equation 1) which classifies between two or more categories. The hyperplane is optimal when it gives the best separation by finding the maximum marginal distance between the classes, and there are no miss classifications. For a binary classification problem, all points in class A will be valued as +1, and Class B will be valued as -1.

$$f(x) = w^t.x + b \quad where, \quad x \,\epsilon \begin{cases} A & if\ f(x) > 0 \\ B & if\ f(x) < 0 \end{cases} \tag{1}$$

$f(x)$ is a decision function which classifies x in its respective category. Its value can be +1 or -1. W is the weight vector, normal to the hyperplane. b is the displacement from the origin. Figure 6 shows the separation and optimal separation planes. The points lying on the margins are called support vectors. The non-linear separations are mapped to a new space using the Kernel function: $x \rightarrow \phi(x)$, usually into higher dimensions than the original, as $f(x) = \phi(w^t.x) + b$.

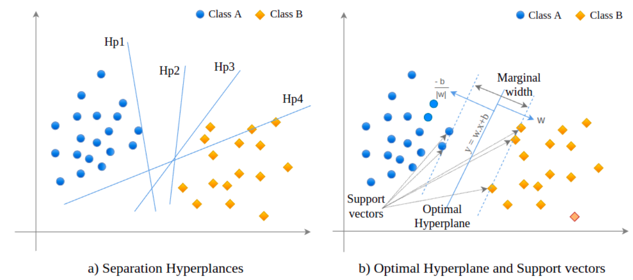

Fig. 6. Support Vector Machines(SVM) a) Separation Hyperplanes b) Optimal Hyperplanes and corresponding support vectors

Some examples of kernel functions are Linear, RBF, Sigmoid, Hyperbolic Tangent, polynomial function, etc. The mapping is such that the data becomes linearly separable in the new dimension. An optimal hyperplane should be such that it maximizes the margin to the nearest points of the classes [169]. Lagrange multipliers solve the above equation to solve the quadratic optimization problem. It determines a hyperplane perpendicular to the weight vector w with an offset b from the origin. The applications of SVM with DAR include binarization [81], single character/symbol/digit classification [97], character recognition [169], [183], word verification (signature verification) [184], text/non-text classification. SVMs have been widely used as the best method of working efficiently on small-sized datasets.

c) Neural Network Architectures The development of artificial neuron-based architectures: Perceptron and MLP, was the first step towards NN architectures. With the advancement of document storage and processing techniques, there was an increase in the use of ANNs for a critical DAR problem: Feature extraction. In early times, ANNs were also used for pre-processing tasks like removing noise (using regression to filter erroneous patterns), and binarization [185]. Some other applications of ANNs in DAR are layout analysis, segmentation and classification [98], [118], [185]. Compared to other techniques like the connected-components approach, the ANNs were more robust to the noisy patterns in the data. A backpropagation algorithm is used to train these networks. However, this error training algorithm has some drawbacks, like overfitting, local optimum, exploding and vanishing gradients, etc. ANNs have been combined with HMM-based methods for optimizing the process for character and word recognition tasks.

CNNs evolved with LeNet-5. It was inspired by Neocognitron, which could recognize visual patterns in the data. A CNN structure matches the visual perception by convolution layers by extracting features with convolution and pooling structures. CNNs are superior to ANNs regarding computational abilities such as: 1) Weight sharing and Local connections- the weights are shared over some local connections. It helps to reduce the parameters. 2) Downsampling: the feature pruning is performed by pooling operation to reduce the dimension of the feature map and thus further reduce the training parameters.

The convolution operation is concerned with local connections only, in contrast to a fully-connected layer, which concerns global connections. In a convolution operation, a kernel is scrolled over the image pixels allowing weight sharing. Consequently, the input is downsampled with the pooling operation and thus resulting in reduced computations. It also helps to overcome overfitting and redundancy. The convolution and pooling layer output is a feature map fed to the fully connected network for classification. The receptive field of a CNN defines the size of the area it perceives at a time. It is proportional to the size of the kernel. To increase the receptive fields with smaller kernel sizes, *Dilated kernels* were introduced. Many variations of convolution operations have been proposed in the literature, such as Deformable convolution, Group convolution, Steerable convolution, Graph convolution, etc. Also, various CNN models have been proposed since LeNet, such as AlexNet, VGGNets, ResNets, MobileNet, InceptionNet, GoogleNet, etc. The OD models like FasrRCNN, Yolo, SSD, etc., which are used for end-to-end text recognition, are also based on the CNN model. Figure 7 present a generic sample of CNN architecture.

CNNs have been used successfully for classification tasks like character and digit recognition tasks [158], [181]. CNNs are invariant in recognizing patterns with changes in scale and distortions. However, some setbacks in a conventional CNN are: the requirement of fixed input and dependency on the receptive field, the dense layer enables pixel-level prediction but is computationally expensive, constraints on output size due to fixed size dense layers and no reusing or sharing of feature maps makes them inefficient for semantic segmentation. Some solutions like Adaptive pooling have been proposed for removing dependency on receptive fields. The convolution network without any fully connected dense layer was proposed as FCN. These networks can produce variable-sized output while preserving the shape of the input. It transforms the input image to a semantically segmented high-resolution image by successive convolution layers stacked together for down-sampling and up-sampling. An example of FCN is a U-Net architecture. Concerning the specific tasks of the DAR task, FCNs are mostly used for semantic segmentation, such as document image reconstruction (binarization [74]) and layout analysis tasks.

d) Classifier Ensemble methods: The high-dimension feature vectors come with an issue of overfitting and the curse of dimensionality. Ensemble methods [186] like AdaBoost, and Random Forest can combine results from various classifiers. These feature selection methods can save from these problems by determining the selective features using weight factoring for the classifier models [82]. The weights of the weak learners are stressed for better classification performance. The ensemble methods are used for DAR tasks like binarization [66] handwriting recognition (HR) [186]

2) Modelling Sequence-to-Sequence Dependencies

Modelling dependencies in a sequence involves extracting the features from sequential observations and relating those features to produce an output sequence. The early era of DAR was mostly using HMMs for modelling sequence-to-sequence dependencies.

a) HMM : Among the very first techniques for character recognition was HMM, which has successfully modelled the structural and statistical characteristics of text even today. There are three categories which build up a recognition model with HMM. 1) Evaluation (Forward-Algorithm) 2) Decoding (Viterbi Algorithm) 3) Training (Baum-Welch-method, EM algorithm). The probabilistic model is based on the Bayesian formula for computing the conditional probability X where $X = x_0, x_1, ...x_t$ for given states C where $C = C_1, C_2...C_n$. A sequence X represents the text line over time t, and the character sequence in line with n elements is represented by C. Then, according to Bayes's rule:

$$\hat{C} = \arg\max_{C} p(C|X) = \arg\max_{C} p(X|C)P(C) \tag{2}$$

HMM for line recognition models the line as a left-right signal, and consequently, the features are sampled from the horizontal and vertical patches. A time series model with a Gaussian mixture model is each state's probability distribution for character recognition. The linearly connected HMM character model is used for word recognition. The success of HMM can be attributed to multiple factors a)The power of optimization, b) the potential of performing segmentation and recognition and c) representing the knowledge sources through its probabilistic model and reduced complexity for the decoding phase. Figure 7 a) shows the character and word HMM model and various combinations of HMM with NNs. Some hybrid models of HMM like ANN/HMM [187], CNN/HMM [181], HMM/BLSTM [188] model is also proposed for improved recognition performance. HMM, models can be used for end-to-end recognition [189] and online recognition [190]. Other tasks of DAR where the use of HMM has been proposed are Character classification, character recognition [183], word recognition [151], [159], [181], [191], [192], line recognition [188], numeral recognition [193] and text recognition [179].

There are some shortcomings of HMM. They are unable to model long-term dependencies. Also, as HMMs are generative, they cannot handle discriminative tasks like sequence labelling [8]. Recurrent Neural Networks (RNN) are alternate options to solve these shortcomings.

b) Recurrent Neural Networks (RNN): The RNNs are networks designed to model the sequential dependencies by taking features as input and converting them into an output sequence. The figure 7c) shows that the output $y^{<t>}$ depends on the current input $x^{<t>}$ and the previous output $a^{<t-1>}$. The distinguishing features of RNNs are 1) Memory elements: which allow RNNs to have long-term dependencies. 2) Share weights among layers. The RNNs produce output for every input feature; this requires pre-segmentation of inputs. To avoid this pre-segmentation, CTC converts the input into target labels directly. There have been many variations to traditional one-directional, one-dimensional RNNs, like Bi-directional RNN (BRNN) and Multi-dimensional RNN (MDRNN). Where a BRNN can model past and future dependencies, the MDRNN can model 2D and 3D inputs like images and videos, respectively. RNNs have some drawbacks. RNN is prone to vanishing gradient and exploding gradient problems causing shorter memory issues. Thus LSTMs were proposed. It integrates a mechanism for filtering what information to remember and what to forget through four gates: Forget gate, Input gate, Output gate and Cell gate. The cell gate is an extra memory component for long memory. The LSTMs can control the flow of information but come at the cost of increased computations for increased weights for training. Gated Recurrent Units were proposed to reduce the number of trainable weights in LSTM.

From a simple 1D RNN to GRU and LSTM, the techniques were getting better at processing single/multi-direction, single/multi-dimensions, longer memory etc. However, along with all these advantages, handling variable length input/output is a setback.

c) Sequence-to-sequence paradigm: The encoder-decoder architectures were proposed to model variable length input and output sequences. These are based on recurrent architectures. The encoder extracts the features from the input sequence to handle the input of variable lengths. A context vector of fixed size represents the extracted features. Consequently, the decoder predicts the output from the context vector. However, the fixed-sized context vector cannot represent longer input sequences, hence poor predictions. The Attention mechanism was introduced to address this issue. With an attention mechanism, the relevant features can be assigned more weight through a probability distribution. This produces a context vector is a dynamic vector that is fed into the decoder. Over the years, many improvements in the mechanism have been proposed. However, the drawback is mostly due to the recurrent layers, which result in higher computational time.

d) Connectionist Temporal Classification (CTC): RNNs are trained to model temporal data as they are robust to the spatial noise of the sequences. However, the outputs of RNNs require sequence labelling for the final output. For this purpose, the RNNS are combined with a CTC module for labelling independent sequences. A CTC model uses dynamic programming, which computes the summation of negative loglikelihood of the output sequence by RNN. For a given input sequence $\mathcal{X} = x_1, x_2, ..., x_n$ the aim is to obtain an output sequence $\mathcal{Y} = y_1, y_2, ..., y_m$ such that it maximizes the probability for the correct answer. The conditional probability of x: $p(\mathcal{Y}|\mathcal{X})$ is computed to infer the most likely $\mathcal{Y}'$ (Equation 5). The per-time-step probabilities of the input sequences X are used to model the conditional probability $p(Y|X)$ (Equation 4) over alignments with the highest probabilities (most active alignments) A.

$$p(\beta_A = \mathcal{Y}|\mathcal{X}) = \prod_{t=1}^{T} p_t(a_t|\mathcal{X}) \tag{3}$$

$$p(\mathcal{Y}|\mathcal{X}) = \sum_{a \epsilon A} p(\beta_A = Y|\mathcal{X}) \tag{4}$$

$$\mathcal{Y}' = argmax_y \ p(\mathcal{Y}|\mathcal{X}) \tag{5}$$

$$\mathcal{L}(\mathcal{X}, \mathcal{Y}') = -\sum ln(p(\mathcal{Y}|\mathcal{X})) \tag{6}$$

A many-to-one mapping set β removes redundant labels with the help of alignments A. This step is also known as *decoding*. The model is trained with a maximum likelihood loss function $\mathcal{L}(\mathcal{X}, \mathcal{Y}')$ (Equation 6). The CTC networks have

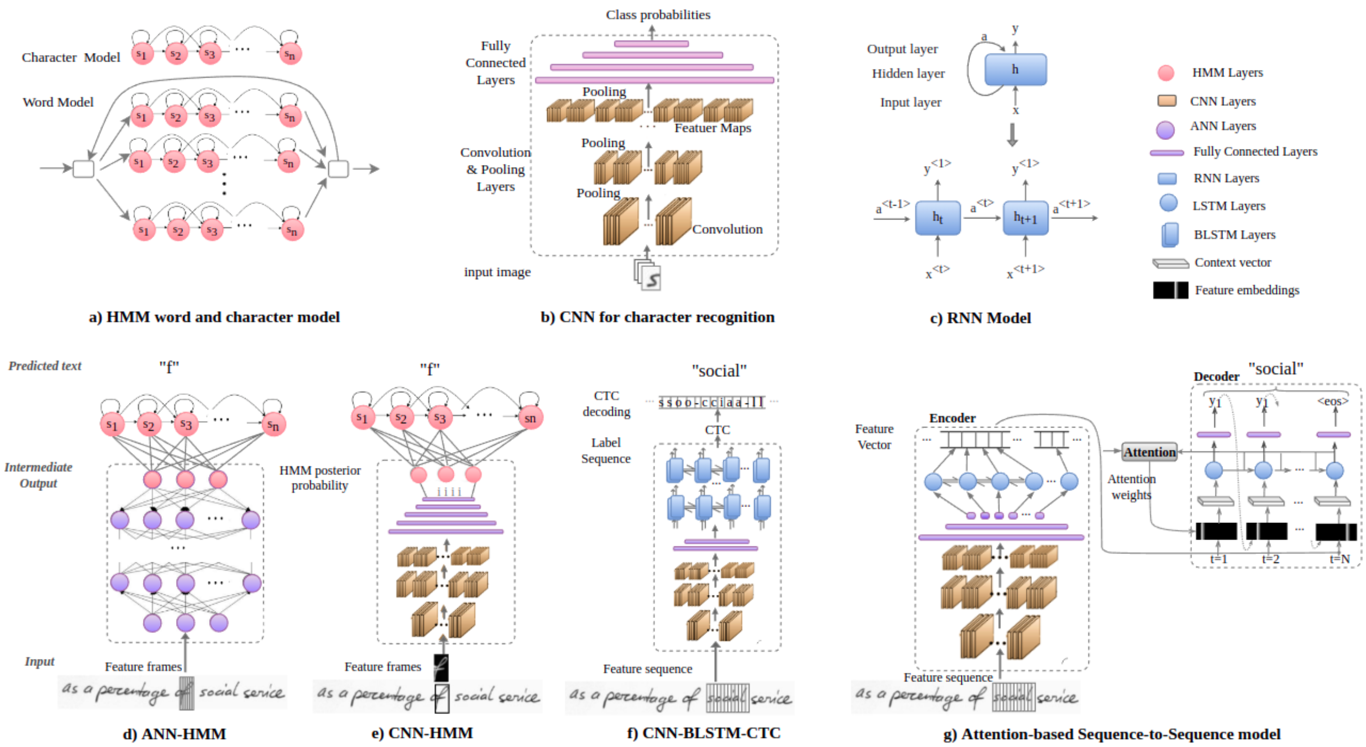

Fig. 7. Text recognition with various visual and sequential models a)HMM model b) CNN model c) RNN model d) ANN-HMM e) CNN-HMM f) CNN-BLSTM-CTC g) Attention-based Sequence-to-Sequence Model

a key advantage over other labelling models like HMM that they do not need to segment the sequences as it uses the outputs of RNNS. Although, it has some drawbacks, such as 1) the assumption of conditional independence of input sequences and 2) It cannot model inter-label dependencies. The different systems integrated with CTC can be seen in figure 7 [194]–[196].

3) End-to-end modelling

The state-of-the-art techniques for end-to-end recognition are Fast Oriented Text Spotting (FOTS), TextSpotter, Mask TextSpotter, TextDragon, etc. Other techniques of end-to-end-recognition systems are discussed below:

a) Hybrid Modelling: The early methods used mostly HMMs for end-to-end recognition, but HMMs were limited in modelling the long-term dependencies [189]. Also, some combination models like HMM-ANN [187], HMM-CNN [181], HMM-BLSTM ANN/HMM and CNN/HMM [188] were proposed to deal with the shortcomings of the former approach. The recent methods using only CNNs have improved accuracy parameters like recall, precision, etc. but are time-consuming in training and computations. These techniques trade off the accuracy parameters and time of the handcrafted segmentation and feature extraction methods. These methods have a localization step and a recognition step. There are usually two approaches to localization. The first approach uses OD techniques, and the other method predicts the start-of-line in a document image.

b) Object detection Methodologies : OD methodologies have paved the way for new techniques to pursue text detection. These methodologies aim to find the target object categories in a given input image by localizing the Region-of-Interest (ROI)s with a Region Proposal Network (RPN) [194]–[196]. ROI pooling is done to predict the object regions with a bounding box. A simple recognition method (OCR) can help recognize the text in the box. The works related to document recognition have used OD methods like Faster R-CNN and Single-shot Multi-box Detector (SSD) to predict bounding boxes for words or lines. A technique termed TextBoxes [197] has used SSD, where the aspect ratios have been increased to a longer text box to match a text area. Various adaptations have been proposed to accomplish the task of recognizing diverse text orientations in different types of documents. There are methods to predict the start-of-line references that have used CNN for feature extraction predicting references and a recurrent model for predicting the next in the line CNN-MDLSTM [198]. There is an end-of-line token to mark the end stop. Such methods are useful for recognizing curved text and multi-column text. Other such methods are pre-trained VGG with CNN-BLSTM [199], [200].

Instance segmentation is another way to detect text in a dense background document. It works on a pixel level to classify it as text or non-text. These approaches are better when it comes to dealing with disoriented texts. TextSnake [201] follows a similar approach and predicts text areas, the centre line, direction, etc.

c) Recognition without segmentation : Some works have recently explored end-to-end recognition of multi-line text or paragraphs without any segmentation step. Earlier, there have been segmentation-free approaches to character or word recognition. They used recurrent attention or one-step procedures to exploit the two-dimensional recognition attributes. The works proposed in [202], [203] have proposed a multi-line recognition with an encoder-decoder architecture. The attention modules are recurrent methods which predict character/line representations. Other recent segmentation-free methods are proposed in [204], [205] A variation of CTC, the Multi-Dimensional Connectionist Classification (MDCC) with MDLSTM architecture predicts probabilities in two dimensions [204]. Another model [206] proposes iterative paragraph processing by hybrid attention model achieves competitive results. A paragraph-unfolding method in [205] uses bi-linear interpolation with FCN encoder and CTC. These segmentation-free approaches have been quite different by accounting for a whole paragraph recognition. The CTC models are limited in modelling the input-output alignments and output CTC code length. A flexible encoder-decoder network has been used to capture the temporal features with an attention scheme. These Sequence-to-sequence (Seq2Seq) models have recently been introduced for DAR. These powerful models deal with the long-term, unaligned sequences in offline handwritten lines or word recognition [207]. The encoder and decoder are RNN-based networks that help establish the temporal relation between the feature sequences generated in intermediate steps (Figure 7).

ONLINE RECOGNITION

The online recognition of handwritten text focuses on spatio-temporal features. The work [28] categorize the recognition process into four classes of procedures used. They are statistical, structural, syntactical, and NN based. The statistical methods are probabilistic and can be parametric or non-parametric. The parametric methods (e.g. HMM) use probability distributions to model the parameters for the variables (samples of handwriting). The parameters are selected during the training procedure. The non-parametric methods (e.g. k-NN) use the input data (samples of handwriting) to estimate the unknown parameters. These methods require more computations as the training data increases. Structural components like graphs, the elasticity of strings, etc., are used to model the structural methods. The topology of shapes of characters and strokes and x-y coordinates are used for feature extraction and classification. Syntactic recognition deals with grammatical aspects of the input samples inferred from training data. It uses primitives and grammatical rules, and the sentences are represented with respect to those rules. The NNs are used for various purposes in online recognition: feature extraction, classification, learning complex input/output relations etc.

The initial online recognition works were done in Arabic, Roman, Chinese and Japanese scripts. A survey of online recognition works has been presented by [9]. It discusses various techniques of the process used widely across all scripts. The pre-processing techniques used were external segmentation, data smoothing, dot reduction, normalization, quantization, signal filtering etc. The feature extraction methods were based on extracting the structures and stroke-based information. The works have identified writing features like the vertical position of the points, writing direction, curvature, pen-up /pen-down, aspect-ratio, curliness, slope, ascender/descender, and context map. The shape-specific features like ascenders, cusps, closures etc., are identified to improve the recognition process.

The techniques used are similar to offline recognition like HMM, k-NN, SVM, NNs etc. However, there are techniques proposed specifically for the online process. The work in [208] proposed NPen++, which uses a multi-state time delay neural network (MSTDNN). It has been tested in Arabic, Roman and Tamil language. The structural techniques use features like lines, curves, diacritic points etc. Fuzzy if-then rules model these. Other techniques include Elastic matching, string matching, Statistical dynamic space warping (SDSW) classifier, Modified Quadratic Discriminant Function (MQDF) classifier, Dynamic Time Warping (DTW) based nearest-neighbour classifier, nearest prototype classifier (NPC), discriminant feature extraction (DFE) and discriminative learning quadratic discriminant function (DLQDF). Deep learning methodologies like a multi-model DNN [209] have also been explored for online recognition. The options like online/offline classifier combination and converting the online images into offline for deep networks have also shown promising results. The literature works for character or digit recognition are [210]–[214]. The HR works are [9], [190], [215]–[219] [153]. The WI works are [220], [221].

V. DATASETS

The major scripts dominating the research on the analysis and recognition of documents are Roman, Chinese-Japanese-and Korean, Arabic-like, Asian and Indian scripts [2], [28]. These scripts originated from very different types of writing systems. Understanding these writing systems can be fruitful in designing an outstanding system for DAR. The development of multi-lingual, multi-script and generic OCRs is still in its infancy due to the need for available standard datasets. The standardization of datasets can help the research community to have a better performance evaluation of the DAR systems [222]. A standard platform of datasets (maybe a possibility in future) would be helpful to the DAR community of researchers for better development and evaluation purposes.

Here we discuss the datasets on major scripts and corresponding research tasks such as pre-processing, segmentation and recognition.

