

Deep Learning Based Cosine Text Feature Extraction from Fuzzy Filter Quality Improved Compressed Image Documents

Sharmila Chidaravalli

Department of Information Science & Engineering,
Global Academy of Technology, VTU, Bangalore, Karnataka, India.
Corresponding author: c.sharmila11@gmail.com
ORCID: <https://orcid.org/0000-0002-3109-2476>

Vimuktha E Salis

Department of Information Science & Engineering,
Global Academy of Technology, VTU, Bangalore, Karnataka, India.
E-mail: vimuktha.ej123@gmail.com
ORCID: <https://orcid.org/0000-0003-3428-936X>

(Received on December 19, 2025; Revised on February 13, 2026 & April 3, 2026 & June 5, 2026 & July 13, 2026;
Accepted on July 26, 2026)

Abstract

With the surge of process modernization and enhanced data management, the world is advancing towards digitization and automation more than was presumed. The Internet is becoming one of the most favored means of communication and multimedia, coupled with data compression methods, plays a crucial role in controlling the huge volume of data that is being produced on a regular basis. Modeling novel methods for the automatic analysis of compressed data without decompression is pressing requirement of the hour. In this work a deep learning method called Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) quality improved compressed image documents is proposed. The LACDNN-FF method is modeled using an agent AI model. To be more specific, three different agent AIs, namely the Image Enhancement Agent (IEA), Layout Analysis Agent (LAA) and Text Feature Extraction Agent (TFEA) are employed in the design of the LACDNN-FF method. Initially, with the scanned image dataset of compressed image documents as input, two dimensional flexible fuzzy filter-based pre-processing is applied by the IEA with the intent of eliminating noise, thereby generating quality-enhanced compressed documents. Next, Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation is applied by the LAA to perform document segmentation and classify the contents using pixel-precise semantic segmentation. Finally, Cosine CNN-BiLSTM (Convolutional Neural Network - Bidirectional Long Short-Term Memory) neural network based Text Feature Extraction is applied by TFEA for accurate text extraction and text feature processing. The proposed deep learning method LACDNN-FF outperforms existing methods and shows a promising solution for accurate text feature extraction with minimal error rate.

Keywords- Two dimension flexible fuzzy filter, Linear activated discriminant analysis, Deep neural network, Cosine convolutional neural network, Bidirectional long short-term memory.

1. Introduction

In the modern world of big data, because documents are increasingly being created and stored in the compressed form to bestow better storage and transmission effectiveness, the issue of recognizing work in the corresponding compressed domain in the absence of decompression becomes very demanding. Conventional method use decompression and then apply learning mechanisms on top of them hence, novel mechanisms are to be modeled with the intent of applying learning mechanisms directly in the compressed domains.

Deep learning based Document layout (DeepDoc) method was proposed (Rajan & Devasena, 2025) for recognizing layout anchor boxes and analyzing text in scanned documents. Owing to divergences in layout,

text orientations and picture quality, scanned documents occasionally present difficulties. Due to this, a reliable deep learning method that can ascertain anchor boxes and extract the foremost data from scanned papers was designed. Then, the data was pre-processed and fed into optimal feature selection model employing Coati's Optimization Algorithm (COA). On the basis of the optimum features learned by the COA the document was analyzed with improved precision and accuracy.

A novel deep learning method called Discrete Cosine Transform-based Compressed Segmentation Network (DTC-CompSegNet) was proposed (Rajesh et al., 2024) that learned features directly from the DCT compressed streams to achieve layout segmentation directly in the JPEG compressed domain. In comparison with the previous layout segmentation methods that operated in the pixel domain, the contribution here is that compressed streams of DCT coefficients extracted from JPEG documents were utilized in training. The feature learning here in turn contributed to improved precision and accuracy.

Over the past few years, digital data has been created rapidly in the form of digital documents. The exponential increase with huge volume of images and textual information has paved for several issues in processing owing to unstructured content. With this inclusion of large number of such images, indexing and ascertaining them becomes a demanding issue. A novel deep learning method called the discrete wavelet transform compressed convolution neural network was proposed (Bisen et al., 2023a). The designed method was found to be efficient both in terms of time and space. Yet another layout analysis and text characterization method employing convolutional neural network (CNN) was presented (Li et al., 2025) with accurate OCR processing. Nevertheless, the accuracy was affected owing to unstructured nature of the data. To address this aspect, a novel layout analysis method was proposed (Horadi et al., 2022) in compressed domain with improved precision.

Image text written in images of varied formats can be found in scanned documents, magazines, newspapers and so on. These image texts are highly available nowadays owing to their ease of access. Nevertheless, the variation of these texts owing to different orientation, alignments and embedded text in complicated colored document images make problem text extraction extremely laborious and demanding problem. To handle the issue, text extraction from image documents employing a connected component and edge based method was investigated (Natei et al., 2018) to enhance performance of and accuracy. However, segmentation was not performed. To overcome the issue, automatic tracing was applied (Rajesh et al., 2020) for segmentation of text lines and words with improved detection and recognition. However, the detection of duplicate images was not addressed. To concentrate on this aspect, information extraction employing deciphering by a hidden Markov Model was proposed (Choudhari & Dakhane, 2018) with improved accuracy.

Text detection from natural scenic imagery plays a crucial aspect as far as Content Based Imagery Analysis is concerned. The paramount and inevitable issues are numerous orientation styles, illumination effects, typical backgrounds and so on. A Text Information Extraction (TIE) method employing a Support Vector Machine (SVM) to detect text in a robust fashion from natural scene images was proposed (Golla et al., 2024) with improved precision and recall. A method to automatically extract PDF table data employing DL with specific concentration on the application of paddleOCR framework was presented (Yang & Yu, 2025).

Nevertheless, conventional image compression methods pose notable issues in terms of both quality and efficiency when applied to high-resolution imagery. To address these issues, a DL based image compression method using Channel-Temporal Attention Module was proposed (Ji et al., 2025), thereby improving the peak signal to noise ratio. The feasibility of applying natural language processing (NLP) with regard to text data extraction for systematic literature reviews was investigated (Du et al., 2025). Here, DL-based NLP was applied with improved scalability. Yet another image quality enhancement method employing edge

extraction along with assessment methods for blurriness was applied (Zhou et al., 2024). Focusing on these two aspects resulted in an improvement of linear coefficient.

Evaluating text features with scanned images is a complicated procedure requiring scientific research and careful analysis, with precise research putting forward prospective for increased results analysis, whereas others studies focus on uncertainties and drawbacks. Earlier, researchers have endeavored to reveal associations between statistical analysis and text feature extraction analysis, nevertheless, for the most part by employing statistical methods. Apart from the above stated issues, our study explores the association between Deep Learning and accurate text feature extraction from compressed images. The major objective of the present study is to develop Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) using scanned image analysis.

1.1 Research Problem

Document analysis plays a vital role in enabling efficient and accurate optical character recognition (OCR). OCR is defined as the recognition and processing of the content of paper documents or image files containing text in order to obtain the text content and layout information. Layout segmentation is still demanding issue in document images. Traditionally, layout segmentation was performed in the pixel domain. However, the required image quality was not obtained with traditional ML methods. Several machine learning and deep learning methods have been introduced for text extraction. Text Extraction is useful for information retrieval, searching, editing, documenting, archiving or reporting of image text. However, variation of these texts owing to differences in size, alignment, text embedded in difficult colored document images, low quality image, and low image contrast and complex background makes text extraction an extremely difficult and challenging problem. To address this, concern the proposed LACDNN-FF is developed with pre-processing, segmentation and text feature extraction to obtain improved compressed image documents.

