

Artificial Intelligence and Machine Learning Approaches to Document Digitization in the Banking Industry: An Analysis



Archana Lopes^{1,2}, Kolla Bhanu Prakash^{1*}

¹ Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur 522 502, Andhra Pradesh, India

² Department Electronics and Computer Science, Fr. Conceicao Rodrigues College of Engineering, Bandra 400 050, Mumbai, India

Corresponding Author Email: drkbp@kluniversity.in

<https://doi.org/10.18280/isi.280521>

ABSTRACT

Received: 22 May 2023

Revised: 2 September 2023

Accepted: 11 October 2023

Available online: 31 October 2023

Keywords:

digitization, OCR, automation, extraction, CNN, artificial intelligence, machine learning, banking industry

Technological advancements have led to a significant evolution in the business landscape, particularly within financial processes and the banking industry. Amidst this transformation, the concept of digitization, although frequently referenced in literature, remains ambiguously defined. This study aims to investigate the implications of digitization in the Indian banking sector and explore the various techniques employed in the process. By transitioning to digital platforms, firms anticipate enhancing their competitive advantage, streamlining financial and operational management, and impacting societal structures. This review examines scholarly articles from the past decade, focusing on the influence of digitization on the economy and the technological trends in document digitization. The review also recognizes the emergence of novel technologies such as automation, artificial intelligence, and machine learning, alongside deep learning algorithms, which are driving a new generation of intelligent business operations. These areas merit further exploration in future research endeavors.

1. INTRODUCTION

Over the past few decades, digital technologies have been increasingly integrated into various facets of daily life. This process, known as digitization, is transforming societies, businesses, and global economies. Identified as a pivotal technology trend, digitization exerts a profound influence on both societal and industrial sectors. Consequently, organizations are under persistent pressure to adopt digital tools and adapt their strategies to thrive in the digital age. Although the transition to digital encompasses a multitude of benefits, it also necessitates substantial investments and time commitments.

In the wake of digital technology's evolution [1], a pertinent question arises regarding the utilization of digitization by professionals in the field and the extent to which scientists and researchers are keeping pace with this development. The digitization of workflow often involves the creation of application-specific software, likely based on the current workflow, thereby necessitating the replication of document structures in their digital storage. However, digitization encompasses more than this minor component; businesses may also need to update their software or hardware, expand their workforce or provide training, or even entirely redefine their business strategy. The complexity of digitization arises from the numerous variables involved, making a robust process model an invaluable tool [1]. In the absence of such a process architecture, businesses may produce suboptimal results.

Thus, our primary objective is to delineate the current state of the art and foster an enhanced understanding of the term

"digitization." Although an abundance of literature exists on digital transformation, studies specifically addressing digitization are notably sparse. It is recognized that the term "digital transformation" was initially introduced by corporate leaders before academic investigations ensued. However, there exists a significant knowledge gap at the federal level, contributing a mere 1% to global research [2].

In the 2020s, artificial intelligence (AI) and machine learning (ML) have emerged as increasingly influential technologies in document digitization. This study strives to elucidate the intellectual progression of AI and ML in financial research, employing a research review with an embedded investigation to analyze the services of these concepts.

Emphasizing the importance of investigating the potential exclusionary effects of economic digitization from both social and political perspectives, this study proposes trends in the digitization of the economy [3, 4]. It draws on research on digital inequality to analyze these developments and their potential impact on individuals. This interdisciplinary approach underscores the necessity for further research on this topic and suggests avenues for exploration from both policy and academic perspectives. Additionally, it examines document digitization technologies such as Convolutional Neural Networks, Artificial Intelligence, and Machine Learning.

This review encompasses the effects of digitization on society and the financial sector, as well as the technologies developed to facilitate digitization. The literature review is divided into two parts: one addressing digitization in the banking and financial sectors and its societal effects, and the other discussing various techniques used for digitization. This

review paper is organized as follows: Section 2 provides an overview of the literature on digitization in the banking and financial sectors, as well as its societal and mobility effects. Section 3 presents a survey on different technologies such as Handwritten Recognition systems using machine learning techniques, CNN techniques, and AI and ML techniques for the digitization of various financial and company documents. Section 4 presents the conclusion of the review paper, and Section 5 discusses the limitations of the literature survey.

2. LITERATURE REVIEW ON DIGITIZATION IN BANKING AND FINANCIAL DOMAIN

Digitization, defined as the conversion of data into a digital format via electronic means, has emerged as a critical process in the banking and financial industry. This transformative process has enabled the shift from traditional manual, paper-based operations to digital workflows. Digitization in this sector encompasses the automation of myriad banking operations and the integration of digital technologies, with the goals of enhancing efficiency, convenience, and customer experience.

Various web and mobile-based payment schemes have been developed, which emphasize improving the customer experience and effectively incorporating financial transactions into the corporate value chain. Distinctively, these schemes often do not apply for banking licenses, instead offering financial services as an adjunct to payment provisions.

The Indian government has demonstrated a commitment to facilitating online banking, with the IT Act of 2000, enacted on October 17, 2000, serving as a notable milestone. Technological advancements seem to be progressively eclipsing the traditional banking system in India, as evidenced by several digitization initiatives undertaken by the country's banks. These initiatives include active engagement on social media platforms such as Twitter and Facebook, the introduction of an online tax accounting system, the provision of a cardless withdrawal option by the Reserve Bank of India, and the launch of a Business Transformation Program by the Bank of Baroda.

Digitization techniques encompass a wide range, including real-time settlement, prepaid payment methods, online clearing services, mobile banking, NEFT, and debit/credit cards [5]. These various methods of digitization underscore the increasingly digital landscape of the banking and financial sector and the ongoing efforts to streamline and modernize financial transactions and customer interactions.

2.1 Literature review on digitization in the Indian banking sector

The necessity for digitization within the Indian banking industry was recognized in the late 1980s to bolster record-keeping, accounting, and customer service strategies. Following this realization, a committee led by Dr. C. Rangarajan was established by the Reserve Bank of India in 1988, with the objective of investigating opportunities for bank computerization. The trend of computerization witnessed noteworthy augmentation coinciding with the reform of the Indian economy during 1991-1992. This escalation was primarily attributed to the increasing influx of private and international banks into the banking framework. To maintain a competitive edge, select commercial banks initiated the

provision of digital services. The introduction of services such as MICR-based cheque processing, electronic cash transfers, branch connectivity, and ATMs significantly enhanced the convenience of banking hours for commercial banks in India.