A. Historical Document datasets

The historical documents are the historic remains of writings on supports like palm leaves, stones, papyri, etc. Some of the oldest surviving manuscripts of these supports are: Sanskrit Shaivism on palm leaves from the ninth-century, bamboo scrolls script from the century to Papyrus scripts from the fourth millennium BCE. Due to the ageing of these documents, their physical structure is affected by degradation caused during storage or digitization. The analysis of these documents requires some prior knowledge of the age and period of the document, as the characters, words and vocabulary vary substantially over the periods. Their diverse layout and formatting styles pose an additional challenge to analyzing these documents. Consequently, developing a single system for analyzing and recognizing historical documents is improbable. The work in [223] has classified over 60 historic documents into the Classification, Structure and Analysis dataset. The

TABLE IV
HISTORIC DATASETS FROM ALL AROUND THE WORLD: PRINTED (PR) AND HANDWRITTEN (HW) MANUSCRIPTS

Dataset	Document Type	Task	Period
Esposalles [227]	Marriage licence books (**Hw**) (Spanish)	HR	-
READ-BAD [228]	European archives(Roman) (**Pr**)	LD	1470-1930
DIVA-HisDB [229]	3 Roman Medival scripts (**Hw**)	DLA	11-14th cen
VML-HD [230]	Arabic books (**Hw**)	Rec	1088-1451
Pinkas	Hebrew scripts	PS	1500-1800
Kuzushiji-MNIST	Kuzushiji characters (**Pr**)	CR	-
GRK-Papyri [231]	Papyri scripts	WI	-
Lontar Sunda	Sudanese Palm scripts (**Hw**)	Binz, HR, LD	15th cen
Sunda AMADI LontarSet [232]	Balinese Palm scripts (**Hw**)	Binz, HR	-
Muscima++ [233]	Music HR		
IAM-HistDB (St-Gall) [234]	Roman scripts (**Hw**)	DLA, Rec	9th cen
IAM-HistDB (Parzival) [234]	German scripts (**Hw**)	DLA, Rec	13th cen
IAM-HistDB (Washington) [234]	English scripts (**Hw**)	DLA, Rec	18th cen
HJ Dataset [235]	Japanese Biography scans (**Hw, Pr**)	IR	
ARDIS [236]	Swedish Digit (**Hw**)	DR	18-19th cen
Hugin-Munin	Norwegian	HR, WI	1820- 1950
POPP [237]	Paris Census	Rec, WI	1926
HTR Benchmarks (ICFHAR-2014, ICDAR2015, ICFHAR-2016, ICDAR-2017) [238]	Bentham and Ratsprotokolle collection (**Hw**)	Rec	-
HIP2013 - HNLA2013 [239]	Historical Newspaper (**Pr**)	DLA	17-20th cen
ICDAR2015 - ANDAR-TL-1K [240]	Ancestral Documents (**Hw**)	Rec, LD	18-19th cen
ICFHR2016 - CLaMM [241]	Roman Medieval scripts (**Hw**)	Clsf	-
IC2017 - CLaMM [242]	Roman Medieval scripts (**Hw**)	Clsf	-
IC2017 - DIVA-His-DB	DLA, LD	Medieval scripts	-
ICDAR2017 - REID2017	Bengali printed books (**Pr**)	Rec, LD	1785-1909
ICDAR2017 - Historical-WI [12]	Handwritten Documents (German, French, Arabic) (**Hw**	WI	13-20th cen
ICFHR2018 - RASM2018 [243]	Scientific Manuscripts(Arabic) (**Hw**)	PS, LD, Rec	8-9th cen CE
ICFHR2018 - Asian Palm leaf [244]	Manuscripts(Balinese, Khumer, Sudanese, Roman) (**Hw**)	Binz, HR, LD	-
ICDAR2019 - DMAS2019	Digitized images Magazines (**Pr**)	HR, LD	1800 - 1938
ICDAR2019 - DIBCO2019 [245]	handwritten and machine printed images, Papyri images (**Hw, Pr**)	Binz	19th cen
ICDAR2019 - cTDaR19 [246]	Accounting records (**Hw**)	TD, Rec	
ICDAR2019 - HDRC-Chinese [247]	Chinese family records (**Hw**)	DLA, LD, Rec	-
ICDAR2019 - REID2019 [248]	Books(Bengali) (**Pr**)	LD, Rec	1713-1914
ICDAR2019 - RASM2019	Scientific manuscripts (Arabic) (**Hw**)	LD, Rec	9-19th CE
ICDAR2019 - HDRC-IR [249]	Document pages (**Hw**)	IR, WI	-
ICFHAR-2020 - HisFra-gIR20 [250]	European middle age books (**Hw**)	IR	9-15th cenCE
ICDAR-2021 - HDC [251]	Images (Roman) (**Hw, Pr**)	Clsf	-

work gives a brief introduction to all the datasets. It provides details regarding the type of documents, Usage of the datasets with their origin and performance measures and other associated metrics. The tables IV give a glimpse of some details of the major historic document datasets. The datasets have a substantial amount of Roman scripts followed by Asian scripts covering documents from worldwide. The recent literature uses deep learning methodologies comprising CNNs. These architectures address miscellaneous tasks such as Binarization, HR(character, word, line), Line detection (LD), DLA, PS, WI, IR, etc. Most datasets are used for recognition (word, line, character) tasks. Recently historic documents are also used for tasks such as the dating of documents and geographic location tracing. Although a wide variety of evaluation metrics is used in the literature, the work in [223] suggests using mIoU and mAP. The annotation formats used chiefly are PAGE XML, COCO and VOC. There is a lack of datasets besides Roman scripts. Also, large-scale datasets are required for Deep learning methodologies such as, Enormous: TableBank [23] (417K) images and PubTabNet [224] (568K+) images, DeepFigures [225] 1.4 million images, Most Categories: DocLayNet [226] (80K+) images, 11 categories.

B. Printed document datasets

A Printed document can be anything like books, scientific documents, newspapers, business letters, magazines, official

TABLE V
PRINTED DATASETS FOR VARIOUS TASKS OF DAR

Dataset	Document type	Images: Class categories	Task
Marmot [252]	Conference papers (English and Chinese)	2,000: Tables, figure	OD, TD
TableBank [23]	latex document from arXiv	4,17,000: Tables	OD, TD and Rec
DeepFigures [225]	Scientific documents	14,00,000: Tables, figure	OD, TD
UNLV	business letters, magazines, reports, newspapers, etc	10,000: 427 images with tables	OD, TD and Rec
FinTabNet	Finacial reports from business org.	90,000: 112,887 tables	TD
NTable-ori [22]	Original Camera images (Textual, electronic, wild)	2,100+: Tables	TD
NTable-cam [22]	Augmented Camera images (Textual, electronic, wild)	17,000+: Tables	TD
NTable-gen [22]	Synthetic dataset	17,000+: Tables	TD
DocBank [253]	latex document from arXiv	5,00,000: Tables, figures, equations, figures, lists, paragraphs, etc.	DLA
IIIT-AR-13k [254]	business type document	13,000: Table, figure, natural image, logo, signature	Page OD, TD
UW-III	Document images	1,600: Tables	TD
ICDAR-13 [255]	(EU & US) Gov. PDF file images	238: 150 tables	OD, TD and Rec
ICDAR-17 POD [256]	Scientific pages (English)	2,417: 2939 figures, 1069 tables, 4707 formulas	OD, TD
ICDAR-19 cTDaR [257]	Forms, financial documents, and scientific papers	3,600: Formulas, tables, figures, graphics	OD, TD and Rec
MediTables [24]	pathology, diagnostic and hospital-related reports	200: 330 tables	TD
PubLayNet [258]	over 1 million PubMed Central PDF articles	3,60,000: Table, Figure, Title, Text, List	Page OD, TD
PubTabNet [224]	scientific articles in PM-COA	5,68,000: Tables	TD and Rec
DocLayNet [226]	PDF pages	80,863: Caption, Footnote, Formula, List-item, Page-footer, Page-header, Picture, Section-header, Table, Text, Title	DLA
Laser-Printed Characters [259]		8,000: 11552 characters	Forgery Detection
NCERT5K-IITRPR [260]	NCERT school books	5,000+: tables, charts, figures, images, equations, circuit diagrams, logos	Non-text component analysis

documents like business letters or government PDFs, or medical documents like Diagnostic reports, pathology papers etc. These documents are heterogeneous regarding their representation, language, formats(scanned, programmatic or both) and reading layout. These documents have textual and graphical content like figures, tables, logos, equations, signatures, etc. Every component of a document, whether it's text or graphics, conveys information. Developing a single system that can help understand these documents is challenging. However, there can be a system to classify these documents into a smaller category of similar documents, which can understand the structure of a single page. With the development of Deep learning methodologies, the automatic processing of documents has achieved attention due to high accuracy in OCR and related methodologies for text recognition. However, deep learning paradigms require large datasets to accomplish the task effortlessly.

Previously the document datasets usually had a limited number of label categories. For instance, a lot of research was focused on Table detection with a lot of datasets at disposal, e.g. ICDAR-2013 Table competition dataset [255], ICDAR-2019 cTDaR [246], UNLV, Marmot table recognition [252], TableBank [23], FinTanNet, NTable [22]. Also, the datasets were limited in size. e.g. ICDAR-2013 Table competition dataset, ICDAR-POD-2017 page OD dataset, ICDAR-2019 cTDaR, Marmot table recognition etc. The contents of most datasets like PubLayNet [258] and DocBank [253] are limited to scientific articles, conference papers or Latex or Docx on arXiV. These documents are styled on a pre-defined uniform template. Thus they lack variation in the samples. Using them when the application is on scientific documents can be fruitful. Due to these imperfections of datasets, Layout analysis and document understanding approaches were also restricted to some specific domains.

TABLE VI
HANDWRITTEN DATASETS CATEGORISED ON DIFFERENT SCRIPTS

Dataset	**Language**			**Statistics**			**Mode**	**Task**
		samples	lines	words	characters	digits		
CEDAR [261]				10570	27835	21179	Off	Cursive DR [262], [263], WS [264], [265], Rec [266]–[270]
NIST [271]		3600			810000		Off	Hw DR [263], [272]–[274], CR [275], [276]
MNIST [277]						70000	Off	DR [278]–[283]
Firemaker [284]		1000					Off	WI, Rec [32]
IAM [285], [286]	Roman(English)	1539	13353	115320			Off	WS [27], [287]–[289] ,WI [32], [290]–[293] , Hw LS [294]–[296] , off-HR [182], [186], [187], [297]
IAM-OnDB [298]		1700	13049	86272			Off	On-HR [190], On-WI [299], GC [300]
IAM on-Do [301]		1000					Off	Content type detection [302], WS [303], TNC [20].
OnHW-chars [304]	Roman(English)				31,275(**U/L**)		On	CR [305], [306]
IBM-UB-1 [307]	Roman(English)	6500 On, 6000 Off			31,275(**U/L**)		On /Off	WI [220], WS, indexing, DAR [219]
GNHK		687	9,363	39,026			Off	
IRONOFF [308]	Roman(French)			50,000	32,000		On	Rec [309]–[312], On-WI [221].
RIMES [313], [314]	Roman(French)	12,723					Off	DLA [315], [316], mail Clsf [317], Rec [184], [318] and WI [291], [319]
RODRIGO [320]	Roman (Spanish)	1853					Off	Rec [321]–[323]
OHASD				3825	19,467		On	LD [324], Rec
AltecOnDB				152,680	644,530		On	Rec [325]
Al-Isra [326]			500	37,000		10,000	Off	HR, WI
IFN/ENIT [327]		2265		26,449			Off	preprocessing [328]–[330], WR [191], [192], WI [30], [331], [332]
Checks DB [333]		7000		29,498		15,000	Off	check HTAR [334]
AHDB [335], [336]		105					Off	HR [337], WI [338]
ARABASE [339]		400					On/Off	On/Off HR, SV
CENPARMI-A [340]	Arabic			11,375	21,426	13,439	Off	CR [341], DR [342], WS [25]
LMCA [343]				500	100,000	30,000	On	WR, DR
ADAB [344]				20,000+			Off	Seg [344]–[346], Rec, On-WI [347]
KHATT [348], [349]		1000					Off	pre-processing, Seg, WI
QUWI [350]		4068					Off	WI [351], writer demographic classification [352]–[354]
AHTID-MW [355]			3710				Off	Seg [356], WI [357]
IAUT/PHCN [358]	Arabic(Farsi)	1140		34,200			Off	pre-processing, WR [359]–[361]
IFN Fars [327]	Arabic(Farsi)			7271			Off	DR, WR
FHT [362]	Arabic(Farsi)	1000	8050	106,600			Off	Seg, Rec, BLD, content discrimination, WI, DLA
CENPARMI-F [363]	Arabic(Farsi)	432,357					Off	DR [364], HR [365], [366]
HaFT [367]	Arabic(Farsi)	1800					Off	Seg, Rec, WI
CENPARMI-U [368]	Arabic(Urdu)				18,000		Off	HR [369], WS [370]
UHSD [143]	Arabic(Urdu)	400					Off	Seg, Rec, WI
PE92 [371]	Korean				235,000		Off	Rec [372]
JPCD [373]	Japanese				1227		Off	Rec [215], [218]
HCL2000 [374], [375]	Chinese				3755		Off	Rec [376]
CASIA [377], [378]	Chinese				1.35M		On/Off	Rec [379], WS [26]
SCUT-COUCH [380], [381]	Chinese				3.6M		On	Rec [382]–[384]
CVL [385]	Roman(Eng, Ger)	2163					Off	WI [386], DR [387]
UNIPEN [388]	Roman, chinese		4563	3298	1423u, 2145l, 1222s, 2735mix	634	On	CR [214], DR
QuWI [350]	Roman(English), Arabic	2034**E**, 2034**A**					Off	WI [351], Writer demographic classification [352]–[354]
Assamese Aksharas [389]	Assamese				147 I		On	CR [389]
Assamese DB [390]	Assamese				8235		On	CR [212]
BanglaWriting [391]	Bangla			21,234	32,787		Off	OCR [19], WI [31], WS
BN-HTRd	Bangla	788					Off	Rec, WS, WSeg, LS
Numerals DB [193]	Bangla, Devanagari	642				22556**De**, 23392**Ba**	Off	Rec [193], [392]
Devanagari DB [393]	Devanagari	642			20305	5137	Off	Rec
Devnagiri-Telugu DB [210]	Devanagari, Telugu				24200**De**, 37817**Te**		On	CR [210]
Multiscript Indian DB [217]	Bangla, Devanagari, Tamil, Telugu				25948**Ba**, 23891**De**, 77609**Ta**, 45217**Te**		On	CR [213]
Gujarati DB [211]	Gujarati	4500			3700	1000	On	CR [211]
KHTD [394]	Kannada	204	4000	26,000			Off	Seg, Rec
Tamil DB [395]	Tamil			265,00			Off	Rec
uTHCD [396]	Tamil	91000		265,00			Off	Preprocessing [397], Rec [398]
HPL-iso-Tamil-char [399]	Tamil			78000			Off	CR
HPL-iso-Tamil-word [400]	Tamil			8500			Off	WR
Tamil-Kannada word corpus [401]	Tamil, Kannada			100000**Ka**, 100000**Ta**			Off	WR
PWDB 13 [402]	13 Indian Scripts			26000			Off	WR [403]
Multiscript DB 11 scripts [404]	11 Indian Scripts			28100			Off	SI [404]
Multiscript DB 5 scripts [216]	Arabic, Cyrillic, Devanagari, Han, Hebrew, or Roman	15 **Ar**, 10 **Cr**, 18 **De**, 15 **Ha**, 10 **He**, 45 **Ro**	209 **Ar**, 276 **Cr**, 382 **De**, 282 **Ha**, 284 **He**, 722 **Ro**	1423**Ar**, 1002**Cr**, 3173**De**, 1981**Ha**, 2261**He**, 3539**Ro**			On	SR [216]
HKR [405]	Cyrillic	1400	63000		715699		Off	Rec [406]
Cyrillic-MNIST [407]		121,234					Off	Rec
KenTrans	Swahili		12,400					Rec
AMHCD	Amazigh				25,740			CR
IRCAM-Tifinagh	Berber	102			3,366			CR
Indonesian-TDB [408]	Indonesian	200					Off	Rec
MRG-OHTC	Tibetan				910		Off	Rec
Handwritten Circuit Diagram Images [409]		1152 of 144 circuits					Off	OD

U: Upper case; L: Lower case; s: small; De: Devanagari; Ba: Bangla; Te: Telugu; Ta: Tamil; Ka: Kannada; Ar: Arabic; Cr: Cyrillic; Ha: Han; He: Hebre; Ro: Roman

Recently, research has shifted toward unique and specific areas of document analysis than the traditional ones. More datasets are now developed with various labels and comprehensive data acquisition sources. Also, there has been an effort to standardize ground truth data format. Common data formats ease the training procedures of OD methods on datasets. Some of the large sized datasets are like TableBank [23], DocBank [253], DeepFigures [225], PubTabNet [224], PubLAyNet [258], etc. Some recent ones are NCERT5k-IITRPR [260] for text/non-text component analysis, Laser-Printed Characters Dataset [259] for document forensics application, DocLayNet [226] for general purpose DLA. Table V enlists the available datasets supporting document classification, analysis and understanding and summarizes some major contributions to printed datasets.

C. Handwritten datasets:

Handwriting has been a mode of communication and information storage means for a long time. With the advancement of digital technologies for reading and writing, printed documents have become more convenient and easy to store. However, it can never replace the convenience of a pen and paper. The need for Handwritten text analysis and recognition (HTAR) systems is inevitable. The standard handwritten datasets were developed as early as the 1990s [222]. The initial phases of DAR were usually concerned with character, word and digit recognition. The evolution of HR techniques paved the way for unconstrained recognition methodologies. Another aspect concerning DAR is the development of techniques to reproduce the human efficiency of recognition. This is possible only when the datasets also replicate real-world scenarios. The unconstrained and flexible data collection environment is essential for creating such datasets.

The handwritten datasets can be Online or Offline. The writings of individuals on a digital device collect the online datasets. These datasets contain writing information, like the x-y coordinates of the pen position, pen pressure, writing speed, stroke order, etc. On the contrary, offline datasets have labelled or unlabelled samples of alphanumeric characters, words, lines or pages categorized into different classes. Written samples and their labels and classes are stored digitally in both cases. The digitization of these samples undergoes several processes like size-normalization, smoothing, and feature extraction before recognition [14]. Labelling the ground truth samples determines what all tasks can be carried out on the dataset. A segmentation system requires coordinates of corresponding segments in the document. A recognition system requires a database with labelling regarding text images' transcriptions (word, character or digit). The WI and recognition systems use Writer Identity-related demographics. Most research is on offline handwritten works; therefore, most datasets have information about the recognition task. There are lesser datasets for tasks such as Word spotting and writer demographic classification as very few datasets support these tasks.

Major handwritten datasets include *IAM* [286], *NIST* [271], *MNIST* [277], *CEDAR* [261], *RIMES* [313], [314], *UNIPEN* , *CENPARMI-Arabic* [340] *PE92* [410], etc. The datasets developed are mostly in languages like *English* like IAM, CEDAR, NIST, MNIST, IAM-OnDB, etc., *Arabic* AHDB, ARABASE, CENPARMI-A, LMCA, KHATT, CENPARMI-F etc., *Chinese* HCL2000, CASIA, SCUT-COUCH, etc., *Indian languages* like Bangla: BN-HTRd, Numerals DB, Devanagari DB, Multiscript Indian DB (Bangla, Devanagari, Tamil, Telugu) [217], Multiscript DB 11 scripts (Roman, Devanagari, Urdu, Kannada, Oriya, Gujarati, Bangla, Gurumukhi, Tamil, Telugu, Malayalam) [404]

The traditional tasks for DAR, supported by most datasets, are pre-processing, segmentation and recognition. Other tasks like DLA, word spotting, and forensic document analysis (WI and verification) have very few datasets concerning them. Tables VI represents various handwritten datasets online/offline with respective languages and supported tasks. The *IAM* dataset [286] is the most widely used dataset for recognition tasks covering samples of around 350 writers. However, this dataset has a problem of imbalanced data distribution in terms of samples per Writer. Due to this reason, IAM is not a probable choice for writer recognition and associated tasks. RIMES [313], [314] dataset is the largest collection of unconstrained samples of over 130 individuals, although it is not free. The Arabic handwriting and writer recognition mostly use *IFN/ENIT* dataset.

Viewing the overall works on DAR, the focus is mainly on English and Arabic scripts. One of the reasons is the availability of free and labelled data samples in this domain. Recent works on DAR have seen a wider spectrum of scripts included in a single research. Creating solutions for multi-script analysis and recognition involves uncovering the commonalities of various writing systems and individuals. Some multi-script databases are like: *UNIPEN* (Roman, Chinese) [388], *QuWI* (Roman, Arabic) [350], *CVL* (English, German) [385], English-French [307], Multiscript Indian databases: Bangla-Devanagari-Tamil-Telugu [217], Arabic-Cyrillic-Devanagari-Han-Hebrew-Roman [216], $PWDB\_13$ 13 Indian scripts database [402], Multi-script DB of 11 indian scripts [404]. This data store information regarding the Writer's identity, such as age, gender, background, etc. The tasks supported by these datasets are writer demographic classification, WI and recognition and multi-script recognition.

VI. Evaluation Metrics

Evaluation metrics analyze the strengths and weaknesses of the techniques and processes. They help in quality control and performance check of the systems. The basic evaluation scheme compares the predicted output with the corresponding ground truth labels of the data. The evolution of DAR systems and algorithms has made it difficult to evaluate them on common ground. The evaluation schemes have been classified into theoretical and experimental approaches [8]. The theoretical approaches are used for low-level algorithms, mostly visual pattern recognition tasks. The experimental methods are further classified into with ground truth data and without ground truth data. When comparisons are made without ground truth, the measures are made to evaluate specific features of the algorithms to test the quality. Considering the complexity of the evaluation procedures concerning the DAR tasks, the segmentation algorithms are harder to evaluate than the classification algorithms. We have categorized the evaluation procedures as per the task of DAR. Table VII lists the metrics to measure the performance of different DAR tasks. These metrics are defined in competitions that seek solutions to trivial and specific problems of DAR. The metrics presented here can be referred from competitions like Handwritten Document Image Binarization Contest [411]–[413], ICDAR 2005 PS Competition [414], etc.