1.2 Objective of this Study

- To achieve higher quality compressed image documents, Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) is developed.
- To attain quality-enhanced compressed documents, two dimension flexible fuzzy filter-based pre-processing is used.
- To accurately identify and classify precise logical regions of document, linear activated discriminant analysis with deep neural network-based segmentation is employed.
- To enable robust text feature extraction, Cosine CNN-BiLSTM neural networks are utilized.

1.3 Novelty and Contributions of the Work

The major novelty and contributions of this paper are as follows:

- A Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) for quality improved compressed image documents is designed. The hybrid method of employing a LACDNN-FF via three different agents for pre-processing, segmentation and text feature extraction.
- Two dimension flexible fuzzy filter-based pre-processing is employed in LACDNN-FF via IEA by using sliding window concept to obtain quality enhanced compressed images.
- Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation is utilized in LACDNN-FF for segmenting documents and classifying contents as headings, paragraphs, tables or images via different classes. Therefore, improving precision and accuracy extensively.

- Cosine CNN-BiLSTM neural networks based Text Feature Extraction is applied in LACDNN-FF via TFEA to extract the features with a minimal error rate.
- The scanned image dataset is utilized to conduct the experimental analysis for the proposed evaluation of text feature extraction. The best prospective combination of text feature extraction that aids in improving the precision and accuracy performance has been identified. The comparative analysis of the proposed LACDNN-FF method has been evaluated based on scanned images in terms of precision, recall, accuracy, error rate and F-measure.

1.4 Organization of the Work

The remainder of this paper is organized as follows: Section 2 describes the recent related works related to text feature extraction mechanisms designed using machine learning and deep learning via scanned and hand written images. Section 3 provides the detailed description of proposed method for text feature extraction analysis using deep learning. Section 4 presents the experimental setup with detailed qualitative and quantitative analysis. In addition, detailed discussion is presented with comparison of the traditional methods. Finally, Section 5 concludes the paper.

2. Related Work

Over the recent few years, it has become paramount to extract text from several formats, consisting of both handwritten and printed documents. The potential to detect and ascertain text is a demanding task in several fields, including analysis of document, optical character recognition, image processing and so on. Hybrid DL architecture was designed (Al-Khafaji & Ramaha, 2025) with improved scalability. Although compressed images are advantageous for communication purposes, their processing becomes laborious owing to the decompression overhead and re-compression operations that are required as many times as the data necessitates to operate on. Hence, direct operation over the JPEG 2000 compressed documents was proposed (Bisen et al., 2023b) with the intent of extracting text and non-text regions in the absence of any segmentation algorithm. CNN and Bi-LSTM were applied (Sinhuja et al., 2024) to handle variability and text complexity involved in handwritten and printed document formats. Applying these techniques resulted in an overall improvement in accuracy performance.

A systematic literature review of key information extraction using DL was investigated (Rombach & Fettke, 2025) for complicated business documents. Nevertheless, conventional methods that depend on extracting keywords directly frequently result in redundancy and fail to signify a strong association between summary and original document. A novel DL-based text summarization method including pre-processing, feature extraction, vectorization and summarization using hybrid Cat Swarm Optimization and Harris Hawk Optimization algorithm were designed (Rautaray et al., 2025) thereby illustrating superiority in text summarization tasks. Integrated DL and text mining tasks were combined to enhance data driven process in an end-to-end data extraction framework (Park et al., 2025).

In the era of digitization, medical document volume continues to increase, presenting challenges in effective information retrieval. Medical reports, frequently available as scanned documents include valuable clinical data prerequisite for decision making in healthcare domain. Nevertheless, manual information extraction and structuring is a laborious and time-consuming process, often susceptible to errors. A DL system for structured recognition of text images based on a differential binarization technique with convolution neural network was designed (Liu et al., 2022). A plethora of AI-driven methods were investigated (Mahadevkar et al., 2024) for analyzing unstructured documents. A novel genetic algorithm was applied (Malashin et al., 2024) along with NLP to fine tune OCR with improved accuracy.

Advanced DL techniques were applied (Sharma et al., 2024) for recognizing low resolution text with improved accuracy. Nevertheless, the peak signal to noise ratio (PSNR) was not focused. To analyze on this aspect, an adaptive image compression algorithm by integration clustering algorithm and DL technique was presented (Liang et al., 2024) with improved PSNR. A method to focus on minimizing error involved in extracting text features employing hybrid learning techniques were investigated (Joodi et al., 2023). Yet another method to focus on the recognition accuracy was proposed (Li, 2021) based on neural network model.

From the scrutiny of related works, limited traditional deep learning solutions have been designed to evaluate the efficiency of text feature extraction. Instead of focusing on entire feature subset of compressed image documents, applying Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation results in minimizing the error rate. With this objective a Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) for quality improved compressed image documents is proposed. The detailed description of the LACDNN-FF method is provided in the following sections.

3. Materials and Method

3.1 Materials

In this work, to extract text feature from quality-improved compressed image documents, a scanned image dataset (Suvradeep, 2026). It is obtained from <https://www.kaggle.com/datasets/suvroo/scanned-images-dataset-for-ocr-and-vlm-finetuning>. The scanned image dataset is a collection of scanned images constituting ten different document categories. The dataset is modeled in such a manner so as to improve Optical Character Recognition (OCR) and enable fine-tuning by providing a substantial heterogeneity of real-world document types that cover numerous domains and textual layouts. **Table 1** given below provides the dataset details.

Table 1. Details of dataset.

S. No.	Categories	Number of images in categories		S. No.	Categories	Number of images in categories
1.	Advertisements	231		6.	News	189
2.	Emails	600		7.	Notes	202
3.	Forms	432		8.	Reports	266
4.	Letters	568		9.	Resumes	121
5.	Memos	621		10.	Scientific papers	262
Total number of images						3492

The features of the considered dataset include, diversity, complex layout, textual variations and scanned image quality. Data augmentation process is performed to generate large volume of synthetic data, thereby augmenting the dataset, especially when the dataset size is small. The main aim of data augmentation is to create new data samples based on the existing data, enhancing the diversity and size of the dataset for training machine learning models. With the objective of simulating real world OCR issues, the dataset covers a broad spectrum of document types by including multi-column texts, tables covering both printed and handwritten, variations in noise and resolution to replicate practical cases. Using the above dataset, the proposed method is designed in the following sub-sections.

3.2 Methodology

This section explains the proposed method for extraction of text feature from quality improved compressed image documents using a deep learning method called Linear Activated Cosine Deep Neural Network based

text feature extraction from Fuzzy Filter (LACDNN-FF). The objective here is to first enhance the image quality of compressed image documents, then analyze the layout and finally label to compressed images of different classes and make the whole process of text feature extraction very efficient in terms of accuracy, precision and error rate. Proposed Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) is integrated with Two Dimension Flexible Fuzzy Filter-based for Pre-processing, Linear Activated Discriminant Analysis with Deep Neural Network-based for Segmentation and Cosine CNN-BiLSTM neural networks based Text Feature Extraction for enhancing the accuracy. Two Dimension Flexible Fuzzy Filter is employed to perform pre-processing to reduce noise and enhance the compressed quality images. Flexible fuzzy filters are effective at handling image uncertainty, fuzziness, and vagueness, allowing for intelligent smoothing that removes noise while preserving the edges of text characters.