A scarcity of cash during the demonetization era compelled individuals to adapt to mobile banking or digital payments, which emerged as more convenient alternatives. In the current scenario, digital payments are being fervently promoted by the Indian government. The National Payments Corporation of India introduced significant steps in the payment process, including the Unified Payments Interface (UPI) and Bharat Interface for Money. The UPI, a mobile platform, facilitates the transfer of money between bank accounts using a virtual address, eliminating the need for actual bank account numbers.

Shetty et al. [6] investigated the challenges and opportunities associated with the development of India's banks through digitization. The study revealed that digitization enhanced customer service, reduced bank costs and customer waiting times, bolstered competitive advantages, and enabled an increase in revenue from promotional efforts. However, the study also identified risks such as cyber-attacks, site selection for ATMs, and the choice of technology. With the rising popularity of smartphones, it was suggested that the banking industry must digitize to align with global expectations. While digitization was found to reduce human error and enhance convenience, banks were urged to exercise heightened vigilance and be prepared to counter cyber threats, which are becoming increasingly prevalent.

This research aims to understand the initial acceptance of mobile banking among current internet banking users. The inclination of Indian internet banking customers to utilize equivalent services, such as mobile banking, has been the focus of limited studies. A theoretical model was developed to consider the acceptance factors influencing internet banking clients to transition to mobile banking. Factors such as perceived convenience, perceived security, mobile self-efficacy, social impact, and customer support were identified as key drivers of adoption. The dependent variable in this study was a customer's likelihood of adopting mobile banking services. The study utilized a partial least squares structural equation modeling approach to analyze data from 420 internet banking clients from public, commercial, foreign, and cooperative banks in India. The study concluded that adoption factors exert a significant impact on customers' social intention to adopt mobile banking, as evidenced by the survey.

The results of this study are expected to provide insights into digital banking channels, increase awareness of digital banking usage, and inform financial institutions about mobile banking usage in India.

2.2 Literature review on the impact of digitization on society and mobility

Defined as "the adoption or increased utilization of digital or computer technology in an organization, industry, or country", digitization represents the transformation of a physical enterprise into a digital one. This process, encompassing the deployment of digital technologies to enhance operations, diminish costs, augment productivity (e.g., operations and maintenance), and generate new business models, opens up new avenues for income and value creation [7]. The integration of digital technology into daily life equates to the digitization of all possible elements. Illustrative manifestations of digitization within the general public include

the revolution of smartphones, the exponential surge in social media usage, and the transition from physical services and infrastructure to online banking, e-government, and e-health services.

Cecchini posits that digitization, specifically Information and Communication Technology (ICT), can potentially alleviate poverty by enhancing access to government, financial, and health services for the underprivileged [8]. The Digital Britain project final report underscores the escalating importance of access to digital infrastructure. The significance of accessibility to advanced devices and understanding their usage becomes evident considering the frequency with which digital technologies are employed to augment the effectiveness and efficiency of major public services such as education, healthcare, and transportation. Without computer and digital media literacy, digital connectivity, or the necessary resources to utilize digital technology, a broad array of contemporary services, commodities, and activities become directly inaccessible, leading to societal exclusion. The uneven global distribution of digital technology's benefits and the varying degrees of users' digital literacy "may have significant implications for economic growth, human development, and wealth creation." Consequently, the ensuing "digital divide" has emerged as a significant concern on both the national and international scale.

Table 1. Summary of literature review on effect of digitization on society and mobility

Ref	Keywords	Summary
[9]	Digital transport, Digitization	This paper demonstrates the value of studying the exclusionary consequences of transportation digitization from both an academic and social policy standpoint.
[10]	Government, information security, costs, switches, television, educational technology, joining procedures, communications technology, Internet telephony	This study confirms the importance of these middlemen in facilitating access to material that meets the most pressing information needs of the community, fostering local ownership, and encouraging participation from the economically disadvantaged.
[11]	Phone use by cyclists, preventative measures, cyclist safety, and traffic laws	Digitization in transportation and travel is already taking place in ways that have altered how people travel.
[12]	Governance, sustainability, smart mobility, and Mobility as a Service	This study offers a thoughtful assessment of Mobility as a Service (MaaS), a recent innovation vying for a place within the Smart Mobility framework. We evaluate a variety of its potential effects on governance and sustainability for urban policymakers in the future (i.e., social and environmental impacts).

With the proliferation of easily accessible customer data, mobility has undergone a substantial transformation. According to John Henson, CEO of Henson Digital, digitization has effectively placed the consumer at the epicenter of advancements, with consumer behavior and technological development serving as key catalysts for

changes leading to the launch of new mobility services. The ubiquitous availability of the internet and connected mobile devices has made travellers increasingly reliant on digital tools and necessitated the skills to navigate the digital world. Digital transformation has introduced a novel perspective to traffic decision-making processes, simplifying data collection and analysis procedures, enhancing accuracy and real-time capabilities, while posing both benefits and challenges to decision-makers, mobility companies, and consumers [13].

Furthermore, digitization enables mobility providers to diversify their income streams through digital advertising and on-board connectivity infrastructure, while simultaneously enhancing customer relationships and reducing costs. Digitization, a paramount social development, influences not only individuals' daily lives but also the future trajectory of enterprises. The effect of digitization on society and mobility is summarized in Table 1.

As Section 2 illustrates, digitization enables the use of mobile devices and computers to conduct financial activities such as payments, banking transactions, and management of investments. This digital revolution has rendered banking and financial services more convenient for customers and widened their accessibility. The use of mobile banking apps for payments from smartphones has reduced the dependency on banks' physical infrastructure. In summary, digitization within the banking and financial sectors has had a transformative impact on society and mobility, making financial services more convenient, accessible, and secure, promoting financial inclusion, and fueling economic growth.