A. Metrics for Pre-processing methods

Image quality metrics are used to evaluate the pre-processing methods like binarization. The binarization methods are evaluated mostly on measures provided by competitions on Binarizations. Even after the standardization of evaluation methods, there still needs to be a standard method that can be applied to most documents. The standard measures to evaluate the binarization methods as provided in various Binarization competitions [411]–[413]are F-Measure, PSNR, NRM, DRD and MPM. The Precision and Recall metrics are the most significant ones for all the tasks of DAR. Recall refers

TABLE VII
EVALUATION METRICS FOR PRE-PROCESSING, SEGMENTATION AND RECOGNITION TASKS OF DAR

Metrics	Description	Metrics	Description
Evaluation of Pre-processing Methods			
Pr: It checks if all objects are extracted	$Pr = \frac{TP}{TP+FP}$; $Pr = \frac{\sum_{x=1}^{Nx}\sum_{y=1}^{Ny} GT(x,y)-B(x,y)}{\sum_{x=1}^{Nx}\sum_{y=1}^{Ny} B(x,y)} \times 100$	Rc: It checks only correct objects which are extracted	$Re = \frac{TP}{TP+FP}$; $Re = \frac{\sum_{x=1}^{Nx}\sum_{y=1}^{Ny} GT(x,y)-B(x,y)}{\sum_{x=1}^{Nx}\sum_{y=1}^{Ny} GT(x,y)} \times 100$
FM: Harmonic mean pr Pr and Re	$FM = \frac{2\times Pr\times Rc}{Pr+Rc}$	pseudo FM: It concerns pseudo Recall and pseudo precision. These pseudo-measures compute distance weights with respect to the ground truth.	$pFM = \frac{2\times p-Pr\times Rc}{p-Pr+p-Rc}$
PSNR: It is a measure of the similarity of images.	$PSNE = 10log\left(\frac{C^2}{MSE}\right)$; $MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y)-I'(x,y))^2}{MN}$	DRD: It gives a measure of visual distortion in a binarized document image.	$\mathcal{D} = \lvert GT_k(i,j) - B_k(x,y)\rvert \times W_{Nm}(i,j)$; $DRD = \frac{\sum_{k=1}^{S}(\sum_{i=-2}^{2}\sum_{-2}^{2}\mathcal{D})}{NUBN}$
NRM: It is the arithmetic mean of foreground pixels changed to the background in relation to all foreground pixels and vice-versa.	$NRM = \frac{\frac{FN}{FN+TP}+\frac{FP}{FP+TN}}{2}$	MPM: It measures the mean distance between contour points of ground truth and binary images.	$MPM = \frac{\sum_{a=1}^{FN} d_{FN}^{a}+\sum_{b=1}^{FP} d_{FP}^{b}}{D}$ d_{FN}^{a}=distance between a^{th} FN $(GT = 1, B = 0)$ and closest contour pixel of B; d_{FP}^{b} = distance between b^{th} FP $(GT = 0, B = 1)$ and closest contour pixel of GT; D =normalization factor
Evaluation of Segmentation Methods			
IoU: The ratio of the area of the intersection between the ground truth GT and the prediction $(GT)'$	$IoU = \frac{GT \cap (GT)'}{GT \cup (GT)'}$	mAP: It is the mean of the Average Precision (AP) of the different classes to be recognized	$mAP = \frac{1}{card(C)}\sum_{c\epsilon C} AP_c$
EDM: A performance metric for detecting each entity can be extracted if we combine the entity's detection rate and recognition accuracy values. N_i = count of ground-truth elements belonging to entity i, M_i = count of result elements belonging to entity i, and $w_1, w_2, w_3, w_4, w_5, w_6$ are pre-determined weights, we can calculate the detection rate (DR) and recognition accuracy (RA) for i as:	$EDM_i = \frac{2DetectRate_j RecognAccuracy_i}{DetectRate_j+RecognAccuracy_i}$; $MatchScore(i,j) = a\frac{T(G_j\cap R_j\cap I)}{T((G_j\cap R_j)\cap I)}$, $where\ a = \begin{cases}1 & if\ g_j = r_i\\ 0 & otherwise\end{cases}$ represents matching results of j GT region, and i result in the region $DR_i = w_1\frac{o2o_i}{N_i} + w_2\frac{g-o2m_i}{N_i} + w_3\frac{g-m2o_i}{N_i}$; $RA_i = w_4\frac{o2o_i}{M_i} + w_5\frac{g-o2m_i}{M_i} + w_6\frac{g-m2o_i}{M_i}$	SM: A global performance metric for detecting all entities can be extracted if we combine all values of detection rate and recognition accuracy.	$SM = \frac{\sum_I N_i EDM_i}{\sum_I N_i}$
FM: FM for segmentation methods concerns the $MatchScore(i,j)$ between S_i and G_i. A match is counted if the $MatchScore(i,j)$ equals or exceeds threshold T_a.	$FM = \frac{2\times DR\times RA}{DR+RA}$; $DR = \frac{No}{N_g}$; $RA = \frac{No}{N_r}$; $No = \#MatchScore(i,j) \geq T_a$	U: It is used to integrate all different types of misclassification into one value, U	$U = \frac{\#(\varphi_1 N_s+\varphi_2 N_m+\varphi_3 N_e+\varphi_4 N_p)}{\#N_g}$
Evaluation of Recognition Methods			
CER: Its the percentage of wrong recognized characters and total characters	$CER = \frac{wrong\ recognized\ characters}{total\ characters}$	CRR: Its the percentage of recognition rate on character level and total characters	$CRR = \frac{correct\ recognized\ characters}{total\ characters}$
WER: Its the percentage of wrong recognized words and total words	$WER = \frac{wrong\ recognized\ words}{total\ words}$	WRR: Its the percentage of correctly recognized words and total words	$WRR = \frac{correct\ recognized\ words}{total\ words}$
Accuracy: the ratio of correctly recognized characters to characters in the ground-truth data	$Accuracy = \frac{correctly\ recognized\ characters}{characters\ in\ the\ ground-truth\ data}$		

Image size = $N_x \times N_y$, B = Binary image; I(x,y) = original image, $(I)'$(x,y) = noisy image, W = weight matrix, NUBN = non-uniform (not all black or white pixels); o2o: one-to-one; o2m: one-to-many; m2o; many-to-one m2m: many-to-many; S_i = i_{th} text-line detected with segmentation, Nr = extracted lines, N_g = text-lines in input image; N_s = split regions, N_m = merged regions, N_e = missed regions, N_p = partially missed regions

to completeness, and precision refers to the purity of the information retrieval task. F-Measure is used to check the quality of the whole system, which is the harmonic mean of Pr and Re. PSNR and NMR can measure the test of similarity of ground truth and binary image. NMR can provide a specific measure of the relationship between foreground and background pixels. The DRD metric is useful for computing the visual distortion in a binary image. MPM considers the contour points to account for the misclassified points between the ground truth and binary image. Among all the techniques of binarization, Adaptive contrast outperforms others. Different binarization methods reported in this work have measured the performance on metrics such as Precision, Recall, FM, and PSNR. These works are Global threshold [53], Local threshold [48], [49], Edge Detection [55], Image transforms [58], Mixture models [60], Conditional Random Fields [62], Game theory [65] and Deep Learning [70].

B. Metrics for Segmentation Methods

The Segmentation methods include layout analysis methods and page/line/word/character segmentation methods. The metrics discussed in this section are referred from PS competition [414], which is mostly based on a $MatchScore(i, j)$. The number of matches detected versus ground truth is defined as the count of matching results of GT region j and result region i. The Detection Rate and Recognition accuracy is defined based on MatchScore. Another metric, EDM, computes detection performance. The performance of all detection rate and recognition accuracy values gives a global measure of detection as SM. Sometimes, the result of segmentation methods is RoI areas bounded by some bounding boxes. Two metrics measure the quality of these segmented regions: IoU and mAP. The segmentation metrics are the most complex to evaluate due to the diversified nature of the problem. Various works for segmentation methods have considered diverse metrics of performance. However, IoU and mAP are the basic measures to evaluate the efficiency of methods. Projection-based methods like projection profiles have used skew correction error [108] for LS. Smearing-based method (RLSA) in [100], [415] have used FM, whereas the Background analysis method Voronoi diagram [107] uses success rate to represent segmentation for word segmentation.

C. Metrics for Recognition Methods

The recognition metrics measure the error rates, whether characters or words. The metrics are based on Levestein distance which uses dynamic programming. The Character error rate (CER) computes the Levestein distance between the predicted and ground truth character. The Word error rate (WER) computes Levestein's distance between predicted and ground truth words. The work in [8] defines Accuracy (Recall) and Precision for recognition systems as the ratio of correctly recognized words to the total number of characters in ground truth or OCR output, respectively. The text recognition methods in the literature have been evaluated for CER, WER

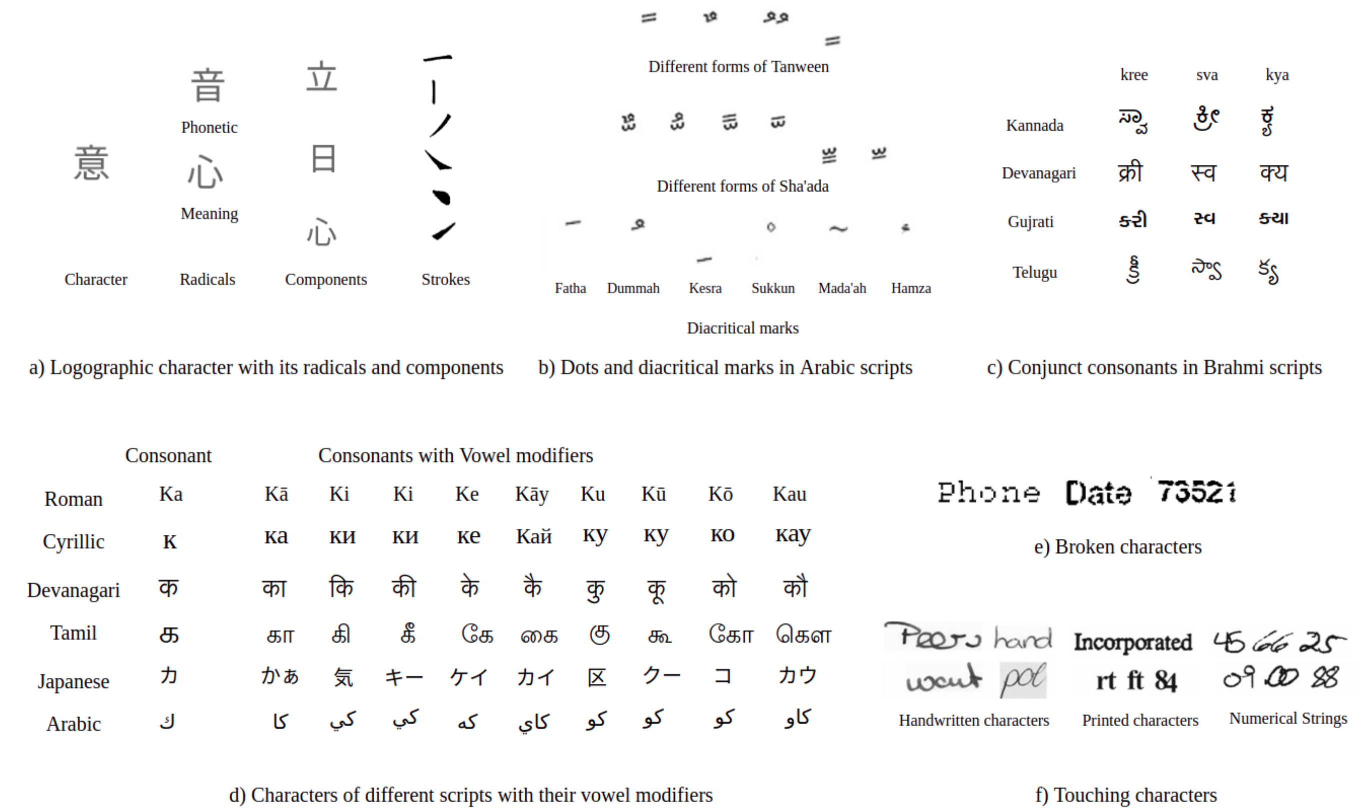

Fig. 8. Challenges of different writing systems a) Logographic characters b) Arabic scripts [421] c) Brahmi scripts d) Multiple scripts characters and vowels e) Broken Characters f) Touching characters [422]

and Accuracy. The techniques based on CNN have usually used Accuracy as a performance metric. Such techniques include CNN (Alexnet) [416] CNN [417], CNN and BLSTM [418], CNN-FCL-SVM [419], Deep Belief Neural Network (DBNN) [420] etc. The networks with recurrent architectures like LSTM and BLSTM have used CER, and WER as their performance measure. These methods include FPN-CRNN-CTC [194] Attention-based Encoder Decoder [206] RPN-CNN-BLSTM-LM [199], CNN-MDLSTM [202], CNN-MDLSTM [203] etc.

VII. Discussion

NNs have been the method of choice for the problems of feature extraction and classification. They are used for a wide variety of problems of DAR, like online or offline, printed or handwritten recognition of single words, characters and digits. Although, for a complete sequence of text like lines, paragraphs, etc., the NNs are combined by models which can process probabilities into character sequences. Such models include CTC [406], HMM, etc. A NN takes an image as an input and produces corresponding feature vectors. A task of DAR may require processing contextual or temporal information based on the application. Simple tasks like character or digit recognition can be accomplished by considering only contextual information. In these cases, ANNs or CNNs are efficient choices. However, the processing sequence of text requires temporal information and additional storage. RNNs, LSTM and Bi-LSTM, etc., are used. The Bi-LSTMs can process past as well as future sequences as they process from both right-to-left and left-to-right directions. As the BLSTMs are usually not integrated with a language model, combining them with HMM or CTC model improved recognition performance. Among all the recognition models, the best-performing recurrent model is BLSTM networks. Although, the downsides of these approaches are the large data and training requirements. An alternative system used in the literature to overcome the drawbacks of the RNNs is a hybrid solution: ANN/CNN with HMM model as shown in figure 7.

Other NN architectures proposed are encoder-decoder with attention, Gated FCN, Gated CNN, The DAR tasks with the use of NNs in the literature are character recognition [169], [184] [213] survey of NN approaches [406], word recognition, line recognition [188], attention based recognition, WI [31], OD models for recognition [115], etc. RNN overtook the road of recognition problems with the development of deep networks. The performance of classification and recognition is based on the proficiency of feature extraction methods. SVM has shown better character recognition results than other tasks of DAR. However, SVM, HMM, and NN have been the most widely used techniques for DAR over all the years of research [423]. The performance of various techniques for DAR tasks like Binarization, DLA, Offline and online recognition is presented in tables VIII, IX, X, XI respectively.

A. *Script-specific challenges*

1) Script-specific Pre-processing: Most of the pre-processing methods above are independent of scripts. However, some documents require processing script-specific characteristics with methods aware of the script idiosyncrasies.

For instance, The Arabic scripts have dots and diacritical marks, and Urdu and Farsi have retroflex marks corresponding to some characteristics. The pre-processing in these scripts removes such markings and adds them in the post-processing phase. Indic scripts have a 'shirorekha' above all characters. It helps identify the document's skew angle and words as a shirorekha connects a word. However, handwritten Indic documents are challenging due to distance variation in lines, as vowel modifiers and other writer-specific traits are scattered throughout. Analysis and recognition of historical documents on ancient elements like palm leaves, stones etc., have also been studied with early and modern techniques. Mostly entropy-based methods have been used for historical documents, but these fail in the case of palm leaves scripts as they have highly degraded.

2) Script-specific segmentation: The word and character representations are diverse across all the world's major writing systems. In Roman scripts, the words usually have more inter-word than inter-word spaces. In contrast, the logographic systems (CJK) have a syllabic system where words do not indicate spaces between words. With these scripts, there is only LS and character segmentation. The LS in CJK scripts is easier and similar to that in Roman scripts. Character segmentation is easier in printed documents with only CJK scripts, as the syllables can easily be segmented with methods like projection profiles. However, the task is complicated for handwritten documents. Radical-based [156], stroke-based [157], and holistic approaches [155] have been proposed for CJK character recognition. There has been significant research for CJK scripts, although the early research is done on smaller specific datasets, making it difficult to compare the performances. With more standard datasets, state-of-the-art assessment has become easier. The research in recognition of printed documents in CJK script has achieved high accuracy levels (THOCR), whereas handwritten DAR is still unresolved.

The Abjad systems include Arabic, Hebrew, Syriac, and Thaana scripts. These are the most complicated ones, with only alphabets and a very large (effectively 4,000 used today) similar character set. Some complications are non-heterogeneous spacing, cursive writing, overlapping words, right-to-left writing style and diacritical marks around characters. Also, the scripts lack a symmetry of ligatures in height or width, making the segmentation more challenging. Hence simple methods such as projection profiles cannot be applied here. Techniques like the detection of peaks and skeleton analysis have helped detect the global baseline; however, these scripts are very complex and therefore extracting local baselines may be difficult. For word segmentation, the vertical profiles could be more fruitful; therefore, specific methods have been proposed for these scripts. Regularities and singularities based, local minima of the upper contour, and finding right-to-left order [150], [151] are some proposed methods. Feature extraction, followed by NNs for classification, is widely used in character segmentation and recognition.

Abugidas' writing system is based on Brahmi script, including Indic scripts (Devnagiri, Gurumukhi, Gujarati, Bengali, Manipuri, Oriya, Tamil, Telugu, Kannada, Malayalam, etc.), Indonesian scripts (Balinese, Buginese), etc. All these scripts have significant discrepancies, which are caused due to regional influence over the years. The Abugidas script system has a small set of consonants and vowels forming a character. The combination of these consonants and vowels varies across all scripts and is a large amount, thus causing difficulties in segmentation and recognition. 8. The research for Indic scripts has been mostly focused on Devanagari and Bangla. The main reason is the less availability of open-access datasets. Recently more standard datasets have been made available (Indic scripts dataset), which has opened gates for research in this field. Many options for segmentation techniques for scripts with a shirorekha like Devanagari have been proposed. Word segmentation methods of Roman scripts can be applied to the Indian scripts. However, character segmentation is quite challenging as added changes in word representations are due to the vowel modifiers and conjunct consonants. 8. Conjunct consonants are used mostly in Brahmi scripts for sophisticated sounds that combine two consonants [424]. Structural character segmentation techniques are not useful for such cases. Graph-based methods have been proposed to handle these cases [425].

3) Scripts-specific recognition: The high-level features of scripts require specific techniques to determine the prime characteristics of their components. However, the low-level features are mostly independent and can be extracted through common feature extraction methods (for all scripts). The logographic scripts have radicals and components as the basic elements of their characters 8. A basic feature extraction method is to skeletonize the radicals and compute principal components by PCA. A more fine characteristic feature of these scripts is Strokes. In Chinese, only six strokes form different components of a character. The geometry and position of these strokes are also used as a feature. Statistical approaches for CJK scripts include Peripheral, Stroke-density,

TABLE VIII
PERFORMANCE OF VARIOUS BINARIZATION METHODS REPRESENTED BY EVALUATION METRICS F-MEASURE (FM) AND PEAK SIGNAL TO NOISE RATIO (PSNR)

Category	Method	Dataset	FM	PSNR
Global threshold	Otsu [53]	DIBCO-11	82.22	16.94
		Santgall	80.71	17.09
Local threshold	Niblack [48]	DIBCO-11	68.52	12.76
	Sauvola [49]	DIBCO-11	82.54	15.78
		Santgall	88.68	19.86
Edge Detection	Su [55]	DIBCO-11	87.8	17.56
Image transforms	Sehad 7 [58]	DIBCO-11	88.90	17.51
Mixture models	FAIR [60]	DIBCO-11	92.36	19.32
Conditional Random Fields	Howe [62]	DIBCO-11	88.74	17.84
Game theory	GiB [65]	DIBCO-11	89.85	18.86
Deep Learning	Pastor [70]	DIBCO-13	87.74	18.91
		Santgall	97.02	27.22

TABLE IX
PERFORMANCE OF DLA METHODS ON VARIOUS EVALUATION METRICS

Category	Method	Dataset	Metric	Result
Projection-based methods	Projection Profile [108]	Multi-script	Skew correction Error	0.12
	Recursive XY cut [101]	private	Subjective	NA
Smearing-based methods	RLSA [100], [415]	private	FM	84.80%
CC-based methods	Docstrum [102]	private	Subjective	NA
	Delaunay triangulation	Mullti-script	FM	100%
Background analysis methods	Voronoi diagram [107]	Multi-script	success rate	99.05%

TABLE X
PERFORMANCE OF OFFLINE RECOGNITION METHODS ON EVALUATION METRICS LIKE ACCURACY, CER, WER

Script - Task	Dataset	Method	Metric	Result	Script - Task	Dataset	Method	Metric	Result
Arabic - HCR	OIHACDB-40	CNN (Alexnet) [416]	Acc	100%	Arabic -HCR	AHCD	CNN (Alexnet) [416]	Acc	99.98%
Arabic - HCR	AHCD, Hijja	CNN [417]	Acc	97.00%,	Arabic - HCR	KHATT	CNN and BLSTM [418]	Acc	Very high
Arabic - HCR	AHCD	CNN+FCL+SVM [419]	Acc	95.07%	Arabic - HCR	HACDB	Deep SVM [420]	ECR	2.1%
Arabic - HCR		DBNN [420]	ER	40%	Arabic, Farsi, Persian - HWR	IAUT/PHCN	DTW [361]	Acc	94.00%
Urdu - HWR	CENPARMI	SVM [369]	Acc	97.00%	Arabic - HWR	IFN/ENIT	MDLSTM [426]	Acc	98.92%
Arabic - HWR	SUST-ARG names	CNN [427]	Acc	99.14%	Chinese - HCR	HWDB1.0-1.2	DFE+DLQDF [428]	Acc	92.72%
Chinese - HCR	HWDB1.0-1.2	CCPR-2010 Winner: HKU [429]	Acc	89.99 %	Chinese - HCR	HWDB1.0-1.2	ICDAR-2011 Winner: [430]	Acc	92.18 %
Chinese - HCR	HWDB1.0-1.2	ICDAR-2013 Winner: [431]	Acc	94.77%	Chinese - HCR	HWDB1.0-1.2	DirectMap+ConvNet [432]	Acc	96.95 %
Chinese - HCR	HWDB1.0-1.2	DirectMap+ConvNet+ Adaptation [432]	Acc	97.37 %	Chinese - HCR	HWDB1.0-1.2	DirectMap+ConvNet+ Ensemble [432]	Acc	97.12 %
Chinese - HCR	HWDB1.0-1.2	HCCR-Gabor-GoogLeNet [433]	Acc	96.35 %	Chinese - HCR	HWDB1.0-1.2	HCCR-Ensemble-GoogLeNet-10 [433]	Acc	96.74 %
Roman - HCR	EMNIST	DWT-DCT+SVM [434]	Acc	89.51%	Roman - HDR				97.74%
Roman - HCR	NIST SD 19	HMM [435]	Acc	90.00%	Roman - HDR				98.00%
Roman - HCR	EMNIST	MRF-CNN [436]	Acc	95.44%	Roman - HDR				99.75%
Roman -HWR	ICDAR-RRC ICDAR 2003	RNN-BLSTM-CTC [437]	Acc	89%	Roman -HWR	ICDAR 2011			87%
Roman -HWR	ICDAR 2013			90%	Roman -HWR	GSVT			89%
Roman - HWR	IRONOFF, SRTP-Cheque, AWS words	NN-HMM [438]	Acc	96.1%	Roman, Oriya - HWR	2500 Words	(MLP-NN) [439]	Acc	99.6%
Devanagari, Bengali - HWR	Bengali	HMM [440]	Acc	90.23%		Devanagari			93.82%
Kannada, Devanagari, Roman - HWR	1850 words	*k*-NN [441]	Acc	98.61%	Devanagari, Roman - HWR	474 words	Heuristic approach [442]	Acc	85.95%
Roman - EER (pg)	IAM	FPN+ CRNN + CTC [194]	C	15.6%	Roman - EER (pr)	RIMES	Attention-based Encoder-Decoder [206]	C, W	1.91%, 6.62%
Roman - EER (pr)	IAM		C, W	4.45%, 14.55%	Roman - EER (pr)	READ 2016		C, W	3.59%, 13.94%
Roman - EER (pr)	RIMES	RPN+CNN+BLSTM+LM [199]	C, W	2.1%, 9.3%	Roman - EER (pr)	IAM		C, W	6.4%, 23.2%
Roman - EER (pr)	RIMES	CNN+MDLSTM [202]	C, W	2.9%, 12.6%					
	IAM		C, W	6.4%, 23.2%	Roman - EER (pr)	IAM	CNN+MDLSTM [203]	C	16.2%