Segmenting text from complex backgrounds or noisy compressed documents is challenging. Linear Activated Discriminant Analysis maximizes class separability between the text and background, reducing the dimensionality and computational cost. Deep Neural Network is employed to provide accurate, precise segmentation of text regions from the enhanced image. Text in images requires capturing both spatial and sequential characteristics for accurate recognition. Cosine CNN-BiLSTM is employed to extract the features and for learning hierarchical features from pixel-level data. This combined approach reduces time, increases accuracy, and text extraction performance. The proposed method termed as Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) is illustrated with a block diagram as shown in **Figure 1**.

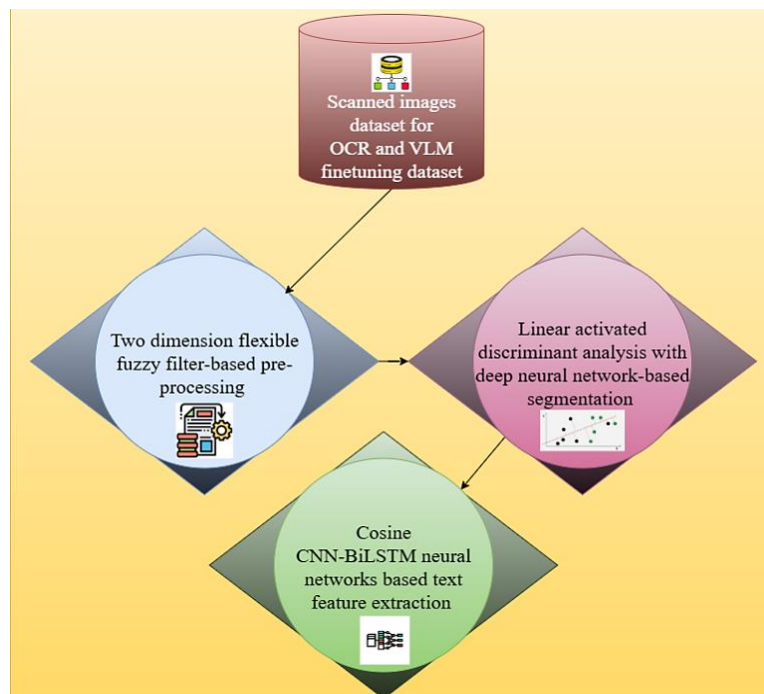


Figure 1. Block diagram of LACDNN-FF method.

Figure 1 demonstrate the block diagram of LACDNN-FF method. At first, the scanned image dataset of compressed image documents is considered as input. Documents and samples are gathered from the dataset.

Next, the presented method employs a novel Two Dimension Flexible Fuzzy Filter-based Pre-processing to prepare the compressed image documents for OCR. After preprocessing, document segmentation and content classification for compressed image documents necessitate ascertaining and classifying distinct content regions such as tables, images, text and so on frequently without full decompression.

In our work, Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation is applied with the aid of class separation that identifies and separates the different logical regions of the document in a precise and accurate fashion. Finally, text feature extraction is performed by combining the CNN-BiLSTM with cosine similarity, thereby ensuring accurate results. The detailed modeling of the proposed method is provided in the following sub-sections.

3.2.1 Two Dimension Flexible Fuzzy Filter-based Pre-processing

Before being pre-processed to discard noise, the compressed document image is initially checked for skew by Image Enhancement Agent (IEA). For optical character recognition systems to ascertain text features, pre-processing is essential. Pre-processing is utilized for minimizing background noise, underscoring the region of interest (i.e. the text feature) and efficiently isolating the foreground from the background in a compressed image. In the case of compressed document processing following data collection, image quality enhancement to isolate a text feature from the background of an image in our work, is essential and is performed by IEA. With this objective Image Quality Enhancement for Compressed Documents using Two Dimensional Flexible Fuzzy Filter-based Pre-processing is designed.

Conventional linear filters tend to produce blurred images. Fuzzy filters are utilized to provide a more precise and robust mechanism for noise reduction but not limited to, low-contrast or noisy, poor-quality image scenarios. Contrary to exiting methods, Two Dimensional Flexible Fuzzy Filter-based Pre-processing is chosen over traditional image enhancement methods because of its superior ability to handle uncertainty, maintain edge integrity and adapt to varying image qualities. Fuzzy filters are applied to preserve crucial image details, edges, and textures. This avoids blurring the images. **Figure 2** shows the structure of Two Dimension Flexible Fuzzy Filter-based Pre-processing.

As shown in the **Figure 2**, to start with the compressed image documents ‘ S ’ from dataset ‘ DS ’ provided as input are subjected to filtering using flexible fuzzy filtration function. Next, the compressed sample document membership function is formulated with the intent of defining the area of filter fuzziness.

With this the filtration function the dark points are eliminated because their weight is small compared to that of the remaining pixels in the sliding window. The image filtration is based on the use of a two dimensional flexible fuzzy filter with a sliding window of size ‘ $P * Q$ ($P = 2A + 1, Q = 2B + 1$)’ pixels. The filter performance for 8bpp images (in BMP format) is presented using the following relations.

$$S_{fil}(i, j) = \frac{\sum_{a=-A}^A \sum_{b=-B}^B \mu(i+a, j+b) S(i+a, j+b)}{\sum_{a=-A}^A \sum_{b=-B}^B \mu(i+a, j+b)} \quad (1)$$

From the above Equation (1), ‘ $S(i, j)$ ’ and ‘ $S_{fil}(i, j)$ ’ denotes the pixels analogous to the original compressed sample document and the filtered compressed sample document respectively. The membership function is activated as given below.

$$\mu[\Delta(i + a, j + b)] = \begin{cases} 1 & \text{for } \Delta(i + a, j + b) \leq \alpha \\ \frac{\Delta(i+a, j+b) - \beta}{\alpha - \beta} & \text{for } \alpha \leq \Delta(i + a, j + b) \leq \beta \\ 0 & \text{for } \Delta(i + a, j + b) \geq \beta \end{cases} \quad (2)$$

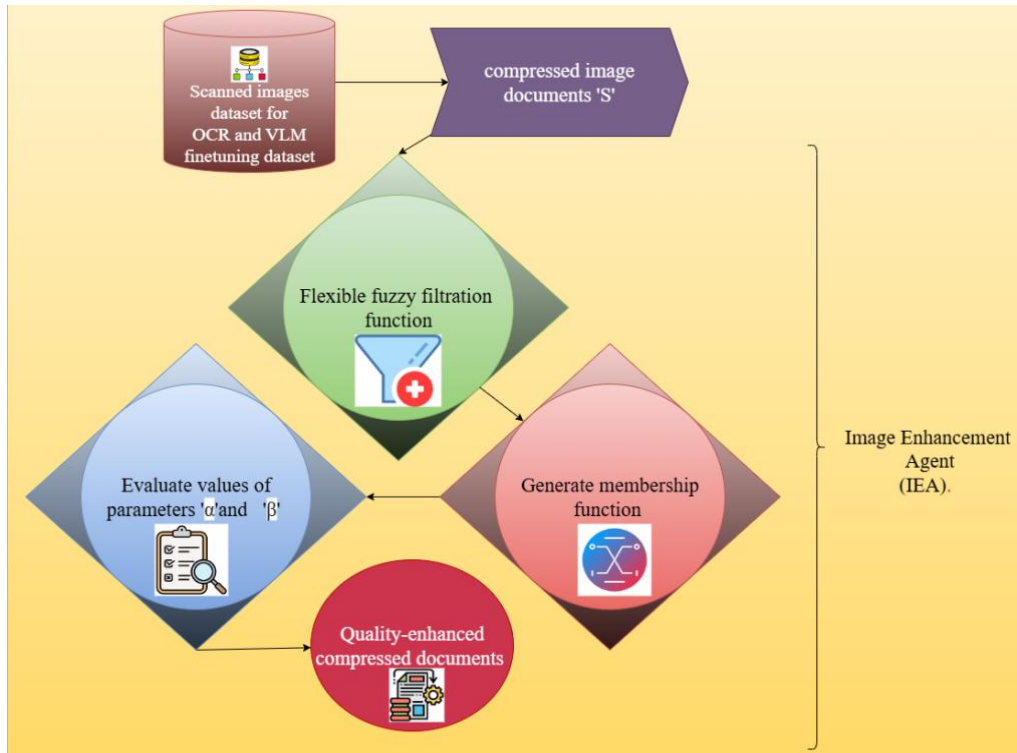


Figure 2. Structure of two dimension flexible fuzzy filter-based pre-processing.