3. LITERATURE REVIEW ON VARIOUS TECHNOLOGIES FOR DIGITIZATION OF DOCUMENTS

The traditional approach in document digitization application is shown in Figure 1. General flow of the process includes acquiring scanned image followed by preprocessing it by different image pre processing techniques. Then features are extracted followed by feature selection step. The selected features are then used for classification of documents. Once the documents are classified, they can be further processed to extract relevant text data.

In this section we present the detail literature review of various technologies like Machine learning and Artificial Intelligence used for Digitization of documents of financial sectors, policies and academics perspective. Table 2 shows the details information of various research work for digitization of documents by different techniques.

We have identified literature evaluations that are pertinent to the application of AI and ML in the policy, financial, and educational sectors of digital transformation, although they vary in terms of methodology and methodologies applied [14]. Table 3 shows a comparison of different OCR engines that are essential for the application of AI and ML in document digitization. Table 4 gives a comparison of different datasets available online.

Moubayed et al. [15] examined the e-learning field's definitions and characteristics and provided a quick overview of the most widely used machine learning and data analytics tools.

Another literature review research that aimed to introduce readers to educational data mining was conducted by Romero and Ventura [16]. Prediction, cluster analysis, outlier

identification, connection mining, social network analysis, process mining, text mining, data reduction for human evaluation, model-based discovery, knowledge tracing, and non-negative matrix factorization were all used in this study. Outside of the IJAIED context, the research did not conform to systematic literature review standards, focus on ML algorithms, or even consider any of those things.

A reflection study on AI by Guan et al. [17] pointed out that

analytics and profiling have lately gained greater traction when examining the themes and their evolution. The article gives a brief introduction to the ways in which deep learning and AI have been used to improve educational practices. However, the paper fails to sufficiently highlight the importance of ML algorithms in online education. Table 2 summarises various technologies for digitization of documents.

Table 2. Summary of review on various technologies for digitization of documents

Ref.	Keywords	Summary
[18]	Convolutional Neural Network, Genetic Algorithm, Sparse Auto encoder	The results of this study show that evolutionary methods can be used to train CNN more effectively
[19]	Numeric Recognition Optical Character Recognition Hybrid Feature Set	It can be concluded that better feature vectors, a hybrid feature set, and training with a larger dataset containing all possible ways of writing a number can improve the accuracy of digit recognition using a neural network
[20]	Pattern recognition, handwriting recognition, number recognition, neural networks, classification algorithms, machine learning, off-line handwriting recognition, and WEKA	This study is a first effort to help in the identification of handwritten numbers without resorting to conventional classification techniques
[21]	CNN; Deep learning; ECOC; OCR; SVM	In order to recognize handwritten characters, this article introduces CNN-ECOC, a Convolutional Neural Network and Embedded Conditional Classifier combo
[22]	TensorFlow (TF), Handwritten Text Recognition (HTR)	In this study, personalities are categorized. A conventional neural network was used to execute this job. Our final precision was more than 90.3%
[23]	Using a consensus-based clustering method, to recognize handwritten digits from the MNIST dataset	The implementation of consensus clustering is presented in this work using the MNIST dataset as a case study. These implementations are done in Python with the help of the Jupyter Notebook and the Anaconda and Consensus Clustering packages
[24]	Features extracted, patterns recognized, matching patterns in a certain direction, and recognition accuracy	The results show that the directed feature extraction algorithms work better than the traditional ones

Table 3. Summary of review on different OCR engines

Ref.	Summary	Advantages	Limitations
[25]	Comparison of techniques for text extraction using OCR	CNN, and LSTM algorithms when used in OCR give better accuracy	For complex dialects, OCR does not give 100% accuracy
[26]	Comparison of Foxit, Tesseract PDF2GO over 8,562 government documents with six document categories	Tesseract is the most suitable solution as it gives the highest F1 score for documents without any table data	Documents involving watermarks and handwritten scanned documents are not used
[27]	Comparison of Google Tesseract with Microsoft OCR with RPA	Microsoft OCR results are more accurate than Google Tesseract, but Tesseract takes very little time compared to Microsoft OCR	The dataset used for comparison consists of only 100 documents
[28]	Comparison of OCR engines such as Google Drive OCR, ABBYY FineReader, GOCR, Tesseract, OCRAD, and Cuneiform for extracting source code from video screenshots	The time required to process video by Google Drive and ABBYY FineReader is 6.02s and 2.51s whereas Tesseract needs 1.56s. The time required to process screenshots by Google drive is 5.69s and by ABBYY is 2.54s and Tesseract needs 4.60s	The bounding boxes around the code in the screenshots were manually done

Table 4. Comparison of datasets

Dataset	Types of Images	Size of the Dataset	Classes
RVL-CDIP	Letters, Memo, Emails, Invoice	400000	16
MNIST	Digits	70000	10
EMNIST	Handwritten Characters	131600	47
Bank Search	Loan documents	11000	11
Turkish Bank	Money transfer request	47846	5
NIBSS	E-payment channels in the form of tables and graph	12942000 Entries	8
FUNSD	Scanned forms	199	4
NIST-SPDB6	Handwritten tax form images	5596	20
Tobacco-3482	The report, memo, form, and email	3482	10
SROIE	Scanned Receipts	1000	-

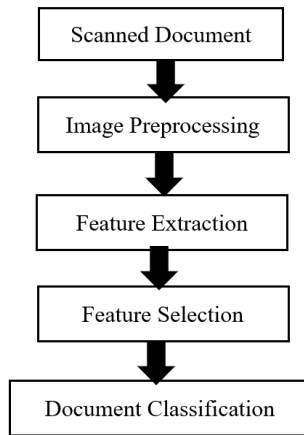


Figure 1. Traditional approach for document digitization

3.1 Review on OCR engines used in document digitisation

To understand how OCR might be enhanced, it is necessary to first grasp its limits. Readability is affected by the following factors:

1) Handwriting or other marks that cover the text hamper its correct recognition.

Furthermore, many components in the file, such as lines and boxes, confuse OCR because it attempts to interpret the lines as part of the text. According to Legal Scans, in this situation, the success rate of OCR lowers swiftly and equals just 60%-80% successful read.