C: CER; W: WER; Acc: Accuracy; ER: Error rate; pg: page; pr: paragraph

TABLE XI
PERFORMANCE OF ONLINE RECOGNITION METHODS ON EVALUATION METRICS: ACCURACY

Script	Method	Database	Acc
Arabic	Data from a single user	A hierarchical classifier with tree structure [443]	98%
Arabic	120 postcode words based on 13 Arabic characters	Structural analysis method: directional features + decision tree [444]	86% WI, 100% WD
Arabic	Videos of 28 isolated Arabic letters. Two different users wrote each letter eight times	The motion information of the hand movement is projected onto two static AD images + video-based approach + KNN classifier [445]	99.11%
Chinese	OLHWDB1.0-1.2	DFE + DLQDF [428]	95.31%
Chinese	OLHWDB1.0-1.2	CCPR-2010 Winner: SCUT-HCII-2 [429]	92.39%
Chinese	OLHWDB1.0-1.2	ICDAR-2011 Winner: VO-3 [430]	95.77%
Chinese	OLHWDB1.0-1.2	ICDAR-2013 Winner: UWarwick [431]	97.39%
Chinese	OLHWDB1.0-1.2	DirectMap + ConvNet [432]	97.55%
Chinese	HOLHWDB1.0-1.2	DirectMap + ConvNet + Adaptation [432]	97.91%
Chinese	OLHWDB1.0-1.2	DirectMap + ConvNet + Ensemble [432]	97.64%
Japanese	881 characters	HMM [446]	95.4%
Japanese	TUAT HANDS, kuchibue-d-97-06, nakayosi-t-98-09	Euclidean distance and modified quadratic discriminant function [447] (MQFD2)	91.0%
Arabic, Roman, Tamil	-	Information Retrieval (IR) model [448]	93.3%
Roman	5000 dictionary words	multi-state time delay neural network (MSTDNN) [208]	96.0%
Assamese	18000 samples, 100 writers	HMM and vector quantization (VQ) classifiers [449]	99.3%
Bangla	-	Distance function based on Levenshtein distance metric [450]	98.4%
Devanagari	460 words	CNN classifier [451]	98.2%
Devanagari	Cursive words	HMM classifier [452]	88.1%
Tamil	7233 word	HMM classifier [453]	98.0%

WI: writer independent; WD: writer dependent

stroke-directional, and gradient features [8], [454]. For Brahmi scripts, particular script-specific features have been utilized to capture structural and statistical aspects. A 'shirrorekha' in the Devanagari script is an example of one such discriminative feature. GSC features have been used in scripts like Roman, Arabic and Brahmi for domain-specific feature extraction. However, the fine peculiarities of Arabic and Brahmi scripts make it difficult to capture the details of vowel modifiers 8.

Roman scripts: The Roman or Latin script is the most widely used script, with over 300,000,000 users worldwide. It is a bicameral script, i.e. it has alphabets in upper case and lower case, with 26 letters. The literature for this script has seen developments since the inception of the problem of character recognition [269], [270]. These methods used HMM and SVM for the recognition task. Later deep learning architectures were used, such as the work [436] uses the EMNIST dataset to train the MRF-CNN model. Even the recent techniques for DAR such as [194]–[196], [198]–[200] are applied to Roman datasets, mostly English. The characteristic features of this script pose some challenges to DAR. 1) all the letters stick to a baseline, making extracting text line segments easy. However, the descenders below the letters like *p* and *q* may extend up to the next line, thus disrupting the process of segmentation. 2) spacing between the characters defines the words: printed documents are well spaced; however, the in some handwriting styles, the system can not differentiate between the inter-word and inter-character spacing. 3) similar letter shapes confuse classifiers: the letters and digits such as *a-d*, *d-l*, *F-E*, *U-V*, *c-e*, *S-5*, *L-1* are some confusing instances.

Chinese-like scripts: Much work has been done for Chinese character recognition. Three categories of recognition approaches have been reported: Radical-based, Stroke-based and holistic approaches. Based on the text component extracted through segmentation, the approaches are Radical-based for radicals and stroke based for individual strokes. However, the component-based methods face difficulties with corresponding segmentation. Instead of recognizing individual text components, the holistic approaches use gradient or directional features to recognize a whole character. Many methods have been proposed for holistic character recognition of CJK scripts like Statistical, structural, HMM, NN, SVM, MRF, etc. Word recognition in CJK scripts is difficult as the words are not separated clearly, as in Roman scripts. More sophisticated techniques of semantic understanding are required for this task. Also, more lexicons are required as the number of characters here is much larger than in Roman and other scripts. Mostly specific domains of word recognition have been proposed for Chinese scripts like Address recognition or Bank checks recognition [455]. Chinese : [26], [376], [382], Korean [371], [372], Japanese [215], [218], [235]

Arabic scripts: have the most complex structure of their components. The words in Arabic have within-word separators caused due to some characters. This space is smaller than usual space between words. Any such word has sub-words which are termed PAW. Thus, Arabic documents cannot be easily segmented into words and characters. The presence of PAWs is challenging for the Recognition of Arabic scripts. Most of the classification and recognition methods accommodate these peculiarities of the script into the system for efficient results. [191], [192], [337], Other tasks of DAR for Arabic scripts studied are Check Recognition [334].

Brahmi scripts: A unique characteristic of some Brahmi scripts is the *shirorekha*, a connecting line for words in scripts like Devanagari, Bengali, Gurmukhi, etc. In printed documents, a *shirorekha* helps segment words; however, variations due to handwriting are not effective in handwritten documents. Projection profiles (horizontal, vertical) are used for removing shirorekha in printed documents. In addition, some script-specific morphological methods are applied to handwritten documents. The general word segmentation methods used for Romans scripts are equally good for these scripts (with/without shirorekha). LS in handwritten documents is also complicated due to conjunct conjugates and vowel modifiers around the characters, due to which straight-line alignment is unclear. The conjunct consonants are partial characters combined with another consonant and represent some special phonemes. Different combinations of these consonants and vowel modifiers exist in Brahmi scripts like Kannada, Tamil, Devanagari, etc. (Figure 8). Thus, Semantics-based classification methods for various scripts have been proposed for character segmentation and recognition. Graph representations have also been proposed for Brahmi scripts to solve the problem of conjunct consonants. Tree-based classifiers for reducing many classes into a few by combining similar ones have also been explored for printed documents. The work reported in the literature for Brahmi scripts is very specific to the characteristic features of the different scripts. Hence no method can be generalized for all. Also, post-processing methods are necessary for improved recognition results. The major DAR works in Brahmi scripts are done for Bangla and Devanagari. The tasks in Bangla scripts are line, word, and character segmentation [147], [148], Online word segmentation [153], character recognition [213], [403], numeral recognition [456]. The Devanagari script DAR research include segmentation [142], [146], [169], [425], Conjunct consonants and overlapping consonants [424], Numeral recognition [193]. Tamil script recognition [398] The online recognition has been done for multi-scripts [14], [210], [216], Gujrati [211], end-to-end bangla recognition [19] etc.

All scripts discussed so far have different characteristics which play a positive or a negative role in the process of DAR. There are diacritics in Arabic and vowel modifiers in south Asian scripts. The presence of vowel modifiers and diacritics is an additional element that can be confusing to recognize or lead to erroneous results when missed. Also, different ways of writing styles, like cursive writing, complicate the learning for even the simpler scripts like Roman in offline and online systems. In some scripts, the context of the conversation decides the shape of the letters; for instance, in Persian and Arabic scripts, the combination of letters changes according to their position in a sentence. At the least, the logographic systems like CJK scripts, a very large set of characters is the major concern for all CJK recognition systems. The initial years of research saw major research in the Roman, Chinese, Japanese and Arabic scripts, focusing mostly on character and numeral recognition. Recent research has seen more work in Indian scripts.

VIII. Future Research Directions

The research for DAR has been a very long journey. It started with the problem of automatic recognition of printed characters for a very specific font. However, there have been extensive advancements since then; with the evolution of deep learning tools, the techniques have achieved state-of-the-art results for diverse tasks of DAR. With new developments, new gates have opened towards the research in various fields of DAR, which are discussed below:

1) Standardization of datasets and evaluation metrics

Standard datasets and evaluation metrics are pre-requisite for achieving common grounds to compare the algorithms. Recently, there have been efforts in the Standardization of datasets and metrics. However, only some metrics are useful for some applications. It requires researchers to test various metrics specific to the problem and conclude on the best one. A common platform for datasets of all scripts would allow researchers to access a wider range of datasets.

2) Integration of Language model

Current online and offline recognition systems do not always include language models. Higher-order language models (n-grams) in integration with the recognition systems have been explored to learn the grammar rules and improve performance. Dictionaries with very large open vocabularies have also been used to enhance recognition. However, these have yet to be tested on full-page printed, handwritten and online documents. Using a large-scale, higher-order language model efficiently would require approximation and optimization methods.

3) Multi-script and Multi-application models

The diverse peculiarities of scripts pose challenges in modelling multi-script recognition systems. The last decade has focused on multi-lingual models with more applications on India-based scripts. There is still ample scope for research towards developing end-to-end models that could switch between specifications per the script requirements. There are many tasks of DAR which could have a combined model. For instance, WI and handwritten Script Identification need to model the intricacies of the writing styles. This advantage could be modelled into a single system for a more generic DAR model.

4) Real-time detection and recognition

With the increased usage of electronic devices like mobile phones, etc., it is obvious to have approaches which could be utilized for real-time recognition tasks. The real-time processing requirements need to be addressed in the DAR models. Such algorithms would involve tracking mechanisms incorporating temporal cues for efficient text analysis and recognition accuracy.

5) Adaptation to new scripts and styles

The development of a recognition model considers the characteristics of the unique styles of the script in the application into the model. Adapting this model to a new script depends on two factors: 1) the degree of similarity of the scripts and 2) the Adaptability of the model for new data [37]. Approaches like Transfer Learning are useful in designing such systems, which is much less the cost of training and computations. These factors can help design models which are more open to adaptations.

6) End-to-end recognition systems

The step-wise segmentation methods have a drawback in that the consequent steps depend on the performance of the previous steps. End-to-end systems for DAR [194]–[196], [198]–[200] have reduced the intermediate steps while producing better transcripts of the text in document images. These models have paved the way for designing systems with full-page recognition abilities. They have the advantage of additional information at the initial steps to guide them better; for instance, the information on transcriptions in the detection steps aids the whole process, thus producing better transcriptions. The system design in [198] keeps the segmentation and recognition systems separate for training, i.e. the backpropagation errors do not concern the segmentation step. This loses the essence of the purpose of having an end-to-end system. Some challenges that need to be addressed are variations in number, location, and patterns of text regions; different reading orders in different scripts; for instance, Arabic script has right-to-left reading order. Better character recognition models with improved feedback mechanisms, language models and optimization techniques are required.

IX. Conclusion

This paper discusses offline or online, printed or handwritten, and constrained or unconstrained DAR methods. We have examined the challenges towards DAR with various perspectives such as 1) DAR perspective, 2) Online recognition perspective, 3) Image acquisition and related perspective. A detailed introduction to the basic tasks of DAR, i.e. 1) Pre-processing, 2)Segmentation, and 3) Recognition, has been presented with a brief insight into the related literary works. Pre-processing and segmentation play an important role in the recognition performance of the systems. However, in contrast to the traditional step-wise DAR, recent times have seen more focus on end-to-end and segmentation-free recognition systems. Lately, there has been a domination of deep learning approaches for different DAR tasks. The most used approaches in the history of DAR include SVM, HMM and NN, with a recent focus on CNN and RNNs. The efficiency of these methods depends on various factors, such as pre-processing and segmentation performance, length of the training samples, feature extraction techniques etc. An extensive analysis of datasets is also presented, covering offline and online, printed and handwritten, and historic and latest datasets. We have elaborated on numerous script-related challenges influencing the proposed systems' effectiveness. Roman, Chinese, Japanese, Arabic, and Brahmi scripts have been the research focus in the last decade. The handwritten text recognition is yet to be solved for Arabic and Asian Scripts. The Arabic scripts have issues relating to character segmentation due to the uncertainty of shapes a character can take in a sentence. Unlike Roman scripts, Asian scripts have a large character base, with most of them slightly different from each other. These scripts have confusing patterns, which are challenging for the recognition systems. Despite a long journey of DAR, along with some high-performance systems, new approaches can pave the way for developing generic systems for document recognition. We have presented some options for future research works in DAR.

References

[1] Y. Y. Tang, S.-W. Lee, and C. Y. Suen, "Automatic document processing: a survey," *Pattern recognition*, vol. 29, no. 12, pp. 1931–1952, 1996.

[2] D. Ghosh, T. Dube, and A. Shivaprasad, "Script recognition—a review," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2142–2161, 2010.

[3] D. Sinwar, V. S. Dhaka, N. Pradhan, and S. Pandey, "Offline script recognition from handwritten and printed multilingual documents: a survey," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 24, no. 1, pp. 97–121, 2021.

[4] A. Chaudhuri, K. Mandaviya, P. Badelia, and S. K. Ghosh, "Optical character recognition systems," in *Optical Character Recognition Systems for Different Languages with Soft Computing*. Springer, 2017, pp. 9–41.

[5] ——, "Optical character recognition systems," in *Optical Character Recognition Systems for Different Languages with Soft Computing*. Springer, 2017, pp. 9–41.

[6] N. Islam, Z. Islam, and N. Noor, "A survey on optical character recognition system," *arXiv preprint arXiv:1710.05703*, 2017.

[7] R. Smith *et al.*, "Tesseract ocr engine," *Lecture. Google Code. Google Inc*, 2007. [Online]. Available: https://github.com/tesseract-ocr/tesseract#tesseract-ocr

[8] D. Doermann, K. Tombre *et al.*, *Handbook of document image processing and recognition*. Springer, 2014, vol. 1.

[9] C. C. Tappert, C. Y. Suen, and T. Wakahara, "The state of the art in online handwriting recognition," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 8, pp. 787–808, 1990.

[10] R. Smith, D. Antonova, and D.-S. Lee, "Adapting the tesseract open source ocr engine for multilingual ocr," in *Proceedings of the International Workshop on Multilingual OCR*, 2009, pp. 1–8.

[11] N. Samadiani and H. Hassanpour, "A neural network-based approach for recognizing multi-font printed english characters," *Journal of Electrical Systems and Information Technology*, vol. 2, no. 2, pp. 207–218, 2015.

[12] F. Slimane, R. Ingold, and J. Hennebert, "Icdar2017 competition on multi-font and multi-size digitally represented arabic text," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 1466–1472.

[13] A. A. Aburas and M. E. Gumah, "Arabic handwriting recognition: Challenges and solutions," in *2008 International Symposium on Information Technology*, vol. 2. IEEE, 2008, pp. 1–6.

[14] H. Singh, R. K. Sharma, and V. Singh, "Online handwriting recognition systems for indic and non-indic scripts: a review," *Artificial Intelligence Review*, vol. 54, no. 2, pp. 1525–1579, 2021.

[15] Y. Zheng, H. Li, and D. Doermann, "Text identification in noisy document images using markov random model," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* IEEE, 2003, pp. 599–603.

[16] A. Farahmand, H. Sarrafzadeh, and J. Shanbehzadeh, "Document image noises and removal methods," *International MultiConference of Engineers and Computer Scientists 2013. Proceedings.*, vol. 1, 2013.

[17] M. Agrawal and D. Doermann, "Stroke-like pattern noise removal in binary document images," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 17–21.

[18] J. Liang, D. Doermann, and H. Li, "Camera-based analysis of text and documents: a survey," *International Journal of Document Analysis and Recognition (IJDAR)*, vol. 7, no. 2, pp. 84–104, 2005.

[19] F. B. Safir, A. Q. Ohi, M. F. Mridha, M. M. Monowar, and M. A. Hamid, "End-to-end optical character recognition for bengali handwritten words," in *2021 National Computing Colleges Conference (NCCC)*. IEEE, 2021, pp. 1–7.

[20] A. Delaye and C.-L. Liu, "Text/non-text classification in online handwritten documents with conditional random fields," in *Pattern Recognition: Chinese Conference, CCPR 2012, Beijing, China, September 24-26, 2012. Proceedings 5*. Springer Berlin Heidelberg, 2012, pp. 514–521.

[21] A. K. Jain and S. Bhattacharjee, "Text segmentation using gabor filters for automatic document processing," *Machine vision and applications*, vol. 5, no. 3, pp. 169–184, 1992.

[22] Z. Zhu, L. Gao, Y. Li, Y. Huang, L. Du, N. Lu, and X. Wang, "Ntable: A dataset for camera-based table detection," in *International Conference on Document Analysis and Recognition*. Springer, 2021, pp. 117–129.

[23] M. Li, L. Cui, S. Huang, F. Wei, M. Zhou, and Z. Li, "Tablebank: Table benchmark for image-based table detection and recognition," in *Proceedings of The 12th language resources and evaluation conference*, 2020, pp. 1918–1925.

[24] A. P. Deshpande, V. R. Potlapalli, and R. K. Sarvadevabhatla, "Meditables: A new dataset and deep network for multi-category table localization in medical documents," in *International Conference on Document Analysis and Recognition*. Springer, 2021, pp. 112–124.

[25] M. Khayyat, L. Lam, and C. Y. Suen, "Arabic handwritten word spotting using language models," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 43–48.

[26] H. Zhang and C.-L. Liu, "A lattice-based method for keyword spotting in online chinese handwriting," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 1064–1068.

[27] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke, "A novel word spotting method based on recurrent neural networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, 2012. [Online]. Available: https://doi.org/10.1109/TPAMI.2011.113

[28] H. Singh, R. K. Sharma, and V. Singh, "Online handwriting recognition systems for indic and non-indic scripts: a review," *Artificial Intelligence Review*, vol. 54, no. 2, pp. 1525–1579, 2021.

[29] S. M. Obaidullah, S. K. Das, and K. Roy, "A system for handwritten script identification from indian document," *Journal of Pattern Recognition Research*, vol. 8, no. 1, pp. 1–12, 2013.

[30] M. N. Abdi, M. Khemakhem, and H. Ben-Abdallah, "A novel approach for off-line arabic writer identification based on stroke feature combination," in *2009 24th International Symposium on Computer and Information Sciences*. IEEE, 2009, pp. 597–600.

[31] M. F. Mridha, A. Q. Ohi, J. Shin, M. M. Kabir, M. M. Monowar, and M. A. Hamid, "A thresholded gabor-cnn based writer identification system for indic scripts," *IEEE Access*, vol. 9, pp. 132 329–132 341, 2021.

[32] M. Bulacu and L. Schomaker, "Text-independent writer identification and verification using textural and allographic features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, 2007. [Online]. Available: https://doi.org/10.1109/TPAMI.2007.1009

[33] A. Shivram, C. Ramaiah, and V. Govindaraju, "A hierarchical bayesian approach to online writer identification," *IET Biometrics*, vol. 2, 2013. [Online]. Available: https://doi.org/10.1049/iet-bmt.2013.0017

[34] M. J. Khan, A. Yousaf, K. Khurshid, A. Abbas, and F. Shafait, "Automated forgery detection in multispectral document images using fuzzy clustering," in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*. IEEE, 2018, pp. 393–398.

[35] W. Wang, Z. Yan, and H. Lin, "A document image quality assessment method based on feature fusion," in *The International Conference on Image, Vision and Intelligent Systems (ICIVIS 2021)*. Springer, 2022, pp. 889–899.

[36] S. Nigam, A. Behera, S. Verma, and P. Nagabhushan, "Deformity removal from handwritten text documents using variable cycle gan," *PREPRINT (Version 1) available at Research Square*, pp. 1–16, 2022.

[37] S. Nigam, A. P. Behera, M. Gogoi, S. Verma, and P. Nagabhushan, "Strike off removal in indic scripts with transfer learning," *Neural Computing and Applications*, pp. 1–17, 2023.

[38] W. Bieniecki, S. Grabowski, and W. Rozenberg, "Image preprocessing for improving ocr accuracy," in *2007 international conference on perspective technologies and methods in MEMS design*. IEEE, 2007, pp. 75–80.

[39] H. P. Le and G. Lee, "Noise removal from binarized text images," in *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, vol. 3. IEEE, 2010, pp. 586–589.