From the above Equation (2), the membership function ' μ ' with parameters ' α ' and ' β ' defines the area of filter fuzziness. The increment function ' Δ ' refers to the brightness difference between the central pixel ' $S(i, j)$ ' in sliding filter window and pixel ' $S(i + a, j + b)$ ' that is at a distance ' (a, b) ' from the subsequent pixel respectively. The increment function ' Δ ' result is formulated as given below.

$$\Delta(i + a, j + b) = |S(i, j) - S(i + a, j + b)| \quad (3)$$

Subsequently the values of parameters ' α ' and ' β ' are defined as given below.

$$\alpha = \gamma - \varepsilon; \beta = \gamma + \varepsilon \quad (4)$$

From the above Equation (4), ' γ ' represent the center of filter fuzziness and ' ε ' designate the margins of the deviation from ' γ '. Finally, the value of center of filter fuzziness ' γ ' is obtained as given below.

$$\gamma = \frac{1}{2} [Diff_{MaxB}] \quad (5)$$

From the above Equation (5), the value of center of filter fuzziness is arrived at based on the maximum brightness difference between background portion and region of interest in the compressed documents via Image Enhancement Agent (IEA). **Figure 3** shows the flow process of two Dimension Flexible Fuzzy Filter-based Pre-processing.

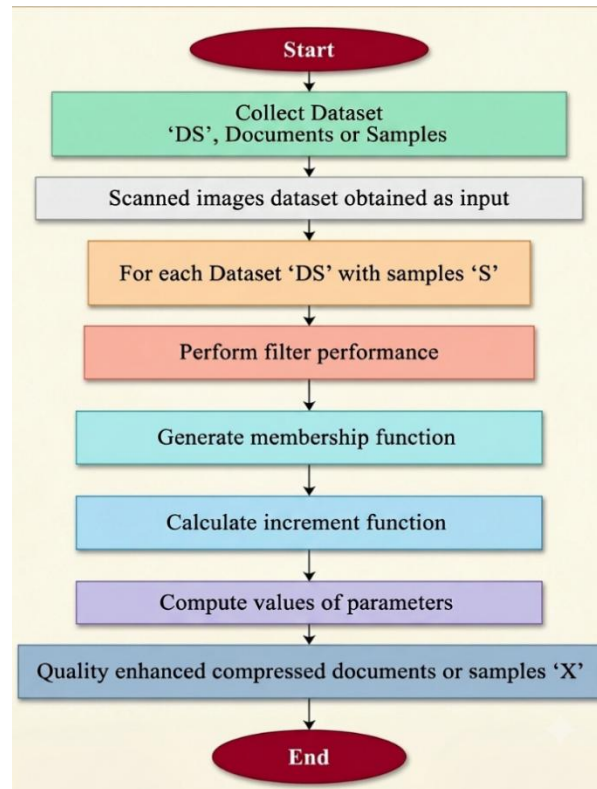


Figure 3. Flow process of two dimension flexible fuzzy filter-based pre-processing.

The pseudo code representation of two dimension flexible fuzzy filter-based pre-processing to ensure image quality enhancement for compressed documents is given below.

Algorithm 1. Two Dimension Flexible Fuzzy Filter-based Pre-processing
Input: Dataset ' DS ', Documents or Samples ' $S = \{S_1, S_2, \dots, S_m\}$ '
Output: Quality-enhanced Compressed Documents
1: Initialize ' $m = 3492$ ' 2: Begin 3: For each Dataset ' DS ' with samples ' S ' 4: Perform filter performance according to Equation (1) 5: Generate membership function according to Equation (2) 6: Evaluate increment function according to Equation (3) 7: Evaluate values of parameters ' α ' and ' β ' according to Equations (4) and (5) 8: Return quality enhanced compressed documents or samples ' X ' 9: End for 10: End

In above algorithm, the scanned image dataset obtained as input. The documents or samples are split into ten different categories and characteristics. For each category and characteristics that consists of several sample instances, the process of image quality enhancement is performed. A filtering function is applied based on the sliding window for each sample set. A Membership function is designed with the objective of

defining the area of filter fuzziness. Next, a brightness difference increment function is formulated. Finally, based on the values of the parameters ' α ' and ' β ' the maximum brightness difference between background portion and region of interest is formulated with higher image quality and a higher peak signal to noise ratio.

3.2.2 Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation

With the generated quality enhanced compressed documents the second step is layout analysis. In our work, the layout analysis is performed by the Layout Analysis Agent (LAA). The LAA performs document segmentation and content classification. Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation is employed to separate different logical regions of the quality enhanced compressed documents such as, header, table, figure, images, etc. utilizing pixel-precise semantic segmentation. Every pixel in the quality enhanced compressed documents is assigned a specific class label such as headings, paragraphs, tables or images via different classes. This in turn provides a dense annotation map covering the entire set of quality enhanced compressed documents a holistic view at a pixel-by-pixel level. It groups all instances of the same class (e.g., all paragraphs) into a single region, without differentiating between individual objects within that class. Although Linear Activated Discriminant Analysis is a Machine Learning (ML) technique, applying non-linearity with deep neural network makes it a Deep Learning (DL) technique. In this work, Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation is applied by the LAA to perform both quality enhanced compressed document segmentation and content classification.

Linear Activated Discriminant Analysis (LDA) principles are applied to a deep learning architecture to learn linearly separable representations. It is particularly effective at maximizing the distance between different classes (text vs. background) while minimizing the variance within a single class. Deep Neural Network (DNN) Segmentation utilizes a deep learning model to handle pixel-level segmentation. This is crucial for segmentation in complex or noisy, high-compression documents. The LAA specifically uses this method to perform accurate layout analysis and content classification on document images. The LAA performs detailed segmentation to distinguish between different types of content (text, images) at the pixel level. **Figure 4** shows the structure of Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation model.

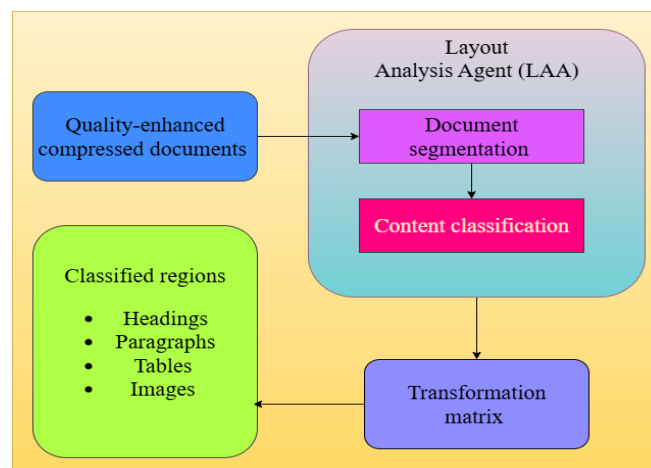


Figure 4. Structure of linear activated discriminant analysis with deep neural network-based segmentation.