2) The quality and condition of a paper.

The first issue can be addressed by AI Machine Learning, making OCR more advanced [29].

OCR mainly has three units: text detection, text recognition, and text labeling. In the industry many OCR tools are available. There are open-source OCR tools like Tesseract, and EasyOCR and commercial tools like ABBY, and Google vision. Open-source tools like Tesseract and EasyOCR have inbuilt text detection, or we can use algorithms like EAST and CTPN for detection using pre-trained models. As the output of the text detection step, a bounding box is created around the portion of the image having text. After text detection, we use machine learning to match those identified images with a particular text in the text recognition step. Tools like Tesseract, Calamari, and CRNN can be used for this purpose. Tesseract is the most popular since it is open source and provides decent accuracy for the document. The last step of OCR is text labeling. During this step, useful information recognized from the particular image is tagged with a particular label. This step aims to develop the linkage between the desired content and its usage. Regular expressions or NER model-based spacy tools are often used for this purpose. Once the relevant data is extracted using the NER methods, we use key-value pair and reserve it in JSON or XML format or we will directly store the relevant extracted value in the database like a particular key. Tesseract is an open-source OCR solution maintained by the community supported by Google. It returns results in plain text or PDF, with text overlaid on the original image, and provides decent accuracy. Tesseract has a neural net-based OCR engine mode that gives improved accuracy even for image documents with high noise. But with Tesseract, a lot of preprocessing is needed specifically to document noise type. ABBY comes with a GUI to support OCR extraction. It comes with ABBY cloud for cloud services. ABBY also has a desktop app that starts at approximately \$200 and provides

access to a software development kit. ABBY returns JSON, XML, or a pdf with a text searchable inline form. But ABBY has errors due to missing character detection and has low accuracy for handwritten text.

The Google vision OCR tool can be used via API. But its setup is quite challenging. It gives JSON as an output, which contains information about character position, and it works well with handwritten text. This tool is not a good choice when working with tabular data as it misses single digits. It also does not support document sizes greater than 10MB. It is a commercial engine that gives free access to the first 1000 pages.

Microsoft OCR provides high accuracy in text extraction, even from noisy image documents. It has an API endpoint to send the image document and return it with JSON output, which contains coordinate information along with text extracted. Microsoft OCR has errors due to missing character detection. PDF2GO is a multiplatform online tool that works with file formats like pdf, Microsoft word, txt, ePub, OpenOffice, etc. But the drawback is the limited free version, which is ad-laden. There is no option for an e-signature and watermark.

Foxit Reader is a multilingual PDF tool that allows you to create, view, edit, digitally sign and print PDF files. However, this engine contains potentially unwanted programs that behave like malware, such as Ask Toolbar and OpenCandy.

From the literature different OCRs like Google cloud vision, Microsoft Azure, ABBYY FineReader, Google Tesseract and ABBYY Flexicapture were studied and compared on different aspects. As Google cloud Vision and Microsoft Azure, uses pre trained models, they have higher image classification accuracy. ABBYY FineReader has market leading OCR capabilities, excelling at digitizing documents and extracting text from images and pdfs with high accuracy. However it lacks more advanced computer vision functionality. Google Tesseract and ABBYY FlexiCapture are outperformed by ABBYY FineReader. Tesseract is open source and focuses more on structured data extraction.

Google cloud vision delivers the highest overall accuracy above 98%. ABBYY FineReader matches their accuracy on clean documents, but lags on handwriting and poor quality scans. Tesseract and Microsoft Azure has accuracy of 90-95%.

3.2 Printed and handwritten text recognition systems by machine learning techniques

To digitize documents, particularly drawings and plans, geometrical forms predominate, making text recognition a significantly more difficult process. Because a standard OCR system cannot recognize the location of text in such images, the text is usually ignored. The objects are first identified using an AI and ML-based. The text is then explicitly searched for in frames. This workflow protects the system from overlooking critical information.

Many financial documents have handwritten text. As discussed in previous section on review of OCR engines, all OCR engines provide limited support on handwritten text detection and recognition. Therefore, it is necessary to employ techniques for handwritten text recognition. Handwriting recognition systems are models used to collect and analyze data in the form of handwritten text from sources such as paper documents and photographs. The success of handwriting recognition systems has largely depended on optical text recognition, which is responsible for segmenting handwritten

digits, and character recognition is the core of that module [21]. Number and character recognition is used to match handwritten numbers and characters with their associated computer records. This is something that any learning paradigm will do.

Using machine learning algorithms, Shamim et al. [20] suggested Handwritten Digit Recognition. Naive Bayes, Multilayer Perceptron, J48, Random Forest, Bayes Net, Support Vector Machine, and Random Tree were among the methods used. The approach detailed in this paper employs many machine learning methods to identify printed numbers without an Internet connection. This paper's main objective is to make sure that techniques for reading handwritten numbers are accurate and effective. Bayes Net, Support Vector Machine, Random Forest, Random Tree, Multilayer Perceptron, J48, and Naive Bayes are used for identification of digits. The Multilayer Perceptron used in this study increased accuracy to a maximum of 90.37%. More accuracy needs to be achieved.

CNN-ECOC was suggested by Bora et al. [21] for Handwritten Character Recognition from Pictures. This study uses a CNN and an Error Correcting Output Code (ECOC) classifier to present OCR. The ECOC classifies the features once they have been isolated using the CNN. The NIST manually written character picture dataset was utilized for preparing and approving the CNN-ECOC. According to this article, CNN-ECOC provides greater precision than the regular CNN classifier. The fact that the ECOC classifier places a higher computing burden on the system and requires more time to train is one of its drawbacks.

Hybrid evolutionary technique for Convolutional Neural Network-based Devanagari handwritten numeral identification was proposed by Trivedi et al. [18]. They employed the Sparse Autoencoder, CNN, Softmax Classifier, and Genetic Algorithm to create the model. A hybrid deep learning method utilizing the L-BFGS system and the Genetic Algorithm has been created for CNN training. The model has been tested using the Devanagari handwritten numeral dataset. This study's conclusion is that CNN training should be improved by utilizing evolutionary approaches. There will be some issues with chromosomal length as the number of iterations rises. Not all layers of the system are applied with genetic algorithms.