[40] L. Xu, E. Oja, and P. Kultanen, "A new curve detection method: randomized hough transform (rht)," *Pattern recognition letters*, vol. 11, no. 5, pp. 331–338, 1990.

[41] A. Chakraborty and M. Blumenstein, "Marginal noise reduction in historical handwritten documents–a survey," in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*. IEEE, 2016, pp. 323–328.

[42] S.-C. Pei, M. Tzeng, and Y.-Z. Hsiao, "Enhancement of uneven lighting text image using line-based empirical mode decomposition," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 1249–1252.

[43] F. Z. A. Bella, M. El Rhabi, A. Hakim, and A. Laghrib, "Reduction of the non-uniform illumination using nonlocal variational models for document image analysis," *Journal of the Franklin Institute*, vol. 355, no. 16, pp. 8225–8244, 2018.

[44] C. Simon, I. K. Park *et al.*, "Correcting geometric and photometric distortion of document images on a smartphone," *Journal of Electronic Imaging*, vol. 24, no. 1, p. 013038, 2015.

[45] O. Nina, B. Morse, and W. Barrett, "A recursive otsu thresholding method for scanned document binarization," in *2011 IEEE Workshop on Applications of Computer Vision (WACV)*. IEEE, 2011, pp. 307–314.

[46] X. Chen, L. Lin, and Y. Gao, "Parallel nonparametric binarization for degraded document images," *Neurocomputing.*, vol. 189, 2016. [Online]. Available: https://doi.org/10.1016/j.neucom.2015.11.040

[47] E. H. B. Smith, L. Likforman-Sulem, and J. Darbon, "Effect of preprocessing on binarization," in *Document Recognition and Retrieval XVII*, vol. 7534. SPIE, 2010, pp. 154–161.

[48] W. Niblack, *An introduction to digital image processing*. Birkerod: Strandberg Publishing Company, 1985.

[49] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognit*, vol. 33, 2000. [Online]. Available: https://doi.org/10.1016/S0031-3203(99)00055-2

[50] N. Mitianoudis and N. Papamarkos, "Document image binarization using local features and gaussian mixture modeling," *Image Vis Comput.*, vol. 38, 2015. [Online]. Available: https://doi.org/10.1016/j.imavis.2015.04.003

[51] P. K. Sahoo, S. Soltani, and A. K. Wong, "A survey of thresholding techniques," *Computer vision, graphics, and image processing*, vol. 41, no. 2, pp. 233–260, 1988.

[52] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic imaging*, vol. 13, no. 1, pp. 146–165, 2004.

[53] N. Otsu, "A threshold selection method from gray-level histograms," *Trans Syst Man Cybern*, vol. 9, 1979. [Online]. Available: https://doi.org/10.1109/TSMC.1979.4310076

[54] R. F. Moghaddam and M. Cheriet, "Adotsu: an adaptive and parameterless generalization of otsu's method for document image binarization," *Pattern Recognit*, vol. 45, 2012. [Online]. Available: https://doi.org/10.1016/j.patcog.2011.12.013

[55] B. Su, S. Lu, and C. L. Tan, "Robust document image binarization technique for degraded document images," *Trans Image Process.*, vol. 22, 2013. [Online]. Available: https://doi.org/10.1109/TIP.2012.2231089

[56] F. Jia, C. Shi, K. He, C. Wang, and B. Xiao, "Degraded document image binarization using structural symmetry of strokes," *Pattern Recognit.*, vol. 74, 2018. [Online]. Available: https://doi.org/10.1016/j.patcog.2017.09.032

[57] M. Valizadeh and E. Kabir, "An adaptive water flow model for binarization of degraded document images," *Int J Doc Anal Recognit (IJDAR).*, vol. 16, 2013. [Online]. Available: https://doi.org/10.1007/s10032-012-0182-z

[58] A. Sehad, Y. Chibani, R. Hedjam, and M. Cheriet, "Gabor filter-based texture for ancient degraded document image binarization," *Pattern Analysis and Applications*, vol. 22, pp. 1–22, 2019.

[59] H. Z. Nafchi, R. F. Moghaddam, and M. Cheriet, "Phase-based binarization of ancient document images: Model and applications," *Trans Image Process.*, vol. 23, 2014. [Online]. Available: https://doi.org/10.1109/TIP.2014.2322451

[60] T. Lelore and F. Bouchara, "Fair: a fast algorithm for document image restoration," *Trans Pattern Anal Mach Intell*, vol. 35, 2013. [Online]. Available: https://doi.org/10.1109/TPAMI.2013.63

[61] A. Mishra, K. Alahari, and C. Jawahar, "Unsupervised refinement of color and stroke features for text binarization," *Int J Doc*

Anal Recognit (IJDAR)., vol. 20, 2017. [Online]. Available: https://doi.org/10.1007/s10032-017-0283-9

[62] N. R. Howe, "A laplacian energy for document binarization," in *2011 International conference on document analysis and recognition*. IEEE, 2011, pp. 6–10.

[63] K. R. Ayyalasomayajula and A. Brun, "Document binarization using topological clustering guided laplacian energy segmentation," in *2014 14th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2014, pp. 523–528.

[64] E. Ahmadi, Z. Azimifar, M. Shams, M. Famouri, and M. J. Shafiee, "Document image binarization using a discriminative structural classifier," *Pattern recognition letters*, vol. 63, pp. 36–42, 2015.

[65] S. Bhowmik, R. Sarkar, B. Das, and D. Doermann, "Gib: a game theory inspired binarization technique for degraded document images," *IEEE Trans Image Process.*, vol. 28, 2018. [Online]. Available: https://doi.org/10.1109/TIP.2018.2878959

[66] F. Kasmin, A. Abdullah, and A. S. Prabuwono, "Ensemble of steerable local neighbourhood grey-level information for binarization," *Pattern Recognit Lett.*, vol. 98, 2017. [Online]. Available: https://doi.org/10.1016/j.patrec.2017.07.014

[67] H. Hamza, E. Smigiel, and E. Belaid, "Neural based binarization techniques," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. IEEE, 2005, pp. 317–321.

[68] J. C. Rabelo, C. Zanchettin, C. A. Mello, and B. L. Bezerra, "A multi-layer perceptron approach to threshold documents with complex background," in *2011 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2011, pp. 2523–2530.

[69] J. Pastor-Pellicer, F. Zamora-Martínez, S. España-Boquera, and M. J. Castro-Bleda, "F-measure as the error function to train neural networks," in *Advances in Computational Intelligence: 12th International Work-Conference on Artificial Neural Networks, IWANN 2013, Puerto de la Cruz, Tenerife, Spain, June 12-14, 2013, Proceedings, Part I 12*. Springer, 2013, pp. 376–384.

[70] J. Pastor-Pellicer, S. España-Boquera, F. Zamora-Martínez, M. Z. Afzal, and M. J. Castro-Bleda, "Insights on the use of convolutional neural networks for document image binarization," in *Advances in Computational Intelligence: 13th International Work-Conference on Artificial Neural Networks, IWANN 2015, Palma de Mallorca, Spain, June 10-12, 2015. Proceedings, Part II 13*. Springer, 2015, pp. 115–126.

[71] J. Calvo-Zaragoza and A. J. Gallego, "A selectional auto-encoder approach for document image binarization," *Pattern Recognit.*, vol. 86, 2019. [Online]. Available: https://doi.org/10.1016/j.patcog.2018.08.011

[72] K. R. Ayyalasomayajula, F. Malmberg, and A. Brun, "Pdnet: semantic segmentation integrated with a primal-dual network for document binarization," *Pattern Recognit Lett.*, vol. 121, 2018. [Online]. Available: https://doi.org/10.1016/j.patrec.2018.05.011

[73] R. Mondal, D. Chakraborty, and B. Chanda, "Learning 2d morphological network for old document image binarization," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 65–70.

[74] S. Kang, B. K. Iwana, and S. Uchida, "Cascading modular u-nets for document image binarization," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 675–680.

[75] C. Tensmeyer, M. Brodie, D. Saunders, and T. Martinez, "Generating realistic binarization data with generative adversarial networks," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 172–177.

[76] J. Zhao, C. Shi, F. Jia, Y. Wang, and B. Xiao, "Document image binarization with cascaded generators of conditional generative adversarial networks," *Pattern Recognit.*, vol. 96, 2019. [Online]. Available: https://doi.org/10.1016/j.patcog.2019.106968

[77] M. Z. Afzal, J. Pastor-Pellicer, F. Shafait, T. M. Breuel, A. Dengel, and M. Liwicki, "Document image binarization using lstm: A sequence learning approach," in *Proceedings of the 3rd international workshop on historical document imaging and processing*, 2015, pp. 79–84.

[78] F. Westphal, N. Lavesson, and H. Grahn, "Document image binarization using recurrent neural networks," in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*. IEEE, 2018, pp. 263–268.

[79] A. K. Bhunia, A. K. Bhunia, A. Sain, and P. P. Roy, "Improving document binarization via adversarial noise-texture augmentation," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 2721–2725.

[80] A. Krantz and F. Westphal, "Cluster-based sample selection for document image binarization," in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, vol. 5. IEEE, 2019, pp. 47–52.

[81] W. Xiong, J. Xu, Z. Xiong, J. Wang, and M. Liu, "Degraded historical document image binarization using local features and support vector machine (svm)," *Optik*, vol. 164, 2018. [Online]. Available: https://doi.org/10.1016/j.ijleo.2018.02.072

[82] I. B. Messaoud, H. El Abed, H. Amiri, and V. Märgner, "New method for the selection of binarization parameters based on noise features of historical documents," in *Proceedings of the 2011 Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data*, 2011, pp. 1–8.

[83] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "A combined approach for the binarization of handwritten document images," *Pattern Recognit Lett*, vol. 35, 2014. [Online]. Available: https://doi.org/10.1016/j.patrec.2012.09.026

[84] Y. Liang, Z. Lin, L. Sun, and J. Cao, "Document image binarization via optimized hybrid thresholding," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2017, pp. 1–4.

[85] C. Tensmeyer and T. Martinez, "Historical document image binarization: a review," *SN Computer Science*, vol. 1, no. 3, pp. 1–26, 2020.

[86] B. Su, S. Lu, and C. L. Tan, "Combination of document image binarization techniques," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 22–26.

[87] W. A. Mustafa and M. M. M. A. Kader, "Binarization of document images: a comprehensive review," in *Journal of Physics: Conference Series*, vol. 1019. IOP Publishing, 2018, p. 012023.

[88] F. Bast, "Science information resources," *Science Reporter*, vol. 52, pp. 40–43, 04 2015.

[89] A. reddit user. (2020) A handwritten letter. [Online]. Available: https://www.reddit.com/r/Handwriting/comments/dnf4r3/a_handwritten_letter_partial/

[90] T.-A. Tran, I.-S. Na, and S.-H. Kim, "Separation of text and non-text in document layout analysis using a recursive filter," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 9, no. 10, pp. 4072–4091, 2015.

[91] V. P. Le, N. Nayef, M. Visani, J.-M. Ogier, and C. De Tran, "Text and non-text segmentation based on connected component features," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 1096–1100.

[92] L. Neumann and J. Matas, "Text localization in real-world images using efficiently pruned exhaustive search," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 687–691.

[93] C. Merino-Gracia, K. Lenc, and M. Mirmehdi, "A head-mounted device for recognizing text in natural scenes," in *International Workshop on Camera-Based Document Analysis and Recognition*. Springer, 2011, pp. 29–41.

[94] K. Chen, F. Yin, and C.-L. Liu, "Hybrid page segmentation with efficient whitespace rectangles extraction and grouping," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 958–962.

[95] K. Kise, A. Sato, and M. Iwata, "Segmentation of page images using the area voronoi diagram," *Computer Vision and Image Understanding*, vol. 70, no. 3, pp. 370–382, 1998.

[96] F. Y. Shih and S.-S. Chen, "Adaptive document block segmentation and classification," *IEEE transactions on systems, man, and cybernetics, part B (cybernetics)*, vol. 26, no. 5, pp. 797–802, 1996.

[97] H. Wei, M. Baechler, F. Slimane, and R. Ingold, "Evaluation of svm, mlp and gmm classifiers for layout analysis of historical documents," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 1220–1224.

[98] S. Marinai, M. Gori, and G. Soda, "Artificial neural networks for document analysis and recognition," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 1, pp. 23–35, 2005.

[99] A. Antonacopoulos, "Page segmentation using the description of the background," *Computer Vision and Image Understanding*, vol. 70, no. 3, pp. 350–369, 1998.

[100] K. Y. Wong, R. G. Casey, and F. M. Wahl, "Document analysis system," *IBM journal of research and development*, vol. 26, no. 6, pp. 647–656, 1982.

[101] G. Nagy and S. C. Seth, "Hierarchical representation of optically scanned documents," *CSE Conference and Workshop Papers*, 1984.

[102] L. O'Gorman, "The document spectrum for page layout analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 11, pp. 1162–1173, 1993.

[103] D. J. Ittner and H. S. Baird, "Language-free layout analysis," in *Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR'93)*. IEEE, 1993, pp. 336–340.

[104] A. Dias, "Minimum spanning trees for text segmentation," in *Proc. of Fifth Annual Symposium on Document Analysis and Information Retrieval, Las Vegas, Nevada*, 1996.

[105] M. Aiello, C. Monz, L. Todoran, M. Worring *et al.*, "Document understanding for a broad class of documents," *International Journal on Document Analysis and Recognition*, vol. 5, no. 1, pp. 1–16, 2002.

[106] J.-P. Mei, Y. Wang, L. Chen, and C. Miao, "Large scale document categorization with fuzzy clustering," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1239–1251, 2016.

[107] Y. Lu, Z. Wang, and C. L. Tan, "Word grouping in document images based on voronoi tessellation," in *Document Analysis Systems VI: 6th International Workshop, DAS 2004, Florence, Italy, September 8-10, 2004. Proceedings 6*. Springer, 2004, pp. 147–157.

[108] M. Arivazhagan, H. Srinivasan, and S. Srihari, "A statistical approach to line segmentation in handwritten documents," in *Document recognition and retrieval XIV*, vol. 6500. SPIE, 2007, pp. 245–255.

[109] A. Alaei, P. Nagabhushan, and U. Pal, "Piece-wise painting technique for line segmentation of unconstrained handwritten text: a specific study with persian text documents," *Pattern Anal Appl*, vol. 14, 2011.

[110] M. K. Jindal, R. K. Sharma, and G. S. Lehal, "Segmentation of horizontally overlapping lines in printed indian scripts," *Int J Comput Intell Res*, vol. 3, 2007.

[111] S. Jindal and G. S. Lehal, "Line segmentation of handwritten gurmukhi manuscripts," in *Proceeding of the workshop on document analysis and recognition*, 2012, pp. 74–78.

[112] R. Saabni, A. Asi, and J. El-Sana, "Text line extraction for historical document images," *Pattern Recognit Lett*, vol. 35, 2014.

[113] C. A. Boiangiu, M. C. Tanase, and R. Ioanitescu, "Handwritten documents text line segmentation based on information energy," *Int J Comput Commun Control*, vol. 9, 2014.

[114] N. Arvanitopoulos and S. Süsstrunk, "Seam carving for text line extraction on color and grayscale historical manuscripts," in *2014 14th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2014, pp. 726–731.

[115] A. Jindal and R. Ghosh, "Text line segmentation in indian ancient handwritten documents using faster r-cnn," *Multimedia Tools and Applications*, pp. 1–20, 2022.

[116] U. Garain, T. Paquet, and L. Heutte, "On foreground-background separation in low quality color document images," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, 2005, pp. 585–589 Vol. 2.

[117] S. S. Bukhari, F. Shafait, and T. M. Breuel, "Text-line extraction using a convolution of isotropic gaussian filter with a set of line filters," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 579–583.

[118] K. Wong, R. Casey, and F. Wahl, "Document analysis systems," *IBM J Res Dev*, vol. 26, 1982.

[119] D. J. Kennard and W. A. Barrett, "Separating lines of text in free-form handwritten historical documents," in *Second International Conference on Document Image Analysis for Libraries (DIAL'06)*. IEEE, 2006, pp. 12–pp.

[120] R. Gomathi, R. S. Uma, and S. Mohanval, "Segmentation of touching, overlapping, skewed and short handwritten text lines," *Int J Comput Appl*, vol. 49, 2012.

[121] E. J. Almazan, R. Tal, Y. Qian, and J. H. Elder, "Mcmlsd: A dynamic programming approach to line segment detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2031–2039.

[122] E. Hussain, A. Hannan, and K. Kashyap, "A zoning based feature extraction method for recognition of handwritten assamese characters," *Int J Comput Sci Technol*, vol. 6, 2015.

[123] L. Likforman-Sulem, A. Hanimyan, and C. Faure, "A hough based algorithm for extracting text lines in handwritten documents," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 2. IEEE, 1995, pp. 774–777.

[124] G. Louloudis, B. Gatos, I. Pratikakis, and C. Halatsis, "Text line and word segmentation of handwritten documents," *Pattern Recognit*, vol. 42, 2009.

[125] K. Sesh Kumar, A. M. Namboodiri, and C. Jawahar, "Learning segmentation of documents with complex scripts," in *Computer Vision, Graphics and Image Processing: 5th Indian Conference, ICVGIP 2006, Madurai, India, December 13-16, 2006. Proceedings*. Springer, 2006, pp. 749–760.

[126] F. Yin and C. L. Liu, "Handwritten chinese text line segmentation by clustering with distance metric learning," *Pattern Recognit*, vol. 42, 2009.

[127] S. Jetley, S. Belhe, V. K. Koppula, and A. Negi, "Two-stage hybrid binarization around fringe map based text line segmentation for document images," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 2012, pp. 343–346.

[128] D. Brodic, "Text line segmentation with water flow algorithm based on power function," *J Electr Eng*, vol. 66, 2015.

[129] B. Gatos, G. Louloudis, and N. Stamatopoulos, "Segmentation of historical handwritten documents into text zones and text lines," in *2014 14th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2014, pp. 464–469.

[130] S. Saha, S. Basu, M. Nasipuri, and D. K. Basu, "A hough transform based technique for text segmentation," *J Comput*, vol. 2, 2010.

[131] K. K. Kim, J. H. Kim, and C. Y. Suen, "Recognition of unconstrained handwritten numeral strings by composite segmentation method," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, vol. 2. IEEE, 2000, pp. 594–597.

[132] S. A. Angadi and M. Kodabagi, "A robust segmentation technique for line, word and character extraction from kannada text in low resolution display board images," *International Journal of Image and Graphics*, vol. 14, no. 01n02, p. 1450003, 2014.

[133] R. Manmatha and J. L. Rothfeder, "A scale space approach for automatically segmenting words from historical handwritten documents," *IEEE Trans Pattern Anal Mach Intell*, vol. 27, 2005.

[134] Y.-H. Tseng and H.-J. Lee, "Recognition-based handwritten chinese character segmentation using a probabilistic viterbi algorithm," *Pattern Recognition Letters*, vol. 20, no. 8, pp. 791–806, 1999.

[135] k. Manjusha, S. Kumar, J. Rajendran, and K. P. Soman, "Hindi character segmentation in document images using level set methods and non-linear diffusion," *Int J Comput Appl*, vol. 44, 2012.

[136] S. Karthik, V. Hemanth, V. Balaji, and K. Soman, "Level set methodology for tamil document image binarization and segmentation," *International Journal of Computer Applications*, vol. 39, no. 9, pp. 7–12, 2012.

[137] G. Yang, Z. Yan, and H. Zhao, "Touching string segmentation using mrf," in *2009 International Conference on Computational Intelligence and Security*, vol. 2. IEEE, 2009, pp. 520–524.

[138] A. Rehman, D. Mohamad, F. Kurniawan, and M. Ilays, "Performance analysis of segmentation approach for cursive handwriting on benchmark database," in *2009 IEEE/ACS International Conference on Computer Systems and Applications*. IEEE, 2009, pp. 265–270.

[139] K. K. Kim, J. H. Kim, and C. Y. Suen, "Segmentation-based recognition of handwritten touching pairs of digits using structural features," *Pattern Recognition Letters*, vol. 23, no. 1-3, pp. 13–24, 2002.

[140] Y. Lei, C. Liu, X. Ding, and Q. Fu, "A recognition based system for segmentation of touching handwritten numeral strings," in *Ninth International Workshop on Frontiers in Handwriting Recognition*. IEEE, 2004, pp. 294–299.

[141] I. B. Messaoud, H. Amiri, H. El Abed, and V. Märgner, "A multilevel text-line segmentation framework for handwritten historical documents," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 515–520.

[142] S. R. Narang, M. Jindal, and M. Kumar, "Line segmentation of devanagari ancient manuscripts," *Proceedings of the national academy of sciences, India section A: physical sciences*, vol. 90, no. 4, pp. 717–724, 2020.

[143] A. Raza, I. Siddiqi, A. Abidi, and F. Arif, "An unconstrained benchmark urdu handwritten sentence database with automatic line segmentation," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 491–496.

[144] I. B. Messaoud, H. Amiri, H. El Abed, and V. Märgner, "A multilevel text-line segmentation framework for handwritten historical documents," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 515–520.

[145] R. Saabni, A. Asi, and J. El-Sana, "Text line extraction for historical document images," *Pattern Recognit. Lett.*, vol. 35, 2014. [Online]. Available: https://doi.org/10.1016/j.patrec.2013.07.007

[146] R. P. Kaur, M. Jindal, and M. Kumar, "Txtlineseg: text line segmentation of unconstrained printed text in devanagari script," in *Computational methods and data engineering*. Springer, 2021, pp. 85–100.

[147] P. Rakshit, C. Halder, S. Ghosh, and K. Roy, "Line, word, and character segmentation from bangla handwritten text—a precursor toward bangla hocr," in *Advanced Computing and Systems for Security*. Springer, 2018, pp. 109–120.

[148] K. Agarwal, A. Mantry, and C. Halder, "Word segmentation of offline handwritten bangla text lines," in *Proceedings of International Conference on Advanced Computing Applications*. Springer, 2022, pp. 551–560.

[149] W. Sun and J. Xu, "Enhancing chinese word segmentation using unlabeled data," in *Proceedings of the 2011 Conference on Empirical Methods in natural language Processing*, 2011, pp. 970–979.

[150] G. Olivier, H. Miled, K. Romeo, and Y. Lecourtier, "Segmentation and coding of arabic handwritten words," in *Proceedings of 13th International Conference on Pattern Recognition*, vol. 3. IEEE, 1996, pp. 264–268.