As shown in the **Figure 4**, with the quality enhanced compressed documents or samples ' X ' obtained as input are subjected to layout analysis with the intent of performing document segmentation and content classification. These two tasks are performed by the Layout Analysis Agent (LAA). Initially, document segmentation is performed by transforming non-linear activation function of ML to linear activation function with deep neural network. After, transformation matrix employing within class of scatter matrix and between classes of scatter matrix is modeled via pixel-precise semantic segmentation to segment accurately. Finally, the classified regions are then identified as headings, paragraphs, tables or images via different classes by LAA.

Let us assume that we have a set of observations (i.e. quality enhanced compressed documents) ' X ' belonging to ' C ' distinct classes. The objective is to identify a projection matrix ' PM ' that converts set of observations ' X ' into another coordinate system ' $RX = X.PM$ ' in such a manner that linear class separability is increased and variance of each class is decreased. Let us consider that a conventional neural network takes as input a vector i.e., quality enhanced compressed documents or samples ' X ', multiplies it by a weight matrix ' W_{NN} ', sums up a bias vector ' B ' and finally models a non-linear activation function ' f ' to generate the output ' Y_{NN} ' as given below.

$$Y_{NN} = f(W_{NN} \cdot X + B) \quad (6)$$

For the above conventional neural network model Linear Discriminant Analysis (LDA) the transformation operation is then formulated as given below.

$$Y_{LDA} = W_{LDA} \cdot X \quad (7)$$

From the above two formulates (6) and (7), ' $Y_{NN} \neq Y_{LDA}$ ' because of the occurrence of non-linear activation function ' f '. Hence, linear activated discriminant analysis with deep neural network is designed forming a DL model in our work as given below.

$$f(W_{NN} \cdot X + B) = f(W_{LDA} \cdot X) \quad (8)$$

$$W_{NN} \cdot X = W_{LDA} \cdot X, \text{ with } b \text{ initialized to } 0 \rightarrow W_{NN} = W_{LDA} \quad (9)$$

From the above Equations (8) and (9), it shows the LDA transformation matrix ' W_{LDA} ' is then said to be used for initialization of neural network weight. This in turn will perform the class separation with regard to the training input sample. Then the matrix to perform class separation is done as given below.

$$CS = SM_{WC} SM_{BC} \quad (10)$$

From the above Equation (10), results the class separation ' CS ' are formed employing the within class results of scatter matrix ' SM_{WC} ' and the between class results of scatter matrix ' SM_{BC} ' respectively.

$$SM_{WC} = N' \sum \frac{1}{N_C} (X - \mu_C)(X - \mu_C)^T \quad (11)$$

$$SM_{BC} = N' \sum \frac{1}{N_C} (\mu_C - \mu)(\mu_C - \mu)^T \quad (12)$$

From the above Equations (11) and (12), the within class results of scatter matrix ' SM_{WC} ' and between class results of scatter matrix ' SM_{BC} ' are arrived at based on the mean number of areas (text, table, images, etc.) per class ' N' ' and the number of areas ' N_C ' belonging to class ' C ' respectively. **Figure 5** describes the flow process of linear activated discriminant analysis with deep neural network-based segmentation.

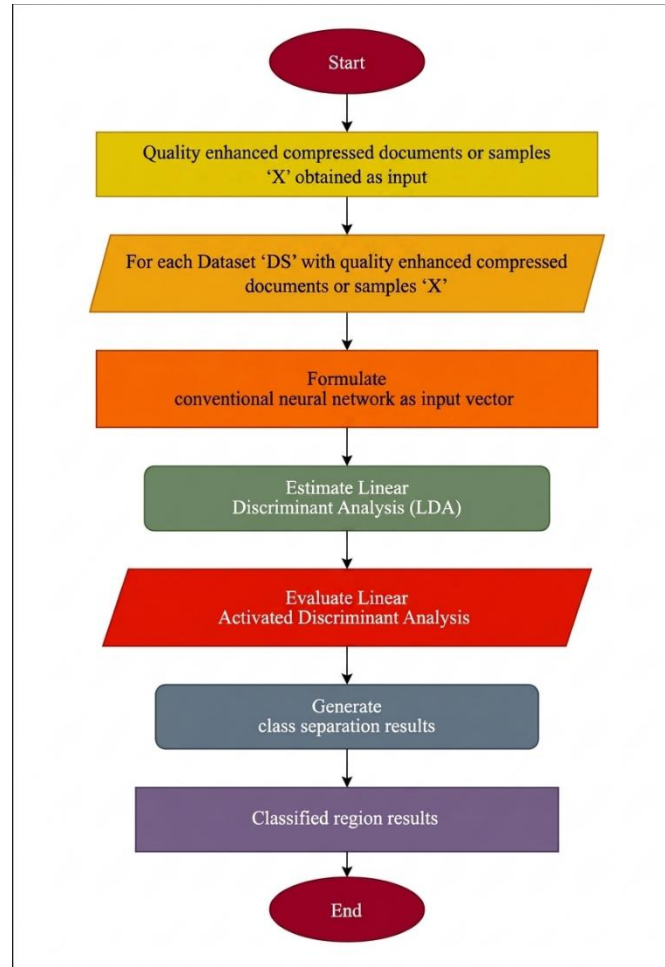


Figure 5. Flow process of linear activated discriminant analysis with deep neural network-based segmentation.

The pseudo code representation of linear activated discriminant analysis with deep neural network-based Segmentation is given below.

Algorithm 2. Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation
Input: Dataset ' DS ', Documents or Samples ' $S = \{S_1, S_2, \dots, S_m\}$ '
Output: Accurate and precise logical regions of document with distinct classes
1: Initialize ' $m = 3492$ ', quality enhanced compressed documents or samples ' X ' 2: Begin 3: For each Dataset ' DS ' with quality enhanced compressed documents or samples ' X ' 4: Formulate conventional neural network as input vector according to Equation (6) 5: Evaluate Linear Discriminant Analysis (LDA) according to Equation (7) 6: Formulate Linear Activated Discriminant Analysis according to Equations (8) and (9) 7: Generate class separation results according to Equations (10), (11) and (12) 8: Return classified region results ' C ' 9: End for 10: End

In the above algorithm, the conventional non-linear activated discriminant analysis of ML is converted to linear activated discriminant analysis of DL for ascertaining and splitting distinct logical regions of the document such as paragraph, images, tables, heading. To perform this task, initially, a conventional neural network is employed to formulate input vector. Following which, Linear Discriminant Analysis (LDA) and Linear Activated Discriminant Analysis results are obtained to generate class separation results. This in turn not only performs document segmentation but also classifies the content in an accurate and precise manner.

3.2.3 Cosine CNN-BiLSTM Neural Networks based Text Feature Extraction

Finally, in this section, with the quality enhanced compressed image documents and the classified logical regions of document with distinct classes as input, text feature extraction is performed by the Text Feature Extraction Agent (TFEA). The quality enhanced compressed image documents with their classified logical regions in distinct classes are transformed into text features using a composite model that combines CNN-BiLSTM neural networks. This composite CNN-Bi-LSTM model influences the advantages of both models, with CNN extracting text features and the Bi-LSTM preserving sequential order in both directions. The fusion strives to optimize text feature extraction from quality enhanced compressed documents, using CNN's text feature extraction capabilities and Bi-LSTM's competence to process data bi-directionally. The proposed Cosine CNN-BiLSTM neural network comprises four distinct layers, namely, input layer, convolution layer, Bi-LSTM layer and output layer. **Figure 6** shows the structure of Cosine CNN-BiLSTM neural network based text feature extraction.