Manchala et al. [22] presented deep learning with TensorFlow for handwritten text recognition. Connectionist temporal classification (CTC), Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN) are the strategies used in this paper to train recurrent neural networks. TensorFlow is used to implement the model. This approach achieves an accuracy of 90.3%. Only the text in this project that has the least amount of noise is given the maximum level of accuracy. The dataset determines the correctness in its entirety. The model's accuracy increases as the amount of data increases. It does not provide the highest level of precision for cursive letters.

Choudhary et al. [19] proposed an offline handwritten character recognition using features extracted from binarization technique. The main purpose of this study is to remove features learned by binarization to recognize handwritten English characters. A multi-layer feedforward neural network was used as a classifier to recognize handwritten character images. Before the characters are labelled, the images are preprocessed by doing things like thinning, reducing foreground and background noise, cropping,

and normalizing size. However, compared to previous research, the precision attained by this study is relatively low.

Hallale and Salunke [24] presented Twelve Directional Feature Extraction for Handwritten English Character Recognition. This paper employs twelve directional properties to distinguish handwritten English alphabets and digits. The features of similarity measurements are studied using directed pattern matching. The detection rates of conventional and twelve directional feature extraction approaches are then compared. It is clear from the results of explore that coordinated capability extraction calculations beat more customary strategies. However, the project's precision is just 88.29%. Accuracy must be enhanced further.

3.3 Document structure recognition

3.3.1 Importance

The analysis of titles, headers, sections, and thematically cohesive pieces is the aim of document logical structure recognition. There is yet another significant benefit that an AI and ML-based OCR system offers over a traditional one, which is significant because:

- It is essential for extracting useful information;
- It permits automated indexing and storage, minimizing the need for extra information retrieval;
- It makes it possible to link related documents (for example, those describing the same object).
- Document structure recognition improves file handling flexibility. If the structure is known, you may simply locate the document as well as others that are related to it.

3.3.2 Realization

Recognizing the document structure may be done in two ways:

(1) Layout analysis

Entry forms, invoices, and other papers are all generated using the same structure and standards and have a comparable format. This frequently provides numerous hints regarding the relationships between various structural units, including headings, body text, tables, references, figures, etc.

(2) Analyzing the content itself

Use keywords to identify the relationships and meanings of text.

Why Using Algorithms is not an option

To help the programme detect certain sections of a document, several software service providers offer to include bespoke algorithms.

However, this strategy isn't flexible. In some projects, it may be necessary to manage a wide range of distinct document layouts. Also, with time, even for the same document type, layouts alter and the distinction between versions becomes increasingly obvious. Thus, using algorithms is ineffective, and a more adaptable tool is needed [30].

It takes a lot of effort, money, and time to refine algorithms so they can adapt to freshly designed document layouts. With its capacity for self-improvement, machine learning is able to adapt to changes swiftly. Document structure identification may be seen as an object detection problem that AI and ML can tackle.

3.4 Deep learning algorithms for feature extraction and document classification

Accuracy of traditional methods depend on the image

preprocessing. Deep learning methods depends on basic Convolutional neural networks. For feature extraction, deep learning methods like VGG-16 and ResNet has proved to be very effective [31].

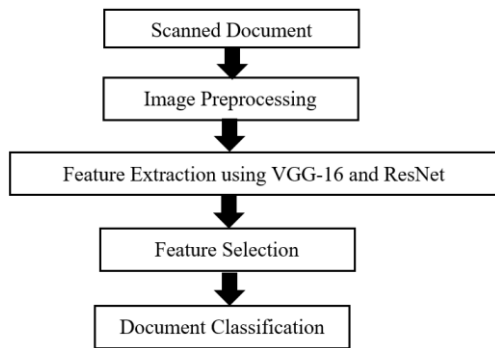


Figure 2. Deep neural network-based approach

VGG-16 (Visual Geometry Group): Input to this network can be color image. This image is passed through stack of convolutional layers. Every convolutional layer in this stack uses a kernel which is very small receptive field and it is of the size 3×3 with stride 1. This network has thirteen convolutional layers, five max pool layers, three fully connected layers. The activation at the last layer is softmax. One thousand channels at the last layer indicates one channel per class. As the depth of the deep neural net goes on increasing, the number of parameters to be trained also increases.

ResNet (Residual Network): This network addresses the problem of vanishing gradient. The core idea in ResNet is, it introduces skip connection or identity shortcut. It skips one or more layers and feeds the input indirectly by passing one or more intermediate layers. This bypass path is an additional path along with regular path. Figure 2 shows the Deep neural net-based approach.

From above discussion, we can conclude that the OCR accuracy on modern documents is very high upto 99% and it drops significantly for older, degraded documents. For improving accuracy of such documents more computational power is needed. OCR processing time mainly depends on image resolution, color mode, document condition and OCR engine design. Grayscale images and cleaner images process faster compared to colored images. Neural networks in OCR engines are very slow but accurate. Post processing like layout analysis and text formatting adds to computational burden. Cloud based OCR services can scale computationally but have latency trade offs [32].

3.5 Comparison of different datasets used in document digitization

The main activities in the bank are cheque clearance, deposits and withdrawals, loan approvals and death claim clearance, and insurance clearance. For automating cheque clearance, different features on the cheque need to be extracted. These features include extraction of bank logo, date, account number, IFSC code, amount, and most importantly verification of signatures on the cheque.

The success of any machine learning algorithm depends on the selection of standard datasets that are properly benchmarked and evaluated. The complete automation of the banking domain has many aspects like automated cheque clearance, data extraction from KYC documents, loan

applications, account opening, and closing forms, etc. There are a lot of complex workflows involving document processing, verification, and authorization, especially in the case of customer onboarding, which is tightly coupled with regulatory requirements such as KYC. With OCR, this workflow can be expedited, and inconsistencies in data can be identified at an early stage. Any CTS (Cheque Truncation System) compliant cheque has mandatory security features like a bank logo in UV ink, void Pantograph, Micro lettering, Date, Rupee Symbol, Signatures, and re, Account number. For the accurate identification of these security features, different datasets are needed. Prateek [33], MNIST dataset of handwritten digits used to identify the legal amount that is digits. This dataset has more than sixty thousand examples to identify handwritten characters, the EMNIST dataset can be used. This dataset has more than 88800 samples. These two datasets contain balanced images. This dataset is normalized. The MNIST dataset has ten classes. The cheque images can be obtained from the IDRBT bank cheque dataset, which contains 112 cheque images.