[151] R. Safabakhsh and P. Adibi, "Nastaaligh handwritten word recognition using a continuous-density variable-duration hmm," *Arabian Journal for Science and Engineering*, vol. 30, no. 1, pp. 95–120, 2005.

[152] M. Makridis, N. Nikolaou, and B. Gatos, "An efficient word segmentation technique for historical and degraded machine-printed documents," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 1. IEEE, 2007, pp. 178–182.

[153] S. Sen, S. Chowdhury, M. Mitra, F. Schwenker, R. Sarkar, and K. Roy, "A novel segmentation technique for online handwritten bangla words," *Pattern Recognition Letters*, vol. 139, pp. 26–33, 2020.

[154] N. Mehta and J. Doshi, "Shirorekha based character segmentation for medieval handwritten devnagari manuscript," *International Journal of Information Technology*, vol. 13, no. 3, pp. 905–909, 2021.

[155] H. Liu and X. Ding, "Handwritten character recognition using gradient feature and quadratic classifier with multiple discrimination schemes," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. IEEE, 2005, pp. 19–23.

[156] C. Suen and E. Huang, "Computational analysis of the structural compositions of frequently used chinese characters," *Computer Processing of Chinese and Oriental Languages*, vol. 1, no. 3, pp. 163–176, 1984.

[157] K.-T. Tang and H. Leung, "Reconstructing strokes and writing sequences from chinese character images," in *2007 International Conference on Machine Learning and Cybernetics*, vol. 1. IEEE, 2007, pp. 160–165.

[158] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional neural network committees for handwritten character classification," in *2011 International conference on document analysis and recognition*. IEEE, 2011, pp. 1135–1139.

[159] A.-L. Bianne-Bernard, F. Menasri, R. A.-H. Mohamad, C. Mokbel, C. Kermorvant, and L. Likforman-Sulem, "Dynamic and contextual information in hmm modeling for handwritten word recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 10, pp. 2066–2080, 2011.

[160] D. Chen, J.-M. Odobez, and H. Bourlard, "Text detection and recognition in images and video frames," *Pattern recognition*, vol. 37, no. 3, pp. 595–608, 2004.

[161] K. Sobottka, H. Kronenberg, T. Perroud, and H. Bunke, "Text extraction from colored book and journal covers," *International Journal on Document Analysis and Recognition*, vol. 2, no. 4, pp. 163–176, 2000.

[162] P. P. Roy, J. Llados, and U. Pal, "Text/graphics separation in color maps," in *2007 International Conference on Computing: Theory and Applications (ICCTA'07)*. IEEE, 2007, pp. 545–551.

[163] N. Nikolaou and N. Papamarkos, "Color reduction for complex document images," *International Journal of Imaging Systems and Technology*, vol. 19, no. 1, pp. 14–26, 2009.

[164] J. Park, G. Lee, E. Kim, J. Lim, S. Kim, H. Yang, M. Lee, and S. Hwang, "Automatic detection and recognition of korean text in outdoor signboard images," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1728–1739, 2010.

[165] P. Shivakumara, W. Huang, and C. L. Tan, "Efficient video text detection using edge features," in *2008 19th International Conference on Pattern Recognition*. IEEE, 2008, pp. 1–4.

[166] M. Li and C. Wang, "An adaptive text detection approach in images and video frames," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 2008, pp. 72–77.

[167] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. Ieee, 2005, pp. 886–893.

[168] W. Jiang, Z.-Y. Lu, J. Li, and X.-P. Liu, "Text localization algorithm in complex scene based on corner-type feature and histogram of oriented gradients of edge magnitude statistical feature," *Journal of Jilin University(Engineering and Technology Edition)*, vol. 43, no. 1, pp. 250–255, 2013.

[169] S. P. Deore and A. Pravin, "Histogram of oriented gradients based off-line handwritten devanagari characters recognition using svm, k-nn and nn classifiers." *Rev. d'Intelligence Artif.*, vol. 33, no. 6, pp. 441–446, 2019.

[170] S. A. Ali and A. T. Hashim, "Wavelet transform based technique for text image localization," *Karbala International Journal of Modern Science*, vol. 2, no. 2, pp. 138–144, 2016.

[171] M. Pavithra and V. M. Aradhya, "A comprehensive of transforms, gabor filter and k-means clustering for text detection in images and video," *Applied computing and informatics*, pp. 1–15, 2014.

[172] O. K. Oyedotun and A. Khashman, "Document segmentation using textural features summarization and feedforward neural network," *Applied Intelligence*, vol. 45, no. 1, pp. 198–212, 2016.

[173] A. Antonacopoulos and R. Ritchings, "Representation and classification of complex-shaped printed regions using white tiles," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 2. IEEE, 1995, pp. 1132–1135.

[174] S. Bhowmik, R. Sarkar, and M. Nasipuri, "Text and non-text separation in handwritten document images using local binary pattern operator," in *Proceedings of the First International Conference on Intelligent Computing and Communication*. Springer, 2017, pp. 507–515.

[175] C. Strouthopoulos, N. Papamarkos, and A. Atsalakis, "Text extraction in complex color documents," *Pattern Recognition*, vol. 35, no. 8, pp. 1743–1758, 2002.

[176] A. L. Spitz, "Determination of the script and language content of document images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 235–245, 1997.

[177] N. Kato, M. Suzuki, S. Omachi, H. Aso, and Y. Nemoto, "A handwritten character recognition system using directional element feature and asymmetric mahalanobis distance," *IEEE transactions on pattern analysis and machine intelligence*, vol. 21, no. 3, pp. 258–262, 1999.

[178] P. Natarajan, Z. Lu, R. Schwartz, I. Bazzi, and J. Makhoul, "Multilingual machine printed ocr," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 15, no. 01, pp. 43–63, 2001.

[179] H. Xue and V. Govindaraju, "Hidden markov models combining discrete symbols and continuous attributes in handwriting recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 3, pp. 458–462, 2006.

[180] J. Chen, H. Cao, R. Prasad, A. Bhardwaj, and P. Natarajan, "Gabor features for offline arabic handwriting recognition," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, 2010, pp. 53–58.

[181] M. Amrouch, M. Rabi, and Y. Es-Saady, "Convolutional feature learning and cnn based hmm for arabic handwriting recognition," in *International conference on image and signal processing*. Springer, 2018, pp. 265–274.

[182] H. Bunke, S. Bengio, and A. Vinciarelli, "Offline recognition of unconstrained handwritten texts using hmms and statistical language models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, 2004. [Online]. Available: https://doi.org/10.1109/TPAMI.2004.14

[183] H. Choudhury, S. Mandal, S. Devnath, S. M. Prasanna, and S. Sundaram, "Combining hmm and svm based stroke classifiers for online assamese handwritten character recognition," in *2015 Annual IEEE India Conference (INDICON)*. IEEE, 2015, pp. 1–6.

[184] L. Guichard, A. H. Toselli, and B. Coüasnon, "Handwritten word verification by svm-based hypotheses re-scoring and multiple thresholds rejection," in *2010 12th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2010, pp. 57–62.

[185] Z. Chi and K. Wong, "A two-stage binarization approach for document images," in *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing. ISIMP 2001 (IEEE Cat. No. 01EX489)*. IEEE, 2001, pp. 275–278.

[186] S. Gunter and H. Bunke, "Ensembles of classifiers for handwritten word recognition," *Int. J. Doc. Anal. Recognit.*, vol. 5, 2003. [Online]. Available: https://doi.org/10.1007/s10032-002-0088-2

[187] P. Dreuw, P. Doetsch, C. Plahl, and H. Ney, "Hierarchical hybrid mlp/hmm or rather mlp features for a discriminatively trained gaussian hmm: a comparison for offline handwriting recognition," in *2011 18th IEEE International Conference on Image Processing*. IEEE, 2011, pp. 3541–3544.

[188] O. Morillot, L. Likforman-Sulem, and E. Grosicki, "Comparative study of hmm and blstm segmentation-free approaches for the recognition of handwritten text-lines," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 783–787.

[189] O. Alsharif and J. Pineau, "End-to-end text recognition with hybrid hmm maxout models," *arXiv preprint arXiv:1310.1811*, 2013.

[190] M. Liwicki and H. Bunke, "Hmm-based on-line recognition of handwritten whiteboard notes," in *Tenth international workshop on frontiers in handwriting recognition*. Suvisoft, 2006.

[191] M. Pechwitz and V. Maergner, "Hmm based approach for handwritten arabic word recognition using the ifn/enit-database," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, vol. 3. IEEE Computer Society, 2003, pp. 890–890.

[192] A. . H. M. R, L. Likforman-Sulem, and C. Mokbel, "Combining slanted-frame classifiers for improved hmm-based arabic handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, 2009. [Online]. Available: https://doi.org/10.1109/TPAMI.2008.136

[193] U. Bhattacharya, S. Parui, B. Shaw, and K. Bhattacharya, "Neural combination of ann and hmm for handwritten devanagari numeral recognition," in *Tenth international workshop on frontiers in handwriting recognition*. Suvisoft, 2006.

[194] M. Carbonell, J. Mas, M. Villegas, A. Fornés, and J. Lladós, "End-to-end handwritten text detection and transcription in full pages," in *2019 International conference on document analysis and recognition workshops (ICDARW)*, vol. 5. IEEE, 2019, pp. 29–34.

[195] M. Carbonell, A. Fornés, M. Villegas, and J. Lladós, "A neural model for text localization, transcription and named entity recognition in full pages," *Pattern Recognition Letters*, vol. 136, pp. 219–227, 2020.

[196] J. Chung and T. Delteil, "A computationally efficient pipeline approach to full page offline handwritten text recognition," in *2019 International conference on document analysis and recognition workshops (ICDARW)*, vol. 5. IEEE, 2019, pp. 35–40.

[197] M. Liao, B. Shi, X. Bai, X. Wang, and W. Liu, "Textboxes: A fast text detector with a single deep neural network," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, 2017.

[198] B. Moysset, C. Kermorvant, and C. Wolf, "Full-page text recognition: Learning where to start and when to stop," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 871–876.

[199] C. Wigington, C. Tensmeyer, B. Davis, W. Barrett, B. Price, and S. Cohen, "Start, follow, read: End-to-end full-page handwriting recognition," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 367–383.

[200] C. Tensmeyer and C. Wigington, "Training full-page handwritten text recognition models without annotated line breaks," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 1–8.

[201] S. Long, J. Ruan, W. Zhang, X. He, W. Wu, and C. Yao, "Textsnake: A flexible representation for detecting text of arbitrary shapes," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 20–36.

[202] T. Bluche, "Joint line segmentation and transcription for end-to-end handwritten paragraph recognition," *Advances in neural information processing systems*, vol. 29, 2016.

[203] T. Bluche, J. Louradour, and R. Messina, "Scan, attend and read: End-to-end handwritten paragraph recognition with mdlstm attention," in *2017 14th IAPR international conference on document analysis and recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 1050–1055.

[204] M. Schall, M.-P. Schambach, and M. O. Franz, "Multi-dimensional connectionist classification: reading text in one step," in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*. IEEE, 2018, pp. 405–410.

[205] M. Yousef and T. E. Bishop, "Origaminet: weakly-supervised, segmentation-free, one-step, full page text recognition by learning to unfold," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 710–14 719.

[206] D. Coquenet, C. Chatelain, and T. Paquet, "End-to-end handwritten paragraph text recognition using a vertical attention network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[207] L. Kang, P. Riba, M. Villegas, A. Fornés, and M. Rusiñol, "Candidate fusion: Integrating language modelling into a sequence-to-sequence handwritten word recognition architecture," *Pattern Recognition*, vol. 112, p. 107790, 2021.

[208] S. Jaeger, S. Manke, J. Reichert, and A. Waibel, "Online handwriting recognition: the npen++ recognizer," *International Journal on Document Analysis and Recognition*, vol. 3, pp. 169–180, 2001.

[209] A. K. Bhunia, S. Mukherjee, A. Sain, A. K. Bhunia, P. P. Roy, and U. Pal, "Indic handwritten script identification using offline-online multi-modal deep network," *Information Fusion*, vol. 57, pp. 1–14, 2020.

[210] H. Swethalakshmi, A. Jayaraman, V. S. Chakravarthy, and C. C. Sekhar, "Online handwritten character recognition of devanagari and telugu characters using support vector machines," in *Tenth international workshop on Frontiers in handwriting recognition*. Suvisoft, 2006.

[211] C. C. Gohel, M. M. Goswami, and V. K. Prajapati, "On-line handwritten gujarati character recognition using low level stroke," in *2015 Third International Conference on Image Information Processing (ICIIP)*. IEEE, 2015, pp. 130–134.

[212] M. Yadav, D. Mangal, S. Natesan, M. Paprzycki, and M. Ganzha, "Assamese character recognition using convolutional neural networks," in *Proceedings of 2nd International Conference on Artificial Intelligence: Advances and Applications*. Springer, 2022, pp. 851–859.

[213] S. Sen, D. Shaoo, S. Paul, R. Sarkar, and K. Roy, "Online handwritten bangla character recognition using cnn: a deep learning approach," in *Intelligent Engineering Informatics*. Springer, 2018, pp. 413–420.

[214] E. H. Ratzlaff, "Methods, reports and survey for the comparison of diverse isolated character recognition results on the unipen database," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* IEEE, 2003, pp. 623–628.

[215] M. Nakagawa, B. Zhu, and M. Onuma, "A model of on-line handwritten japanese text recognition free from line direction and writing format constraints," *IEICE Trans. Inf. Syst*, vol. 88, 2005. [Online]. Available: https://doi.org/10.1093/ietisy/e88-d.8.1815

[216] A. M. Namboodiri and A. K. Jain, "Online handwritten script recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 124–130, 2004.

[217] T. Mondal, U. Bhattacharya, S. K. Parui, K. Das, and D. Mandalapu, "On-line handwriting recognition of indian scripts - the first benchmark," in *2010 12th International Conference on Frontiers in Handwriting Recognition*, 2010, pp. 200–205.

[218] B. Zhu, X.-. D. Zhou, C.-. L. Liu, and M. Nakagawa, "A robust model for on-line handwritten japanese text recognition," *Int. J. Doc. Anal. Recognit.*, vol. 13, 2010. [Online]. Available: https://doi.org/10.1007/s10032-009-0111-y

[219] V. Carbune, P. Gonnet, T. Deselaers, H. A. Rowley, A. Daryin, M. Calvo, L.-L. Wang, D. Keysers, S. Feuz, and P. Gervais, "Fast multi-language lstm-based online handwriting recognition," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 23, no. 2, pp. 89–102, 2020.

[220] V. Venugopal and S. Sundaram, "An online writer identification system using regression-based feature normalization and codebook descriptors," *Expert Systems with Applications*, vol. 72, pp. 196–206, 2017.

[221] G. X. Tan, C. Viard-Gaudin, and A. C. Kot, "Automatic writer identification framework for online handwritten documents using character prototypes," *Pattern Recognit.*, vol. 42, 2009. [Online]. Available: https://doi.org/10.1016/j.patcog.2008.12.019

[222] R. Hussain, A. Raza, I. Siddiqi, K. Khurshid, and C. Djeddi, "A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 1–24, 2015.

[223] K. Nikolaidou, M. Seuret, H. Mokayed, and M. Liwicki, "A survey of historical document image datasets," *arXiv preprint arXiv:2203.08504*, 2022.

[224] X. Zhong, E. ShafieiBavani, and A. Jimeno Yepes, "Image-based table recognition: data, model, and evaluation," in *European Conference on Computer Vision*. Springer, 2020, pp. 564–580.

[225] N. Siegel, N. Lourie, R. Power, and W. Ammar, "Extracting scientific figures with distantly supervised neural networks," in *Proceedings of the 18th ACM/IEEE on joint conference on digital libraries*, 2018, pp. 223–232.

[226] B. Pfitzmann, C. Auer, M. Dolfi, A. S. Nassar, and P. W. Staar, "Doclaynet: A large human-annotated dataset for document-layout analysis," *arXiv preprint arXiv:2206.01062*, 2022.

[227] V. Romero, A. Fornés, N. Serrano, J. A. Sánchez, A. H. Toselli, V. Frinken, E. Vidal, and J. Lladós, "The esposalles database: An ancient marriage license corpus for off-line handwriting recognition," *Pattern Recognition*, vol. 46, no. 6, pp. 1658–1669, 2013.

[228] T. Grüning, R. Labahn, M. Diem, F. Kleber, and S. Fiel, "Read-bad: A new dataset and evaluation scheme for baseline detection in archival documents," in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*. IEEE, 2018, pp. 351–356.

[229] F. Simistira, M. Seuret, N. Eichenberger, A. Garz, M. Liwicki, and R. Ingold, "Diva-hisdb: A precisely annotated large dataset of challenging medieval manuscripts," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2016, pp. 471–476.

[230] M. Kassis, A. Abdalhaleem, A. Droby, R. Alaasam, and J. El-Sana, "Vml-hd: The historical arabic documents dataset for recognition systems," in *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*. IEEE, 2017, pp. 11–14.

[231] H. Mohammed, I. Marthot-Santaniello, and V. Märgner, "Grk-papyri: a dataset of greek handwriting on papyri for the task of writer

identification," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 726–731.

[232] M. W. A. Kesiman, J.-C. Burie, G. N. M. A. Wibawantara, I. M. G. Sunarya, and J.-M. Ogier, "Amadi_lontarset: The first handwritten balinese palm leaf manuscripts dataset," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2016, pp. 168–173.

[233] J. Hajič and P. Pecina, "The muscima++ dataset for handwritten optical music recognition," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 39–46.

[234] A. Fischer, "Iam-histdb a dataset of handwritten historical documents," in *HANDWRITTEN HISTORICAL DOCUMENT ANALYSIS, RECOGNITION, AND RETRIEVAL—STATE OF THE ART AND FUTURE TRENDS*. World Scientific, 2021, pp. 11–23.

[235] Z. Shen, K. Zhang, and M. Dell, "A large dataset of historical japanese documents with complex layouts," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 548–549.

[236] H. Kusetogullari, A. Yavariabdi, A. Cheddad, H. Grahn, and J. Hall, "Ardis: a swedish historical handwritten digit dataset," *Neural Computing and Applications*, vol. 32, no. 21, pp. 16 505–16 518, 2020.

[237] T. Constum, N. Kempf, T. Paquet, P. Tranouez, C. Chatelain, S. Brée, and F. Merveille, "Recognition and information extraction in historical handwritten tables: Toward understanding early century paris census," in *International Workshop on Document Analysis Systems*. Springer, 2022, pp. 143–157.

[238] J. A. Sánchez, V. Romero, A. H. Toselli, M. Villegas, and E. Vidal, "A set of benchmarks for handwritten text recognition on historical documents," *Pattern Recognition*, vol. 94, pp. 122–134, 2019.

[239] A. Antonacopoulos, C. Clausner, C. Papadopoulos, and S. Pletschacher, "Icdar 2013 competition on historical newspaper layout analysis (hnla 2013)," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 1454–1458.

[240] M. Murdock, S. Reid, B. Hamilton, and J. Reese, "Icdar 2015 competition on text line detection in historical documents," in *2015 13th international conference on document analysis and recognition (ICDAR)*. IEEE, 2015, pp. 1171–1175.

[241] F. Cloppet, V. Eglin, D. Stutzmann, N. Vincent *et al.*, "Icfhr2016 competition on the classification of medieval handwritings in latin script," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2016, pp. 590–595.

[242] F. Cloppet, V. Eglin, M. Helias-Baron, C. Kieu, N. Vincent, and D. Stutzmann, "Icdar2017 competition on the classification of medieval handwritings in latin script," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 1371–1376.

[243] C. Clausner, A. Antonacopoulos, N. Mcgregor, and D. Wilson-Nunn, "Icfhr 2018 competition on recognition of historical arabic scientific manuscripts–rasm2018," in *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2018, pp. 471–476.

[244] M. W. A. Kesiman, D. Valy, J. C. Burie, E. Paulus, M. Suryani, S. Hadi, M. Verleysen, S. Chhun, and J.-M. Ogier, "Icfhr 2018 competition on document image analysis tasks for southeast asian palm leaf manuscripts," in *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2018, pp. 483–488.

[245] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "Icdar2017 competition on document image binarization (dibco 2017)," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 1395–1403.

[246] L. Gao, Y. Huang, H. Déjean, J.-L. Meunier, Q. Yan, Y. Fang, F. Kleber, and E. Lang, "Icdar 2019 competition on table detection and recognition (ctdar)," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 1510–1515.

[247] R. Saini, D. Dobson, J. Morrey, M. Liwicki, and F. S. Liwicki, "Icdar 2019 historical document reading challenge on large structured chinese family records," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 1499–1504.

[248] C. Clausner, A. Antonacopoulos, T. Derrick, and S. Pletschacher, "Icdar2019 competition on recognition of early indian printed documents–reid2019," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 1527–1532.

[249] V. Christlein, A. Nicolaou, M. Seuret, D. Stutzmann, and A. Maier, "Icdar 2019 competition on image retrieval for historical handwritten documents," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 1505–1509.

[250] M. Seuret, A. Nicolaou, D. Stutzmann, A. Maier, and V. Christlein, "Icfhr 2020 competition on image retrieval for historical handwritten fragments (hisfrag20) dataset," *ICFHR 2020*, 2021.

[251] A. Scius-Bertrand, S. Gabay, J. Janes, L. Petkovic, C. Corbieres, and T. Clérice, "The bir database–identifying typographic emphasis in list-like historical documents," in *The 6th International Workshop on Historical Document Imaging and Processing*, 2021, pp. 37–42.

[252] J. Fang, X. Tao, Z. Tang, R. Qiu, and Y. Liu, "Dataset, ground-truth and performance metrics for table detection evaluation," in *2012 10th IAPR International Workshop on Document Analysis Systems*. IEEE, 2012, pp. 445–449.

[253] M. Li, Y. Xu, L. Cui, S. Huang, F. Wei, Z. Li, and M. Zhou, "Docbank: A benchmark dataset for document layout analysis," *arXiv preprint arXiv:2006.01038*, 2020.

[254] A. Mondal, P. Lipps, and C. Jawahar, "Iiit-ar-13k: a new dataset for graphical object detection in documents," in *International Workshop on Document Analysis Systems*. Springer, 2020, pp. 216–230.