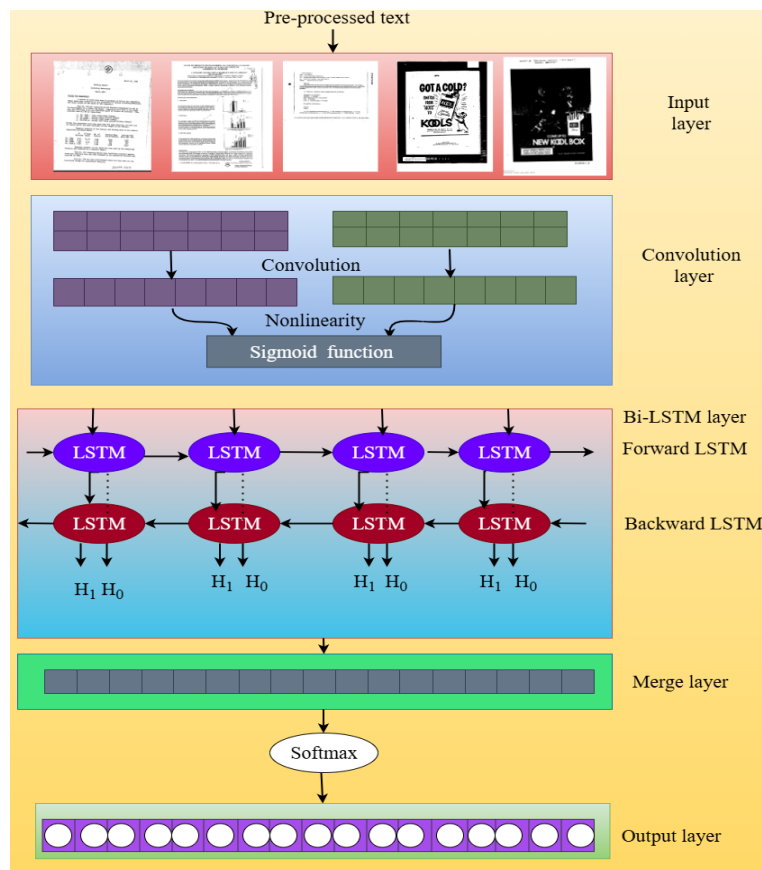


Figure 6. Structure of Cosine CNN-BiLSTM neural network based text feature extraction.

In above figure, the Cosine CNN-BiLSTM neural network includes an input layer, a convolution layer, Bi-LSTM layer, merge layer and an output layer. Quality enhanced compressed documents and classified region outcomes are considered as input. At the convolution layer, sigmoid activation is used to create feature maps. At the Bi-LSTM layer, text images are extracted through forward hidden state and backward hidden state. The Merge layer combines the forward and backward LSTM hidden states outcome through concatenation or averaging to create a single comprehensive feature representation. Lastly, the extracted text feature results are obtained at the output layer in less time.

Input Layer (Word Embedding): The TFEA with the aid of pre-trained word embedding functionalities such as Word2Vec, captures semantic and syntactic relationships between words in a sentence to represent the extracted text numerically for downstream tasks using cosine similarity. This is represented mathematically as given below.

$$Sim(a, b) = \cos(\theta) = \frac{a \cdot b}{||a|| \cdot ||b||} \quad (13)$$

From the above Equation (13), the result of '1' for the cosine similarity between two words ' a ' and ' b ' corresponds to the case in which the vectors are the same, which in turn indicates that the words are equivalent in the corpus. On the other hand, the result of '0' for the cosine similarity between two words ' a ' and ' b ' corresponds to the case in which the vectors are orthogonal, which in turn denotes that the two words are not related to each other. The proposed model obtains text data input, divides it into words or tokens and transforms each word into a numerical integer matrix, therefore performing efficiently with similar words facilitating several degrees of similarity.

Convolution layer within a CNN is where the convolution operation performed by the TFEA, which necessitates moving a small kernel across the quality enhanced compressed documents. The kernel being a matrix of trainable weights, at each stack pointer measures the dot product between its weights and corresponding region of quality enhanced compressed documents. This process is repeated for the entire portion, leading to the formation of a feature map. Let us represent the quality enhanced compressed documents or samples as ' X ' and the kernel as ' K ' then the convolution operation at a position ' (a, b) ' is mathematically formulated as given below.

$$FM(a, b) = Sum(X(a + i) * (b + j) * K(i, j)) \quad (14)$$

Convolution operation as given above a sigmoid activation function is employed within the feature map to introduce nonlinearity to the model as given below.

$$FM(a, b) = Sigmoid(FM(a, b)) \quad (15)$$

From the above Equation (15), the outcomes of all the activated results are combined to construct the input for the Bi-LSTM. The bidirectional long short term memory (Bi-LSTM) is then hybridized with the results of CNN by the TFEA via binary hidden states ' H_0 and H_1 ' thereby ensuring processing in both forward and backward direction. This in turn ensures that both the historical and upcoming data are retained, resulting in the output as given below.

$$H = \{H_0 + H_1\} \quad (16)$$

From the above Equation (16), ' H ' denotes the output of the Bi-LSTM based on the forward hidden state ' H_1 ' and backward hidden state ' H_0 ' respectively. The Bi-LSTM takes the extracted text features as input while also taking into consideration the context of the entire sequence. Finally, the Bi-LSTM processes the text features, therefore producing a series of words as output. The generated series constitutes the textual description for the input quality enhanced compressed documents. The output layer is the final step in this

model generated by the TFEA, and it provides text feature extraction results based on the quality enhanced compressed documents and classified regions extracted by the earlier layers. This is mathematically formulated as given below.

$$Z = \text{Softmax}(w * H + b) \quad (17)$$

From the above Equation (17), results the output layer 'Z' (i.e. text feature extracted) is activated via the 'Softmax' function employing the weight 'w' along with the bias 'b' and output results of Bi-LSTM 'H' respectively. **Figure 7** demonstrates the flow process of Cosine CNN-BiLSTM neural network based text feature extraction.