Every dataset has certain challenges and limitations. RVL-CDIP has limited diversity of documents, also it has class imbalance as certain class have more sample data compared to other classes. The tables, figures and headings are not annotated in RVL-CDIP and the contextual data like matadata, source authorship and relationship between documents is not available for entire dataset. In MNIST dataset images are grayscale and contains basic handwritten digits. EMNIST dataset mainly has english and latin handwritten characters. This dataset is not suitable for non english and non latin documents. It has limited variations in writing style, shapes and sizes of characters. It does not capture full range of variations found in real world handwriting This impacts models generalization capabilities. The data available for training is very limited. This dataset has alphabets written as isolated characters. But handwritten characters are often written in the context of words, sentences or paragraphs. Bank Search dataset has loan documents which has unstructured data and contains text heavy documents. These documents are prone to errors, inconsistencies and missing information. Therefore, data cleaning and preprocessing is necessary. Loan documents can vary in format, structure and loan types. These documents lack contextual information about borrower’s credit history, or the purpose of the loan. Turkish bank dataset does not include all types of loan documents. This dataset lacks diversity in terms of loan purposes, borrower profiles or loan terms. This dataset has inconsistencies due to missing values and errors. Therefore extensive data cleaning and preprocessing steps are necessary. As loan documents contain complex legal and financial terminology, understanding and interpreting loan documents can be complex. This dataset lacks temporal information like when the loans were issued or when the repayments occurred. NIBSS is a financial dataset related to Nigeria’s banking industry. It mainly consists of inter bank transactions, therefore has limited diversity in the dataset. This dataset has limited contextual information. The FUNSD dataset contains scanned copies of invoices, resumes and application forms. It does not cover all varieties of real world documents. As this dataset is very small, therefore it has limited generalisation capability. This dataset provides annotations at the word and character level, but it has annotations inconsistency. NIST- SPDB6 has limited diversity in the dataset with varying image quality. Tobacco 3482 is also a small dataset which has class imbalance. SROIE dataset

contains images which needs preprocessing to improve quality of the images. The images are complex due to logos, stamps or graphics in the background.

Table 4 shows a comparison of different datasets used for evaluating the performance of OCR in the banking domain. Engin et al. [34] developed a multimodal deep neural network for classifying Turkish Banking Customer Order Documents. They have created a dataset with approximately twenty-seven thousand documents of forty-five different classes. The main challenge while creating this dataset was wrongly labeled documents by the back office, also many documents were visually similar. The presence of support documents like identity proofs of customers, driving licenses, property papers, and blank documents also creates problems in the development of clean datasets.

Campos et al. [7] present a dataset of Portugal's public sector banks. This dataset contains information about stocks, loans, and, net transactions. This dataset has details like the type of banks whether they are domestic or international banks ownership details, country, balance sheet structure, liquidity, profitability, and size. Sinka and Corne [35] present a method of developing a dataset. The first step in this is to select the web documents that have some content in them. To save time, the Open Directory Project and the Yahoo! Directory are used to provide web documents that are manually categorized. This dataset contains almost one thousand documents in each of the eleven categories.

This research is about document processing in the banking and financial sector. The success of document digitization and automation lies in the quality of the dataset. For the preparation of high-quality datasets, only those documents that are easily categorized by OCR are chosen.

Oral et al. [36] used a dataset of 3500 Turkish bank documents. These documents are extracted from email, fax, and scanner. To get different layouts, these documents are collected from different customers. It is also concluded that as the dataset increase, the performance also increases. There are datasets available for different domains in banking. Malkadi et al. [30] have developed a dataset for electronic banking in the Central bank of Nigeria. The data is present in the form of tables and graphs. This data is used for reviewing the performance of e-payment channels.

The FUNSD dataset is the dataset used for text detection in scanned documents, OCR, Layout Analysis, and form understanding. This dataset consists of 199 documents, all of which are in image format. SROIE (Scanned Receipt OCR and Information Extraction) consists of around 1000 images of scanned receipts. RVL-CDIP is a dataset of letters, forms, emails, resumes, memos, etc. in scanned format. There are sixteen different classes in the RVL-CDIP dataset. But the images are low quality, noisy and low resolution.

BCSD is bank filled cheque dataset. The data is in image Jpeg format. A total of one hundred and fifty-eight cheque images of the same bank are present in the dataset.

For loan applications as well as death claim clearance, certain supporting documents are required. For example, aadhar card or PAN card copies, driving license, etc. Kaggle provides an image dataset for aadhar card images and synthetic PAN card data [36].

3.6 Performance metrics for evaluation of OCR

Performance metrics used for evaluating OCR performance are mainly Precision, Recall and F1-score. Many text

classification algorithms use precision and recall as performance metrics. But these metrics mainly depend on true positives. In document classification, the count of the negative classes is very high due to many unrelated documents. Precision and recall fail to detect changes in the true negative.

Confusion Matrix: To measure classification quality by measuring the count of how correctly and incorrectly the text is classified. Consider input text has C_j number of classes.

$$X = \{C_1, C_2, \dots, C_j\}$$

Correct classification of positive examples is used to determine precision, recall, F1 score, and Break-Even point.

These measures consider only the classification of positive examples but do not consider the classification of negative examples.

Accuracy: It can be calculated from the confusion matrix as the ratio of the sum of true positive and true negative with the sum of true positive, true negative, false positive, and false negative.

Recall: It is the ratio of the specific true positive document to all documents that are true. Accuracy and recall measure the effectiveness of class identification by a classifier. It is the extent of the error caused by false negatives.

Precision: It is the ratio of the true positive document to the sum of true positive and true negative documents. It is the extent of error that is caused by false positives.