[255] M. Göbel, T. Hassan, E. Oro, and G. Orsi, "Icdar 2013 table competition," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 1449–1453.

[256] L. Gao, X. Yi, Z. Jiang, L. Hao, and Z. Tang, "Icdar2017 competition on page object detection," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 1. IEEE, 2017, pp. 1417–1422.

[257] L. Gao, Y. Huang, H. Déjean, J.-L. Meunier, Q. Yan, Y. Fang, F. Kleber, and E. Lang, "Icdar 2019 competition on table detection and recognition (ctdar)," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 1510–1515.

[258] X. Zhong, J. Tang, and A. J. Yepes, "Publaynet: largest dataset ever for document layout analysis," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2019, pp. 1015–1022.

[259] T. Furukawa, "Recognition of laser-printed characters based on creation of new laser-printed characters datasets," in *International Conference on Document Analysis and Recognition*. Springer, 2021, pp. 407–421.

[260] H. S. Kawoosa, M. Singh, M. M. Joshi, and P. Goyal, "Ncert5k-iitrpr: A benchmark dataset for non-textual component detection in school books," in *International Workshop on Document Analysis Systems*. Springer, 2022, pp. 461–475.

[261] J. J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, 1994. [Online]. Available: https://doi.org/10.1109/34.291440

[262] S. Singh and M. Hewitt, "Cursive digit and character recognition in cedar database," in *Proceedings 15th international conference on pattern recognition. ICPR-2000*, vol. 2. IEEE, 2000, pp. 569–572.

[263] T. M. Ha and H. Bunke, "Off-line, handwritten numeral recognition by perturbation method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, 1997. [Online]. Available: https://doi.org/10.1109/34.589216

[264] M. Blumenstein and B. Verma, "Analysis of segmentation performance on the cedar benchmark database," in *Proceedings of sixth international conference on document analysis and recognition*. IEEE, 2001, pp. 1142–1146.

[265] M. Cheriet, R. Thibault, and R. Sabourin, "A multi-resolution based approach for handwriting segmentation in gray-scale images," in *Proceedings of 1st International Conference on Image Processing*, vol. 1. IEEE, 1994, pp. 159–163.

[266] B. Verma, P. Gader, and W. Chen, "Fusion of multiple handwritten word recognition techniques," *Pattern Recognit. Lett.*, vol. 22, 2001. [Online]. Available: https://doi.org/10.1016/S0167-8655(01)00046-0

[267] M. A. Mohamed and P. Gader, "Generalized hidden markov models. ii. application to handwritten word recognition," *IEEE Trans. Fuzzy Syst.*, vol. 8, 2000. [Online]. Available: https://doi.org/10.1109/91.824774

[268] H. Yamada and Y. Nakano, "Cursive handwritten word recognition using multiple segmentation determined by contour analysis," *IEICE Trans. Inf. Syst.*, vol. 79, 1996.

[269] F. H-C and X. Y-Y, "Multilinguistic handwritten character recognition by bayesian decision-based neural networks," *IEEE Trans. Signal Process*, vol. 46, 1998. [Online]. Available: https://doi.org/10.1109/78.720379

[270] M. Blumenstein, X. Y. Liu, and B. Verma, "A modified direction feature for cursive character recognition," in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, vol. 4. IEEE, 2004, pp. 2983–2987.

[271] R. A. Wilkinson, J. Geist, S. Janet, P. J. Grother, C. J. Burges, R. Creecy, B. Hammond, J. J. Hull, N. Larsen, T. P. Vogl *et al.*, *The first census optical character recognition system conference*. US Depart-

ment of Commerce, National Institute of Standards and Technology, 1992, vol. 184.

[272] M. Shi, Y. Fujisawa, T. Wakabayashi, and F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image," *Pattern Recognit.*, vol. 35, 2002. [Online]. Available: https://doi.org/10.1016/S0031-3203(01)00203-5

[273] L. C-L, H. Sako, and H. Fujisawa, "Effects of classifier structures and training regimes on integrated segmentation and recognition of handwritten numeral strings," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, 2004. [Online]. Available: https://doi.org/10.1109/TPAMI.2004.104

[274] A. K. Jain and D. Zongker, "Representation and recognition of handwritten digits using deformable templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, 1997. [Online]. Available: https://doi.org/10.1109/34.643899

[275] S.-W. Lee and E.-J. Lee, "Integrated segmentation and recognition of connected handwritten characters with recurrent neural network," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1. IEEE, 1995, pp. 413–416.

[276] S. J. Smith, M. O. Bourgoin, K. Sims, and H. L. Voorhees, "Handwritten character classification using nearest neighbor in large databases," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, 1994. [Online]. Available: https://doi.org/10.1109/34.310689

[277] Y. LeCun, "The mnist database of handwritten digits," *http://yann. lecun. com/exdb/mnist/*, 1998.

[278] S. Benzoubeir, A. Hmamed, and H. Qjidaa, "Hypergeometric laguerre moment for handwritten digit recognition," in *2009 International Conference on Multimedia Computing and Systems*. IEEE, 2009, pp. 449–453.

[279] C.-. L. Liu, K. Nakashima, H. Sako, and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," *Pattern Recognit.*, vol. 36, 2003. [Online]. Available: https://doi.org/10.1016/S0031-3203(03)00085-2

[280] L. Z, C. Z, and S. W-C, "Extraction and optimization of b-spline pbd templates for recognition of connected handwritten digit strings," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, 2002. [Online]. Available: https://doi.org/10.1109/34.982890

[281] Z. Wang, Y. Huang, S. Luo, and L. Wang, "A biologically inspired system for fast handwritten digit recognition," in *2011 18th IEEE International Conference on Image Processing*. IEEE, 2011, pp. 1749–1752.

[282] F. Lauer, C. Y. , and G. Bloch, "A trainable feature extractor for handwritten digit recognition," *Pattern Recognit.*, vol. 40, 2007. [Online]. Available: https://doi.org/10.1016/j.patcog.2006.10.011

[283] E. Kussul and T. Baidyk, "Improved method of handwritten digit recognition tested on mnist database," *Image Vision Comput.*, vol. 22, 2004. [Online]. Available: https://doi.org/10.1016/j.imavis.2004.03.008

[284] L. Schomaker, L. Vuurpijl, and L. Schomaker, "Forensic writer identification: A benchmark data set and a comparison of two systems," 2000.

[285] U.-V. Marti and H. Bunke, "A full english sentence database for off-line handwriting recognition," in *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)*. IEEE, 1999, pp. 705–708.

[286] U. V. Marti and H. Bunke, "The iam-database: An english sentence database for offline handwriting recognition," *Int. J. Doc. Anal. Recognit*, vol. 5, 2002. [Online]. Available: https://doi.org/10.1007/s100320200071

[287] S. Wshah, G. Kumar, and V. Govindaraju, "Script independent word spotting in offline handwritten documents based on hidden markov models," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 14–19.

[288] ——, "Multilingual word spotting in offline handwritten documents," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 2012, pp. 310–313.

[289] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "Lexicon-free handwritten word spotting using character hmms," *Pattern Recognit. Lett.*, vol. 33, 2012. [Online]. Available: https://doi.org/10.1016/j.patrec.2011.09.009

[290] A. Bensefia, T. Paquet, and L. Heutte, "A writer identification and verification system," *Pattern Recognit. Lett.*, vol. 26, 2005. [Online]. Available: https://doi.org/10.1016/j.patrec.2005.03.024

[291] I. Siddiqi and N. Vincent, "Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features," *Pattern Recognit.*, vol. 43, 2010. [Online]. Available: https://doi.org/10.1016/j.patcog.2010.05.019

[292] Z. A. Daniels and H. S. Baird, "Discriminating features for writer identification," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 1385–1389.

[293] R. Jain and D. Doermann, "Offline writer identification using k-adjacent segments," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 769–773.

[294] D. Salvi, J. Zhou, J. Waggoner, and S. Wang, "Handwritten text segmentation using average longest path algorithm," in *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. IEEE, 2013, pp. 505–512.

[295] R. P. dos Santos, G. S. Clemente, T. I. Ren, and G. D. Cavalcanti, "Text line segmentation based on morphology and histogram projection," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 651–655.

[296] M. Zimmermann and H. Bunke, "Automatic segmentation of the iam off-line database for handwritten english text," in *2002 International Conference on Pattern Recognition*, vol. 4. IEEE, 2002, pp. 35–39.

[297] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Hybrid off-line cursive handwriting word recognition," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 2. IEEE, 2006, pp. 998–1002.

[298] M. Liwicki and H. Bunke, "Iam-ondb-an on-line english sentence database acquired from handwritten text on a whiteboard," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. IEEE, 2005, pp. 956–961.

[299] A. Schlapbach, M. Liwicki, and H. Bunke, "A writer identification system for on-line whiteboard data," *Pattern Recognit.*, vol. 41, 2008. [Online]. Available: https://doi.org/10.1016/j.patcog.2008.01.006

[300] M. Liwicki, A. Schlapbach, and H. Bunke, "Automatic gender detection using on-line and off-line information," *Pattern Anal. Appl.*, vol. 14, 2011. [Online]. Available: https://doi.org/10.1007/s10044-010-0178-6

[301] E. Indermühle, M. Liwicki, and H. Bunke, "Iamondo-database: an online handwritten document database with non-uniform contents," in *Proceedings of the 9th IAPR international workshop on document analysis systems*, 2010, pp. 97–104.

[302] E. Indermühle, V. Frinken, and H. Bunke, "Mode detection in online handwritten documents using blstm neural networks," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 302–307.

[303] E. Indermuhle, V. Frinken, A. Fischer, and H. Bunke, "Keyword spotting in online handwritten documents containing text and non-text using blstm neural networks," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 73–77.

[304] F. Ott, M. Wehbi, T. Hamann, J. Barth, B. Eskofier, and C. Mutschler, "The onhw dataset: Online handwriting recognition from imu-enhanced ballpoint pens with machine learning," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–20, 2020.

[305] H. Azimi, S. Chang, J. Gold, and K. Karabina, "Improving accuracy and explainability of online handwriting recognition," *arXiv preprint arXiv:2209.09102*, 2022.

[306] M. Bronkhorst, "A pen is all you need: online handwriting recognition using transformers," B.S. thesis, University of Twente, 2021.

[307] A. Shivram, C. Ramaiah, S. Setlur, and V. Govindaraju, "Ibm_ub_1: A dual mode unconstrained english handwriting dataset," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 13–17.

[308] C. Viard-Gaudin, P. M. Lallican, S. Knerr, and P. Binter, "The ireste on/off (ironoff) dual handwriting database," in *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)*. IEEE, 1999, pp. 455–458.

[309] Y. H. Tay, P.-M. Lallican, M. Khalid, C. Viard-Gaudin, and S. Knerr, "Offline handwritten word recognition using a hybrid neural network and hidden markov model," in *Proceedings of the Sixth International Symposium on Signal Processing and its Applications (Cat. No. 01EX467)*, vol. 2. IEEE, 2001, pp. 382–385.

[310] M. Dinesh and M. K. Sridhar, "A feature based on encoding the relative position of a point in the character for online handwritten character recognition," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2. IEEE, 2007, pp. 1014–1017.

[311] C. O. Freitas, L. S. Oliveira, F. Bortolozzi, and S. B. Aires, "Handwritten character recognition using nonsymmetrical perceptual zoning," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 21, 2007. [Online]. Available: https://doi.org/10.1142/S021800140700534X

[312] E. Poisson, C. V. Gaudin, and P.-M. Lallican, "Multi-modular architecture based on convolutional neural networks for online handwritten character recognition," in *Proceedings of the 9th International Con-*

ference on Neural Information Processing, 2002. ICONIP'02., vol. 5. IEEE, 2002, pp. 2444–2448.

[313] E. Augustin, M. Carré, E. Grosicki, J.-M. Brodin, E. Geoffrois, and F. Prêteux, "Rimes evaluation campaign for handwritten mail processing," in *International Workshop on Frontiers in Handwriting Recognition (IWFHR'06),*, 2006, pp. 231–235.

[314] ——, "Rimes evaluation campaign for handwritten mail processing," in *International Workshop on Frontiers in Handwriting Recognition (IWFHR'06),*, 2006, pp. 231–235.

[315] F. Montreuil, E. Grosicki, L. Heutte, and S. Nicolas, "Unconstrained handwritten document layout extraction using 2d conditional random fields," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 853–857.

[316] F. Montreuil, S. Nicolas, E. Grosicki, and L. Heutte, "A new hierarchical handwritten document layout extraction based on conditional random field modeling," in *2010 12th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2010, pp. 31–36.

[317] C. Kermorvant and J. Louradour, "Handwritten mail classification experiments with the rimes database," in *2010 12th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2010, pp. 241–246.

[318] E. Grosicki and H. El Abed, "Icdar 2009 handwriting recognition competition," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 1398–1402.

[319] U. Garain and T. Paquet, "Off-line multi-script writer identification using ar coefficients," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 991–995.

[320] N. Serrano, F. Castro, and A. Juan, "The rodrigo database." in *LREC*, 2010, pp. 19–21.

[321] N. Serrano, A. Sanchis, and A. Juan, "Balancing error and supervision effort in interactive-predictive handwriting recognition," in *Proceedings of the 15th international conference on Intelligent user interfaces*, 2010, pp. 373–376.

[322] V. Romero, N. Serrano, A. H. Toselli, J.-A. Sánchez, and E. Vidal, "Handwritten text recognition for historical documents," in *Proceedings of the Workshop on Language Technologies for Digital Humanities and Cultural Heritage*, 2011, pp. 90–96.

[323] N. Serrano, A. Gimenez, J. Civera, A. Sanchis, and A. Juan, "Interactive handwriting recognition with limited user effort," *Int. J. Doc. Anal. Recognit.*, vol. 17, 2014. [Online]. Available: https://doi.org/10.1007/s10032-013-0204-5

[324] R. I. Elanwar, M. Rashwan, and S. Mashali, "On-line arabic handwriting text line detection using dynamic programming," in *Int. Conf. Comput. Math. Nat. Comput*, 2011, pp. 588–593.

[325] Y. Hamdi, H. Boubaker, and A. M. Alimi, "Online arabic handwriting recognition using graphemes segmentation and deep learning recurrent neural networks," in *Enabling Machine Learning Applications in Data Science*. Springer, 2021, pp. 281–297.

[326] N. Kharma, M. Ahmed, and R. Ward, "A new comprehensive database of handwritten arabic words, numbers, and signatures used for ocr testing," in *Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No. 99TH8411)*, vol. 2. IEEE, 1999, pp. 766–768.

[327] M. Pechwitz, S. S. Maddouri, V. Märgner, N. Ellouze, H. Amiri *et al.*, "Ifn/enit-database of handwritten arabic words," in *Proc. of CIFED*, vol. 2. Citeseer, 2002, pp. 127–136.

[328] H. M. Eraqi and S. Abdelazeem, "A new efficient graphemes segmentation technique for offline arabic handwriting," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 95–100.

[329] S. S. Maddouri, F. B. Samoud, K. Bouriel, N. Ellouze, and H. El Abed, "Baseline extraction: comparison of six methods on ifn/enit database," in *The 11th International Conference on Frontiers in Handwriting Recognition*. Citeseer, 2008.

[330] F. Farooq, V. Govindaraju, and M. Perrone, "Pre-processing methods for handwritten arabic documents," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. IEEE, 2005, pp. 267–271.

[331] M. Bulacu, L. Schomaker, and A. Brink, "Text-independent writer identification and verification on offline arabic handwriting," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2. IEEE, 2007, pp. 769–773.

[332] D. Chawki and S.-M. Labiba, "A texture based approach for arabic writer identification and verification," in *2010 International Conference on Machine and Web Intelligence*. IEEE, 2010, pp. 115–120.

[333] S. Impedovo, G. Facchini, and F. M. Mangini, "A new cursive basic word database for bank-check processing systems," in *2012 10th IAPR International Workshop on Document Analysis Systems*. IEEE, 2012, pp. 450–454.

[334] M. Cheriet, Y. Al-Ohali, N. Ayat, and C. Y. Suen, "Arabic cheque processing system: Issues and future trends," *Digital Document Processing: Major Directions and Recent Advances*, pp. 213–234, 2007.

[335] S. Al-Maadeed, D. Elliman, and C. A. Higgins, "A data base for arabic handwritten text recognition research," *Int. Arab J. Inf. Technol.*, vol. 1, 2004.

[336] S. Al-Ma'adeed, D. Elliman, and C. A. Higgins, "A data base for arabic handwritten text recognition research," in *Proceedings eighth international workshop on frontiers in handwriting recognition*. IEEE, 2002, pp. 485–489.

[337] S. Alma'adeed, C. Higgins, and D. Elliman, "Off-line recognition of handwritten arabic words using multiple hidden markov models," in *Research and Development in Intelligent Systems XX: Proceedings of AI2003, the Twenty-third SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*. Springer, 2004, pp. 33–40.

[338] S. Al-Maadeed, "Text-dependent writer identification for arabic handwriting," *Journal of Electrical and Computer Engineering*, vol. 2012, pp. 13–13, 2012.

[339] N. B. Amara, O. Mazhoud, N. Bouzrara, and N. Ellouze, "Arabase:a relational database for arabic ocr systems," *Int. Arab J. Inf. Technol*, vol. 2, 2005.

[340] H. Alamri, J. Sadri, C. Y. Suen, and N. Nobile, "A novel comprehensive database for arabic off-line handwriting recognition," in *Proceedings of 11th international conference on frontiers in handwriting recognition, ICFHR*, vol. 8, 2008, pp. 664–669.

[341] A. T. Sahlol, C. Y. Suen, M. R. Elbasyouni, and A. A. Sallam, "A proposed ocr algorithm for the recognition of handwritten arabic characters," *J. Pattern Recognit. Intell. Syst*, vol. 2, pp. 8–22, 2014.

[342] H. Alamri, C. L. He, and C. Y. Suen, "A new approach for segmentation and recognition of arabic handwritten touching numeral pairs," in *Computer Analysis of Images and Patterns: 13th International Conference, CAIP 2009, Münster, Germany, September 2-4, 2009. Proceedings 13*. Springer, 2009, pp. 165–172.

[343] M. K. A. E. H. El and A. A. M. Alimi, "The on/off (lmca) dual arabic handwriting database," in *11th International conference on frontiers in handwriting recognition (ICFHR)*, 2008.

[344] H. El-Abed, M. Kherallah, V. Margner, and A. M. Alimi, "On-line arabic handwriting recognition competition: Adab database and participating systems," *Int. J. Doc. Anal. Recognit.*, vol. 14, 2011. [Online]. Available: https://doi.org/10.1007/s10032-010-0124-6

[345] S. A. Azeem and H. Ahmed, "Recognition of segmented online arabic handwritten characters of the adab database," in *2011 10th International Conference on Machine Learning and Applications and Workshops*, vol. 1. IEEE, 2011, pp. 204–207.

[346] I. Hosny, S. Abdou, and A. Fahmy, "Using advanced hidden markov models for online arabic handwriting recognition," in *The First Asian Conference on Pattern Recognition*. IEEE, 2011, pp. 565–569.

[347] H. M. Eraqi and S. A. Azeem, "An on-line arabic handwriting recognition system: Based on a new on-line graphemes segmentation technique," in *2011 international conference on document analysis and recognition*. IEEE, 2011, pp. 409–413.

[348] S. A. Mahmoud, I. Ahmad, M. Alshayeb, W. G. Al-Khatib, M. T. Parvez, G. A. Fink, V. Märgner, and H. El Abed, "Khatt: Arabic offline handwritten text database," in *2012 International conference on frontiers in handwriting recognition*. IEEE, 2012, pp. 449–454.

[349] S. A. Mahmoud, I. Ahmad, W. G. Al-Khatib, M. Alshayeb, M. T. Parvez, V. Märgner, and G. A. Fink, "Khatt: An open arabic offline handwritten text database," *Pattern Recognition*, vol. 47, no. 3, pp. 1096–1112, 2014.

[350] S. Al Maadeed, W. Ayouby, A. Hassaine, and J. M. Aljaam, "Quwi: an arabic and english handwriting dataset for offline writer identification," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 746–751.

[351] A. Hassaïne and S. Al Maadeed, "Icfhr 2012 competition on writer identification challenge 2: Arabic scripts," in *2012 International Conference on Frontiers in Handwriting Recognition*. IEEE, 2012, pp. 835–840.

[352] S. Al Maadeed and A. Hassaine, "Automatic prediction of age, gender, and nationality in offline handwriting," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, pp. 1–10, 2014.

[353] I. Siddiqi, C. Djeddi, A. Raza, and L. Souici-Meslati, "Automatic analysis of handwriting for gender classification," *Pattern Analysis and Applications*, vol. 18, pp. 887–899, 2015.

[354] A. Hassaïne, S. Al Maadeed, J. Aljaam, and A. Jaoua, "Icdar 2013 competition on gender prediction from handwriting," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 1417–1421.

[355] A. Mezghani, S. Kanoun, M. Khemakhem, and H. El Abed, "A database for arabic handwritten text image recognition and writer identification," in *2012 international conference on frontiers in handwriting recognition*. IEEE, 2012, pp. 399–402.

[356] F. Slimane and V. Märgner, "A new text-independent gmm writer identification system applied to arabic handwriting," in *2014 14th International conference on frontiers in handwriting recognition*. IEEE, 2014, pp. 708–713.

[357] F. Khan, A. Bouridane, F. Khelifi, R. Almotaeryi, and S. Almaadeed, "Efficient segmentation of sub-words within handwritten arabic words," in *2014 International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, 2014, pp. 684–689.

[358] A. M. Bidgoli and M. Sarhadi, "Iaut/phcn: Islamic azad university of tehran/persian handwritten city names, a very large database of handwritten persian word," *ICFHR*, vol. 11, pp. 192–197, 2008.

[359] F. Nadi, J. Sadri, and A. Foroozandeh, "A novel method for slant correction of persian handwritten digits and words," in *2013 First Iranian Conference on Pattern Recognition and Image Analysis (PRIA)*. IEEE, 2013, pp. 1–7.

[360] R. Ravani, P. Nooralishahi, and A. S. Amani, "A novel approach for persian/arabic intelligent word recognition," in *3rd European Workshop on Visual Information Processing*. IEEE, 2011, pp. 292–297.