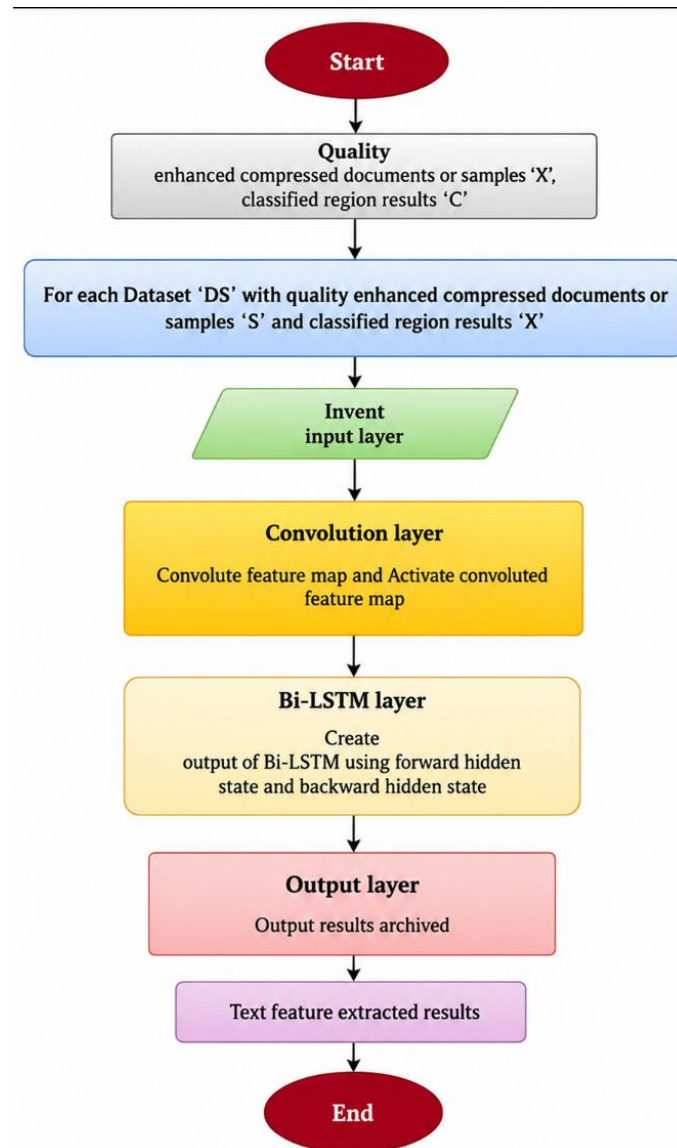


Figure 7. Flow process of Cosine CNN-BiLSTM neural network based text feature extraction.

The pseudo code representation of Cosine CNN-BiLSTM neural network based text feature extraction is given below.

Algorithm 3. Cosine CNN-BiLSTM neural network based text feature extraction
Input: Dataset ‘ DS ’, Documents or Samples ‘ $S = \{S_1, S_2, \dots, S_m\}$ ’
Output: robust text feature extraction
1: Initialize ‘ $m = 3492$ ’, quality enhanced compressed documents or samples ‘ X ’, classified region results ‘ C ’ 2: Begin 3: For each Dataset ‘ DS ’ with quality enhanced compressed documents or samples ‘ X ’ and classified region results ‘ C ’ // Input layer 4: Formulate input layer according to (13) // Convolution layer 5: Convolute feature map according to (14) 6: Activate the convoluted feature map according to (15) // Bi-LSTM layer 7: Generate the output of Bi-LSTM using forward hidden state and backward hidden state according to (16) // Output layer 8: Model the output results according to (17) 9: Return text feature extracted results 10: End for 11: End

In the above algorithm, text features are extracted with minimal error rate, the entire process is split into four layers. They are input layer, convolution layer, Bi-LSTM layer and output layer. In the input layer, quality enhanced compressed documents or samples and classified region results are provided as input. Following which, feature maps are generated and activated in the convolution layer via the sigmoid activation function. Next, Bi-LSTM layer text feature results are created via forward hidden state and backward hidden state. In the output layer, by employing this forward hidden state and backward hidden state for extracting text features in turn minimizes the error rate and time.

4. Experimental Setup

The proposed deep learning method called Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) quality improved compressed image documents is evaluated using Python running on the MS Windows platform on a computer with an i3-2350 processor and 8 GB RAM. The results are compared with three previously existing methods, Deep learning based Document layout (DeepDoc) (Rajan & Devasena, 2025) and Discrete Cosine Transform-based Compressed Segmentation Network (DCT-CompSegNet) (Rajesh et al., 2024) and CNN (Li et al., 2025) using Scanned Images Dataset. The dataset is taken from <https://www.kaggle.com/datasets/suvroo/scanned-images-dataset-for-ocr-and-vlm-finetuning>. After the data collection, data augmentation is carried out to artificially increase the size of the dataset. The number of samples ranging from 500 to 5000 is taken from dataset for conducting the experiments.

An objective of the proposed LACDNN-FF method is to obtain quality improved compressed image documents with higher accuracy, precision and recall and F-measure with minimum time and error. The

existing methods such as the DeepDoc (Rajan & Devasena, 2025), DCT-CompSegNet (Rajesh et al., 2024) and CNN (Li et al., 2025) are taken as base papers. These three base papers are explained in order to understand the proposed method. The proposed method concept is derived by considering the problems of these base papers. The drawbacks of these methods are effectively overcome by implementing the proposed method. Moreover, the results are analyzed and validated based on seven metrics, such as precision, recall, accuracy, F-measure, error rate, time complexity and L2 loss function.

Training settings are involved for optimizing the hyperparameters such as batch size, learning rate, epoch and optimizer. **Table 2** shows the hyperparameter of proposed LACDNN-FF.

Table 2. Hyperparameter of proposed LACDNN-FF.

Parameters	Description
CNN-BiLSTM	Four distinct layers such as input layer, convolution layer, Bi-LSTM layer and output layer
Input layer	Consider quality enhanced compressed documents or samples 'X' and classified region results 'C'
Convolution layer	Sigmoid activation function employed via convolution operation
Bi-LSTM layer	Perform forward and backward direction
Output layer Activation function	Softmax activation function used
Similarity metric	Cosine similarity
Learning rate	0.001
Batch size	32
Epoch	100
Optimizer	Adam

Cross-validation is used to compute the performance of the model ability to carry out unseen data. By using cross-validation, the dataset is divided into two sets such as training and testing. Most data samples (70%) were used for training and the remaining (30%) were taken for testing. In our work, the 10-fold cross-validation is utilized for measuring the results.

4.1 Case Study and Inferences

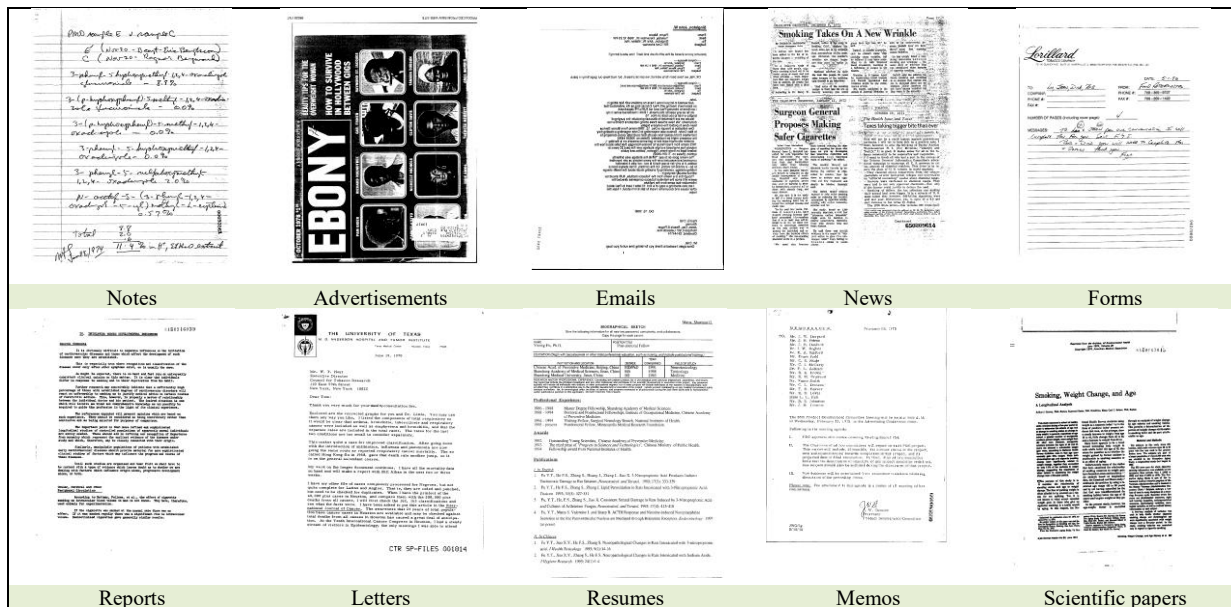


Figure 8. Ten different classes sample scanned images.