F1 score: If precision and recall values are given for different models, the F1 score is used to decide which model is best. It works well for imbalanced datasets. It is the harmonic mean of precision and recall.

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (1)$$

Character Recognition Accuracy: This is nothing but the number of characters correctly recognised by the OCR system.

Word Recognition Accuracy: It gives number of words that are correctly recognised by OCR.

Recognition Time: This is the most important measure for comparing the time taken by OCR to process the document and recognise the text in the document.

Language Support: This metric measures versatility of the OCR engines. Different OCR engines support different multiple languages.

Font and Style Compatibility: Different documents have different fonts and styles which impact overall OCR accuracy. Therefore, OCR performance is tested on different fonts and formats for consistency.

Document Layout Accuracy: A good OCR recognises the texts as well as different elements in the document like headers, footers, tables and images.

4. CONCLUSIONS

This overview of the literature makes use of studies that show how new technologies assist organizations and customers in generating value. Digitization has had a significant impact on the Indian banking industry. It is difficult to thwart the growth and services of digital banking in the digital age. Everybody uses a smartphone, a smart mobile device, to access digital banking services whenever they want and from wherever they are. Digitalization in the financial

industry is thus now inescapable. Since they want quick access to financial services, people prefer this digital banking system to traditional banking, which necessitates users to go to a bank branch.

In this study, a survey is conducted on various aspects of the digitization of banking documents. There are different datasets available for handwritten digits, signatures, cheques, and scanned documents. There is no complete and free dataset for all bank documents like KYC, loan documents, cheque processing, and tax calculations. The documents after scanning may have noise. The algorithms used for printed document differs from the algorithms used for handwritten documents. The comparison results indicate that the performance accuracy of deep learning methods is greater than all the other methods. A convolutional neural network gives the best segmentation accuracy. After segmentation and feature extraction, documents are classified. For classification, neural networks and deep learning methods can be used. Machine learning methods like Naive Bayes and support vector machine gives good accuracy in document classification. These methods work by learning the correlation between features like keywords and document categories. Deep learning methods like CNN can automatically learn features from the pixel data in scanned documents and are able to learn complex patterns.

Additionally, this work adds to the digitization literature by illuminating the principles underlying it in light of the developments made in recent years. The following argument is supported by a number of authors [30] have reached the conclusion that further discussion on the Digitization agenda is required to fully grasp how digital initiatives are changing existing business models.

5. LIMITATIONS

This literature survey has some limitations:

1. The articles that are included in the review depend on the search phrase we employ; for instance, when we used the word "Digitization," numerous contributions that utilized synonyms were left out of this study. On the other hand, while doing systematic reviews, we must recognize that there is an order of proof and that the thing can be deductively said about the world is drawn from research with an explicit and strict design;

2. Despite the fact that we acknowledge that works in other languages may have different findings when eliminated, we opted to prioritize a correct interpretation of the articles, hence eliminating any kind of confusion;

3. We are also known that typically systematic reviews choose publications from many databases for more depth, but for this study, we chose to emphasize openness and simple replication;

4. The existing OCR engines are less accurate for noisy or low resolution. The accuracy drops for small fonts, highly styled fonts and Arabic scripts. To overcome these limitations high resolution scans, more training data are needed. Also, combining OCR with NLP methods can help to achieve better accuracy.

REFERENCES

- [1] Reddy, S.K., Reinartz, W. (2017). Digital transformation and value creation: Sea change ahead. NIM Marketing Intelligence Review, 9(1): 10-17. <https://doi.org/10.1515/gfkmir-2017-0002>
- [2] Gebayew, C., Hardini, I.R., Panjaitan, G.H.A., Kurniawan, N.B. (2018). A systematic literature review on digital transformation. In 2018 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE, pp. 260-265. <https://doi.org/10.1109/ICITSI.2018.8695912>
- [3] Reis, J., Amorim, M., Melão, N., Matos, P. (2018). Digital transformation: A literature review and guidelines for future research. Trends and Advances in Information Systems and Technologies, 1(6): 411-421. https://doi.org/10.1007/978-3-319-77703-0_41
- [4] Mergel, I., Edelman, N., Haug, N. (2019). Defining digital transformation: Results from expert interviews. Government Information Quarterly, 36(4): 101385. <https://doi.org/10.1016/j.giq.2019.06.002>
- [5] Jagtap, D.M. (2018). The impact of digitalisation on indian banking sector. International Journal of Trend in Scientific Research and Development. Digital, pp. 118-122. <https://doi.org/10.31142/ijtsrd18688>
- [6] Shetty, V.V., Apoorva, K., Nikhitha, K.V. (2019). Digital revolution in the indian banking sector. International Journal of Social and Economic Research, 9(3): 15-23. <http://doi.org/10.5958/2249-6270.2019.00015.1>
- [7] Campos, M.M., Mateus, A.R., Pina, Á. (2019). Sovereign exposures in the Portuguese banking system: Evidence from an original dataset. Search IDEAS, No. o201903.
- [8] Ceechini, S. (2002). Information and communications technology for poverty reduction. lessons from rural India. In IEEE 2002 International Symposium on Technology and Society (ISTAS'02). Social Implications of Information and Communication Technology. Proceedings (Cat. No. 02CH37293), pp. 93-99. <https://doi.org/10.1109/ISTAS.2002.1013801>
- [9] Fathima, J. (2020). Digital revolution in the Indian banking sector. International Journal of Commerce, 8(1): 56-64. <https://doi.org/10.34293/commerce.v8i1.1619>
- [10] Gogulamudi, S., Pinnela, V.K., Pathuri, L.S.T., Borra, R. (2020). Handwritten digit recognition by using pattern recognition & consensus clustering. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(6): 2263-2267. <https://doi.org/10.35940/ijitee.F3879.049620>
- [11] Stahl, B.C. (2021). Artificial intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies. Springer Nature, p. 124. <https://doi.org/10.1007/978-3-030-69978-9>
- [12] Tatsat, H., Puri, S., Lookabaugh, B. (2020). Machine learning and data science blueprints for finance. O'Reilly Media.
- [13] Pangbourne, K., Stead, D., Mladenović, M., Milakis, D. (2018). The case of mobility as a service: A critical reflection on challenges for urban transport and mobility governance. In Governance of the Smart Mobility Transition. Emerald Publishing Limited, pp. 33-48. <https://doi.org/10.1108/978-1-78754-317-120181003>
- [14] Chen, L., Chen, P., Lin, Z. (2020). Artificial intelligence in education: A review. IEEE Access, 8: 75264-75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- [15] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H., Shami, A. (2018). E-learning: Challenges and research