[361] R. Ravani and P. Nooralishahi, "Using dynamic time warping for persian handwriting recognition," in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*. The Steering Committee of The World Congress in Computer Science, Computer …, 2011, p. 1.

[362] M. Ziaratban, K. Faez, and F. Bagheri, "Fht: An unconstraint farsi handwritten text database," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 281–285.

[363] P. J. Haghighi, N. Nobile, C. L. He, and C. Y. Suen, "A new large-scale multi-purpose handwritten farsi database," in *Image Analysis and Recognition: 6th International Conference, ICIAR 2009, Halifax, Canada, July 6-8, 2009. Proceedings 6*. Springer, 2009, pp. 278–286.

[364] N. Nobile, C. L. He, M. W. Sagheer, L. Lam, and C. Y. Suen, "Digit/symbol pruning and verification for arabic handwritten digit/symbol spotting," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 648–652.

[365] P. J. Haghighi and C. Y. Suen, "Handwritten farsi word recognition using hidden markov models," *Guide to OCR for Arabic Scripts*, pp. 273–295, 2012.

[366] Z. Imani, A. Ahmadyfard, A. Zohrevand, and M. Alipour, "offline handwritten farsi cursive text recognition using hidden markov models," in *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*. IEEE, 2013, pp. 75–79.

[367] R. Safabaksh, A. R. Ghanbarian, and G. Ghiasi, "Haft: A handwritten farsi text database," in *2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP)*. IEEE, 2013, pp. 89–94.

[368] M. W. Sagheer, C. L. He, N. Nobile, and C. Y. Suen, "A new large urdu database for off-line handwriting recognition," in *Image Analysis and Processing–ICIAP 2009: 15th International Conference Vietri sul Mare, Italy, September 8-11, 2009 Proceedings 15*. Springer, 2009, pp. 538–546.

[369] ——, "Holistic urdu handwritten word recognition using support vector machine," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 1900–1903.

[370] M. W. Sagheer, N. Nobile, C. L. He, and C. Y. Suen, "A novel handwritten urdu word spotting based on connected components analysis," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 2013–2016.

[371] S.-H. Jeong, Y.-S. Nam, and H.-K. Kim, "Non-similar candidate removal method for off-line handwritten korean character recognition," in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* IEEE, 2003, pp. 323–328.

[372] K. Seo, J. Kim, J. Yoon, and K. Chung, "Comparison of feature performance and its application to feature combination in off-line handwritten korean alphabet recognition," *Int. J. Pattern Recognit. Artif. Intell*, vol. 12, 1998. [Online]. Available: https://doi.org/10.1142/S0218001498000178

[373] M. Nakagawa and K. Matsumoto, "Collection of on-line handwritten japanese character pattern databases and their analyses," *Int. J. Doc. Anal. Recognit.*, vol. 7, 2004. [Online]. Available: https://doi.org/10.1007/s10032-004-0125-4

[374] H. Zhang and J. Guo, "Introduction to hcl2000 database," in *Proceedings of Sino-Japan Symposium on Intelligent Information Networks, Beijing*, 2000.

[375] H. Zhang, J. Guo, G. Chen, and C. Li, "Hcl2000-a large-scale handwritten chinese character database for handwritten character recognition," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 286–290.

[376] Z. Zhang, L. Jin, K. Ding, and X. Gao, "Character-sift: a novel feature for offline handwritten chinese character recognition," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 763–767.

[377] D.-H. Wang, C.-L. Liu, J.-L. Yu, and X.-D. Zhou, "Casia-olhwdb1: A database of online handwritten chinese characters," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 1206–1210.

[378] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "Casia online and offline chinese handwriting databases," in *2011 international conference on document analysis and recognition*. IEEE, 2011, pp. 37–41.

[379] D.-. H. Wang, C.-. L. Liu, and X.-. D. Zhou, "An approach for real-time recognition of online chinese handwritten sentences," *Pattern Recognit.*, vol. 45, 2012. [Online]. Available: https://doi.org/10.1016/j.patcog.2012.04.020

[380] Y. Li, L. Jin, X. Zhu, and T. Long, "Scut-couch2008: A comprehensive online unconstrained chinese handwriting dataset," *ICFHR*, vol. 2008, pp. 165–170, 2008.

[381] L. Jin, Y. Gao, G. Liu, Y. Li, and K. Ding, "Scut-couch2009 a comprehensive online unconstrained chinese handwriting database and benchmark evaluation," *Int. J. Doc. Anal. Recognit*, vol. 14, 2011. [Online]. Available: https://doi.org/10.1007/s10032-010-0116-6

[382] G. Liu, L. Jin, K. Ding, and H. Yan, "A new approach for synthesis and recognition of large scale handwritten chinese words," in *2010 12th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2010, pp. 571–575.

[383] D. Tao, L. Liang, L. Jin, and Y. Gao, "Similar handwritten chinese character recognition using discriminative locality alignment manifold learning," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 1012–1016.

[384] S. Huang, L. Jin, and J. Lv, "A novel approach for rotation free online handwritten chinese character recognition," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 1136–1140.

[385] F. Kleber, S. Fiel, M. Diem, and R. Sablatnig, "Cvl-database: An off-line database for writer retrieval, writer identification and word spotting," in *2013 12th international conference on document analysis and recognition*. IEEE, 2013, pp. 560–564.

[386] S. Fiel and R. Sablatnig, "Writer identification and writer retrieval using the fisher vector on visual vocabularies," in *2013 12th International conference on document analysis and recognition*. IEEE, 2013, pp. 545–549.

[387] M. Diem, S. Fiel, A. Garz, M. Keglevic, F. Kleber, and R. Sablatnig, "Icdar 2013 competition on handwritten digit recognition (hdrc 2013)," in *2013 12th international conference on document analysis and recognition*. IEEE, 2013, pp. 1422–1427.

[388] I. Guyon, L. Schomaker, R. Plamondon, M. Liberman, and S. Janet, "Unipen project of on-line data exchange and recognizer benchmarks," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5)*, vol. 2. IEEE, 1994, pp. 29–33.

[389] S. Prasanna, R. Devi, D. Das, S. Ghosh, and K. Naik, "Online stroke and akshara recognition gui in assamese language using hidden markov model," *arXiv preprint arXiv:1407.2390*, 2014.

[390] U. Baruah and S. M. Hazarika, "A dataset of online handwritten assamese characters," *Journal of Information Processing Systems*, vol. 11, no. 3, pp. 325–341, 2015.

[391] M. F. Mridha, A. Q. Ohi, M. A. Ali, M. I. Emon, and M. M. Kabir, "Banglawriting: A multi-purpose offline bangla handwriting dataset," *Data in Brief*, vol. 34, p. 106633, 2021.

[392] M. J. K. Singh, R. Dhir, and R. Rani, "Performance comparison of devanagari handwritten numerals recognition," *Int. J. Comput. Appl*, vol. 22, 2011.

[393] V. J. Dongre and V. H. Mankar, "Development of comprehensive devnagari numeral and character database for offline handwritten character recognition," *Applied Computational Intelligence and Soft Computing*, vol. 2012, pp. 29–29, 2012.

[394] A. Alaei, P. Nagabhushan, and U. Pal, "A benchmark kannada handwritten document dataset and its segmentation," in *2011 International*

Conference on Document Analysis and Recognition. IEEE, 2011, pp. 141–145.

[395] S. Thadchanamoorthy, N. Kodikara, H. Premaretne, U. Pal, and F. Kimura, "Tamil handwritten city name database development and recognition for postal automation," in *2013 12th International Conference on Document Analysis and Recognition*. IEEE, 2013, pp. 793–797.

[396] N. Shaffi and F. Hajamohideen, "uthcd: A new benchmarking for tamil handwritten ocr," *IEEE Access*, vol. 9, pp. 101 469–101 493, 2021.

[397] B. Karthick, H. Singh, and M. Malarvel, "Pre-processing techniques for offline tamil handwritten character recognition," in *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, 2022, pp. 976–981.

[398] N. Shaffi and F. Hajamohideen, "Few-shot learning for tamil handwritten character recognition using deep siamese convolutional neural network," in *International Conference on Applied Intelligence and Informatics*. Springer, 2021, pp. 204–215.

[399] H.-P. Company. (2013) The hpl isolated handwritten tamil character dataset. [Online]. Available: https://lipitk.sourceforge.net/datasets/tamilchardata.htm

[400] M. Agrawal, A. S. Bhaskarabhatla, and S. Madhvanath, "Data collection for handwriting corpus creation in indic scripts," in *International Conference on Speech and Language Technology and Oriental CO-COSDA (ICSLT-COCOSDA 2004), New Delhi, India (November 2004)*. Citeseer, 2004.

[401] B. Nethravathi, C. Archana, K. Shashikiran, A. G. Ramakrishnan, and V. Kumar, "Creation of a huge annotated database for tamil and kannada ohr," in *2010 12th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2010, pp. 415–420.

[402] S. Obaidullah, C. Halder, N. Das, K. Roy *et al.*, "$pwdb_1 3$: A corpus of word-level printed document images from thirteen official indic scripts," in *Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA) 2015*. Springer, 2016, pp. 233–242.

[403] P. Rakshit, S. Chatterjee, C. Halder, S. Sen, S. M. Obaidullah, and K. Roy, "Comparative study on the performance of the state-of-the-art cnn models for handwritten bangla character recognition," *Multimedia Tools and Applications*, pp. 1–22, 2022.

[404] R. Pardeshi, B. Chaudhuri, M. Hangarge, and K. Santosh, "Automatic handwritten indian scripts identification," in *2014 fourteenth international conference on frontiers in handwriting recognition*. IEEE, 2014, pp. 375–380.

[405] D. Nurseitov, K. Bostanbekov, D. Kurmankhojayev, A. Alimova, A. Abdallah, and R. Tolegenov, "Handwritten kazakh and russian (hkr) database for text recognition," *Multimedia Tools and Applications*, pp. 1–23, 2021.

[406] G. A. Daniyar Nurseitov, Kairat Bostanbekov, Maksat Kanatov, Anel Alimova, Abdelrahman Abdallah, "Classification of Handwritten Names of Cities and Handwritten Text Recognition using Various Deep Learning Models," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 5, pp. 934–943, 2020.

[407] B. Tleubayev, Z. Zhexenova, K. Koishybay, and A. Sandygulova, "Cyrillic-mnist: A cyrillic version of the mnist dataset," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 4767–4773.

[408] P. R. Aryan, I. Supriana, and A. Purwarianti, "Development of indonesian handwritten text database for offline character recognition," in *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*. IEEE, 2011, pp. 1–4.

[409] F. Thoma, J. Bayer, Y. Li, and A. Dengel, "A public ground-truth dataset for handwritten circuit diagram images," in *International Conference on Document Analysis and Recognition*. Springer, 2021, pp. 20–27.

[410] D.-H. KIM, Y.-S. Hwang, S.-T. Park, E.-J. Kim, S.-H. Paek, and S.-Y. BANG, "Handwritten korean character image database pe92," *IEICE transactions on information and systems*, vol. 79, no. 7, pp. 943–950, 1996.

[411] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "H-dibco 2010-handwritten document image binarization competition," in *2010 12th International Conference on Frontiers in Handwriting Recognition*. IEEE, 2010, pp. 727–732.

[412] I. Pratikakis, K. Zagoris, G. Barlas, and B. Gatos, "Icfhr2016 handwritten document image binarization contest (h-dibco 2016)," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2016, pp. 619–623.

[413] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "Icfhr2014 competition on handwritten document image binarization (h-dibco 2014)," in *2014 14th International conference on frontiers in handwriting recognition*. IEEE, 2014, pp. 809–813.

[414] A. Antonacopoulos, B. Gatos, and D. Bridson, "Icdar2005 page segmentation competition," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. IEEE, 2005, pp. 75–79.

[415] N. Nikolaou, M. Makridis, B. Gatos, N. Stamatopoulos, and N. Papamarkos, "Segmentation of historical machine-printed documents using adaptive run length smoothing and skeleton segmentation paths," *Image and Vision Computing*, vol. 28, no. 4, pp. 590–604, 2010.

[416] A. El-Sawy, M. Loey, and H. El-Bakry, "Arabic handwritten characters recognition using convolutional neural network," *WSEAS Transactions on Computer Research*, vol. 5, no. 1, pp. 11–19, 2017.

[417] N. Altwaijry and I. Al-Turaiki, "Arabic handwriting recognition system using convolutional neural network," *Neural Computing and Applications*, vol. 33, no. 7, pp. 2249–2261, 2021.

[418] A. AL-Saffar, S. Awang, W. AL-Saiagh, S. Tiun, and A. S. Al-Khaleefa, "Deep learning algorithms for arabic handwriting recognition: A review," *International Journal of Engineering & Technology*, vol. 7, no. 3.20, 2018.

[419] M. Shams, A. Elsonbaty, W. ElSawy *et al.*, "Arabic handwritten character recognition based on convolution neural networks and support vector machine," *arXiv preprint arXiv:2009.13450*, 2020.

[420] M. Elleuch, N. Tagougui, and M. Kherallah, "Arabic handwritten characters recognition using deep belief neural networks," in *2015 IEEE 12th International Multi-Conference on Systems, Signals & Devices (SSD15)*. IEEE, 2015, pp. 1–5.

[421] Y. M. Alginahi, "A survey on arabic character segmentation," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 16, no. 2, pp. 105–126, 2013.

[422] T. Saba, A. Rehman, and M. Elarbi-Boudihir, "Methods and strategies on off-line cursive touched characters segmentation: a directional review," *Artificial Intelligence Review*, vol. 42, no. 4, pp. 1047–1066, 2014.

[423] V. Ruiz-Parrado, R. Heradio, E. Aranda-Escolastico, Á. Sánchez, and J. F. Vélez, "A bibliometric analysis of off-line handwritten document analysis literature (1990-2020)," *Pattern Recognition*, p. 108513, 2021.

[424] B. Thakral and M. Kumar, "Devanagari handwritten text segmentation for overlapping and conjunct characters-a proficient technique," in *Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization*. IEEE, 2014, pp. 1–4.

[425] M. Sonkusare, R. Gupta, and A. Moghe, "A review on character segmentation approach for devanagari script," *Intelligent Systems*, pp. 181–189, 2021.

[426] R. S. Alkhawaldeh, "Arabic (indian) digit handwritten recognition using recurrent transfer deep architecture," *Soft Computing*, vol. 25, no. 4, pp. 3131–3141, 2021.

[427] M. Musa, "Towards building standard datasets for arabic recognition," *International Journal of Engineering and Advanced Research Technology (IJEART)*, vol. 2, no. 2, pp. 16–19, 2016.

[428] C.-L. Liu, F. Yin, D.-H. Wang, and Q.-F. Wang, "Online and offline handwritten chinese character recognition: benchmarking on new databases," *Pattern Recognition*, vol. 46, no. 1, pp. 155–162, 2013.

[429] ——, "Chinese handwriting recognition contest 2010," in *2010 Chinese Conference on Pattern Recognition (CCPR)*. IEEE, 2010, pp. 1–5.

[430] C.-L. Liu, F. Yin, Q.-F. Wang, and D.-H. Wang, "Icdar 2011 chinese handwriting recognition competition," in *2011 International Conference on Document Analysis and Recognition*, 2011, pp. 1464–1469.

[431] F. Yin, Q.-F. Wang, X.-Y. Zhang, and C.-L. Liu, "Icdar 2013 chinese handwriting recognition competition," in *2013 12th international conference on document analysis and recognition*. IEEE, 2013, pp. 1464–1470.

[432] X.-Y. Zhang, Y. Bengio, and C.-L. Liu, "Online and offline handwritten chinese character recognition: A comprehensive study and new benchmark," *Pattern Recognition*, vol. 61, pp. 348–360, 2017.

[433] Z. Zhong, L. Jin, and Z. Xie, "High performance offline handwritten chinese character recognition using googlenet and directional feature maps," in *2015 13th international conference on document analysis and recognition (ICDAR)*. IEEE, 2015, pp. 846–850.

[434] P. Ghadekar, S. Ingole, and D. Sonone, "Handwritten digit and letter recognition using hybrid dwt-dct with knn and svm classifier," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. IEEE, 2018, pp. 1–6.

[435] P. R. Cavalin, A. de Souza Britto Jr, F. Bortolozzi, R. Sabourin, and L. E. S. Oliveira, "An implicit segmentation-based method for recognition of handwritten strings of characters," in *Proceedings of the 2006 ACM symposium on Applied computing*, 2006, pp. 836–840.

[436] Y. Peng and H. Yin, "Markov random field based convolutional neural networks for image classification," in *Intelligent Data Engineering and Automated Learning–IDEAL 2017: 18th International Conference, Guilin, China, October 30–November 1, 2017, Proceedings 18*. Springer, 2017, pp. 387–396.
[437] B. Su and S. Lu, "Accurate recognition of words in scenes without character segmentation using recurrent neural network," *Pattern Recognition*, vol. 63, pp. 397–405, 2017.
[438] Y. H. Tay, P.-M. Lallican, M. Khalid, C. Viard-Gaudin, and S. Kneer, "An offline cursive handwritten word recognition system," in *Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology. TENCON 2001 (Cat. No. 01CH37239)*, vol. 2. IEEE, 2001, pp. 519–524.
[439] K. Roy and U. Pal, "Word-wise hand-written script separation for indian postal automation," in *Tenth international workshop on frontiers in handwriting recognition*. Suvisoft, 2006.
[440] R. Ghosh and P. P. Roy, "Comparison of zone-features for online bengali and devanagari word recognition using hmm," in *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*. IEEE, 2016, pp. 435–440.
[441] B. Dhandra, H. Mallikarjun, R. Hegadi, and V. Malemath, "Word-wise script identification from bilingual documents based on morphological reconstruction," in *2006 1st International Conference on Digital Information Management*. IEEE, 2006, pp. 389–394.
[442] R. S. Zinjore and R. Ramteke, "Identification and removal of devanagari script and extraction of roman words from printed bilingual text document," *International Journal of Computer Applications*, vol. 975, p. 8887, 2015.
[443] T. S. El-Sheikh and S. El-Taweel, "Real-time arabic handwritten character recognition," *Pattern recognition*, vol. 23, no. 12, pp. 1323–1332, 1990.
[444] S. Al-Emami and M. Usher, "On-line recognition of handwritten arabic characters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 704–710, 1990.
[445] K. Assaleh, T. Shanableh, and H. Hajjaj, "Recognition of handwritten arabic alphabet via hand motion tracking," *Journal of the Franklin Institute*, vol. 346, no. 2, pp. 175–189, 2009.
[446] K. Takahashi, H. Yasuda, and T. Matsumoto, "A fast hmm algorithm for on-line handwritten character recognition," in *Proceedings of the fourth international conference on document analysis and recognition*, vol. 1. IEEE, 1997, pp. 369–375.
[447] C.-L. Liu and X.-D. Zhou, "Online japanese character recognition using trajectory-based normalization and direction feature extraction," in *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft, 2006.
[448] G. X. Tan, C. Viard-Gaudin, and A. C. Kot, "Information retrieval model for online handwritten script identification," in *2009 10th International Conference on Document Analysis and Recognition*. IEEE, 2009, pp. 336–340.
[449] G. S. Reddy, P. Sharma, S. Prasanna, C. Mahanta, and L. Sharma, "Combined online and offline assamese handwritten numeral recognizer," in *2012 National Conference on Communications (NCC)*. IEEE, 2012, pp. 1–5.
[450] S. D. Chowdhury, U. Bhattacharya, and S. K. Parui, "Online handwriting recognition using levenshtein distance metric," in *2013 12th international conference on document analysis and recognition*. IEEE, 2013, pp. 79–83.
[451] K. Mehrotra, S. Jetley, A. Deshmukh, and S. Belhe, "Unconstrained handwritten devanagari character recognition using convolutional neural networks," in *Proceedings of the 4th International Workshop on Multilingual OCR*, 2013, pp. 1–5.
[452] N. Bhattacharya, P. P. Roy, and U. Pal, "Sub-stroke-wise relative feature for online indic handwriting recognition," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 18, no. 2, pp. 1–16, 2018.
[453] A. Bharath and S. Madhvanath, "Hidden markov models for online handwritten tamil word recognition," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 1. IEEE, 2007, pp. 506–510.
[454] H. Liu and X. Ding, "Handwritten character recognition using gradient feature and quadratic classifier with multiple discrimination schemes," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*. IEEE, 2005, pp. 19–23.
[455] H. Tang, E. Augustin, C. Y. Suen, O. Baret, and M. Cheriet, "Spiral recognition methodology and its application for recognition of chinese bank checks," in *Ninth International Workshop on Frontiers in Handwriting Recognition*. IEEE, 2004, pp. 263–268.
[456] U. Bhattacharya and B. B. Chaudhuri, "Handwritten numeral databases of indian scripts and multistage recognition of mixed numerals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, 2009. [Online]. Available: https://doi.org/10.1109/TPAMI.2008.88

Shivangi Nigam Shivangi Nigam received her M. Tech in computer sciences from NIT Hamirpur, HP, in 2015. She received her B.Tech degree in Computer Sciences from Priyadarshini College of Computer Science, Greater Noida, UP, in 2012. She is currently a research scholar in IIIT Allahabad, Prayagraj, UP. Her research interests include machine learning, deep learning, computer vision and wireless sensor networks.

Shekhar Verma Shekhar Verma received the Ph.D. degree in computer networks from the Indian Institute of Technology, Varanasi, India, in 1993. He is currently working as a Professor (IT) with the Indian Institute of Information Technology, Allahabad, India. His research interests include machine learning, computer networks, wireless sensor networks, wireless networks, information and network security, vehicular technology, and cryptography.

P Nagabhushan P. Nagabhushan (B.E.-1980, M.Tech–1983, Ph.D.-1989) is the Vice-Chancellor of the Vignan's Foundation for Science, Technology & Research, Guntur, India. He is an active researcher in the theme discipline of Computer Cognition and Recognition with interests in Dimensionality Reduction, Document Image Analysis, Knowledge Mining, Incremental Learning, Symbolic Data Analysis, Image and Object Recognition and related areas. Till now he has successfully supervised 21 Ph.D. candidates. He has over 400 publications in journals and conferences of International repute. He has chaired several international conferences. He is a visiting professor to USA, Japan and France. He is a fellow of Institution of Engineers and Institution of Telecommunication and Electronics Engineers, India.