- opportunities using machine learning & data analytics. *IEEE Access*, 6: 39117-39138. <https://doi.org/10.1109/ACCESS.2018.2851790>
- [16] Romero, C., Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, 40(6): 601-618. <https://doi.org/10.1109/TSMCC.2010.2053532>
- [17] Guan, C., Mou, J., Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, 4(4): 134-147. <https://doi.org/10.1016/j.ijis.2020.09.001>
- [18] Trivedi, A., Srivastava, S., Mishra, A., Shukla, A., Tiwari, R. (2018). Hybrid evolutionary approach for devanagari handwritten numeral recognition using convolutional neural network. *Procedia Computer Science*, 125: 525-532. <https://doi.org/10.1016/j.procs.2017.12.068>
- [19] Choudhary, A., Rishi, R., Ahlawat, S. (2013). Off-line handwritten character recognition using features extracted from binarization technique. *Aasri Procedia*, 4: 306-312. <https://doi.org/10.1016/j.aasri.2013.10.045>
- [20] Shamim, S.M., Miah, M.B.A., Sarker, A., Rana, M., Al Jobair, A. (2018). Handwritten digit recognition using machine learning algorithms. *Global Journal of Computer Science and Technology*, 18(1): 17-23. <https://doi.org/10.17509/ijost.v3i1.10795>
- [21] Bora, M.B., Daimary, D., Amitab, K., Kandar, D. (2020). Handwritten character recognition from images using CNN-ECOC. *Procedia Computer Science*, 167: 2403-2409. <https://doi.org/10.1016/j.procs.2020.03.293>
- [22] Manchala, S.Y., Kinthali, J., Kotha, K., Kumar, J.J.K.S., Jayalaxmi, J. (2020). Handwritten text recognition using deep learning with Tensorflow. *International Journal of Engineering and Technical Research*, 9(5): IJERTV9IS050534. <https://doi.org/10.17577/IJERTV9IS050534>
- [23] Yang, T., Pasquier, N., Precioso, F. (2022). Semi-supervised consensus clustering based on closed patterns. *Knowledge-Based Systems*, 235: 107599.. <http://doi.org/10.1016/j.knosys.2021.107599>
- [24] Hallale, S.B., Salunke, G.D. (2013). Twelve directional feature extraction for handwritten English character recognition. *International Journal of Recent Technology and Engineering*, 2(2): 39-42.
- [25] Seshadripuram, K.K.M. (2020). Digitization of banks: An evidence from India. *International Journal of Research and Analytical Reviews*, 2(9): 681-689.
- [26] Singh, S., Srivastava, R.K. (2020). Understanding the intention to use mobile banking by existing online banking customers: An empirical study. *Journal of Financial Services Marketing*, 25(3-4): 86-96. <https://doi.org/10.1057/s41264-020-00074-w>
- [27] Mittal, R., Garg, A. (2020). Text extraction using OCR: A systematic review. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, Coimbatore, India, pp. 357-362. <https://doi.org/10.1109/ICIRCA48905.2020.9183326>
- [28] Ramdhani, T.W., Budi, I., Purwandari, B. (2021). Optical character recognition engines performance comparison in information extraction. *International Journal of Advanced Computer Science and Applications*, 12(8): 120-127. <https://doi.org/10.14569/IJACSA.2021.0120814>
- [29] Malathi, T., Selvamuthukumar, D., Chandar, C.D., Niranjan, V., Swashtika, A.K. (2021). An experimental performance analysis on robotics process automation (RPA) with open source OCR engines: Microsoft ocr and google tesseract OCR. In *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 1059(1): 012004. <https://doi.org/10.1088/1757-899X/1059/1/012004>
- [30] Malkadi, A., Alahmadi, M., Haiduc, S. (2020). A study on the accuracy of ocr engines for source code transcription from programming screencasts. In *Proceedings of the 17th International Conference on Mining Software Repositories*, pp. 65-75. <https://doi.org/10.1145/3379597.3387468>
- [31] Fadoju, O.S., Evbuomwan, G., Olokoyo, F., Oyedele, O., Ogunwale, O., Kolawole, O.O. (2018). Dataset for electronic payment performance in Nigerian banking system: A trend analysis from 2012 to 2017. *Data in Brief*, 20: 85-89. <https://doi.org/10.1016/j.dib.2018.07.046>
- [32] Sethy, A., Rout, A.K., Uriti, A., Yalla, S.P. (2023). A comprehensive machine learning framework for automated book genre classifier. *Revue d'Intelligence Artificielle*, 37(3): 745-751. <https://doi.org/10.18280/ria.370323>
- [33] Agrawal, P., Deepak, C., Vishu, M., Anatoliy, Z., Radu, P., Dragi, K., Christian, T. (2021). Automated bank cheque verification using image processing and deep learning methods. *Multimedia Tools and Applications*, 80(4): 5319-5350. <http://doi.org/10.1007/s11042-020-09818-1>
- [34] Engin, D., Emekligil, E., Oral, B., Arslan, S., Akpınar, M. (2019). Multimodal deep neural networks for banking document classification. In *International Conference on Advances in Information Mining and Management*, 21-25.
- [35] Sinka, M.P., Corne, D.W. (2005). The banksearch web document dataset: Investigating unsupervised clustering and category similarity. *Journal of Network And Computer Applications*, 28(2): 129-146. <https://doi.org/10.1016/j.jnca.2004.01.002>
- [36] Oral, B., Emekligil, E., Arslan, S., Eryiğit, G. (2020). Information extraction from text intensive and visually rich banking documents. *Information Processing & Management*, 57(6): 102361. <https://doi.org/10.1016/j.ipm.2020.102361>