In this section case analysis of text feature extraction using Scanned Images dataset are simulated by applying Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) quality improved compressed image documents. **Figure 8** provides the sample of the scanned images of ten different classes.

As shown in the above figure, with the presence of ten different categories also as stated in **Table 1**, for simulation process we have used three different categories, namely, emails, reports and scientific paper. **Figure 9** given below provides three different sample images along with the pre-processed results.

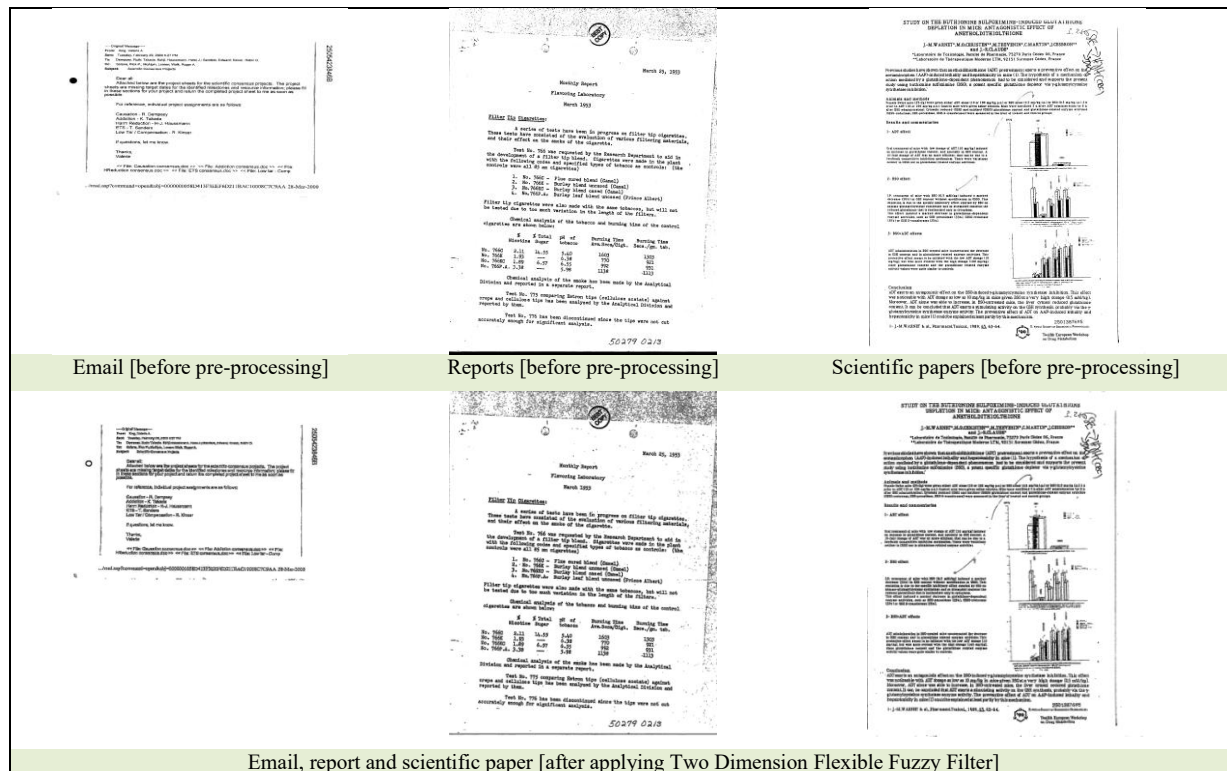


Figure 9. Pre-processed results of email, report and scientific paper.

As shown in the above figure with the pre-processing performed by applying Two Dimension Flexible Fuzzy Filter using flexible fuzzy filtration function, the dark points are discarded. This is because their weight is small compared to that of the remaining pixels in the sliding window. This in turn results in the minimization of error rate during the text feature extraction process. Following this, layout analysis is performed by applying Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation algorithm. **Figure 10** shows the layout analysis results.

As shown in **Figure 10**, employing the transformation matrix, the layout analysis agent, splits the samples into corresponding regions. Obtaining these segmentation regions in turn aids in improving the overall accuracy and precision involved in the text extraction process. Finally, with the pre-processed sample results and the layout analysis performed for each sample the text extraction result is shown in **Figure 11**.

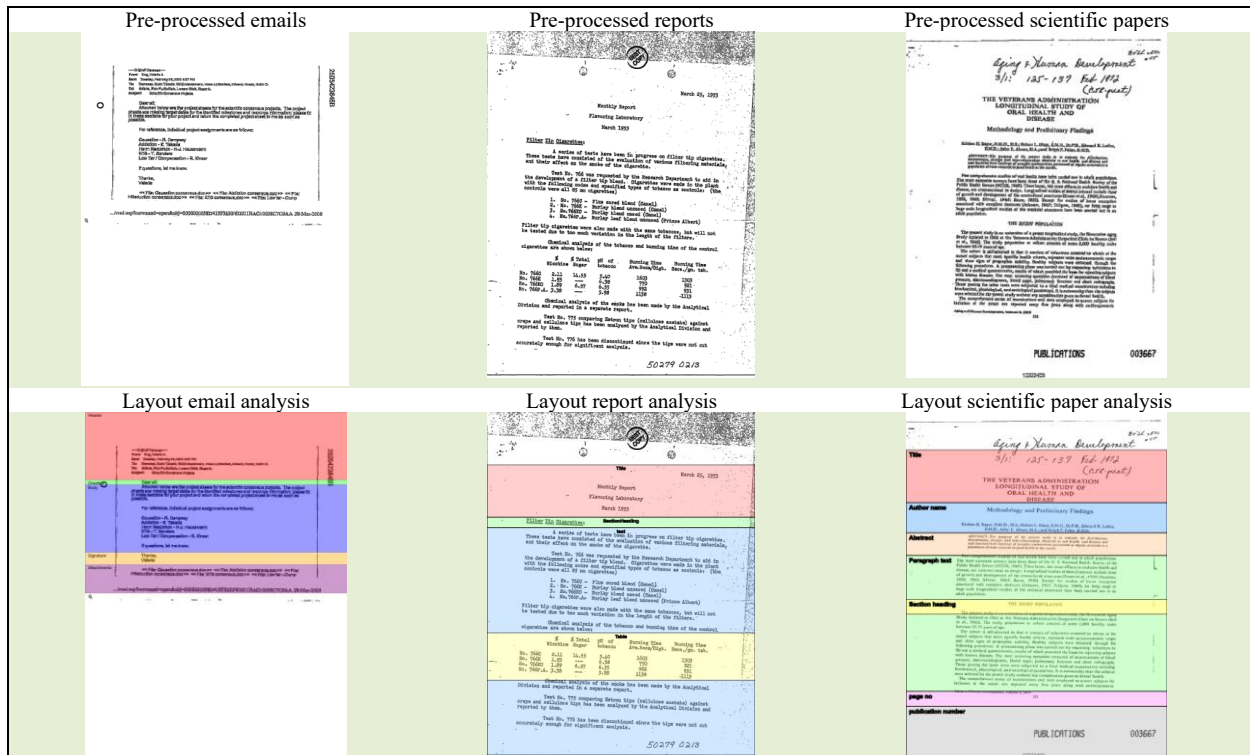


Figure 10. Layout analysis results of email, report and scientific paper.

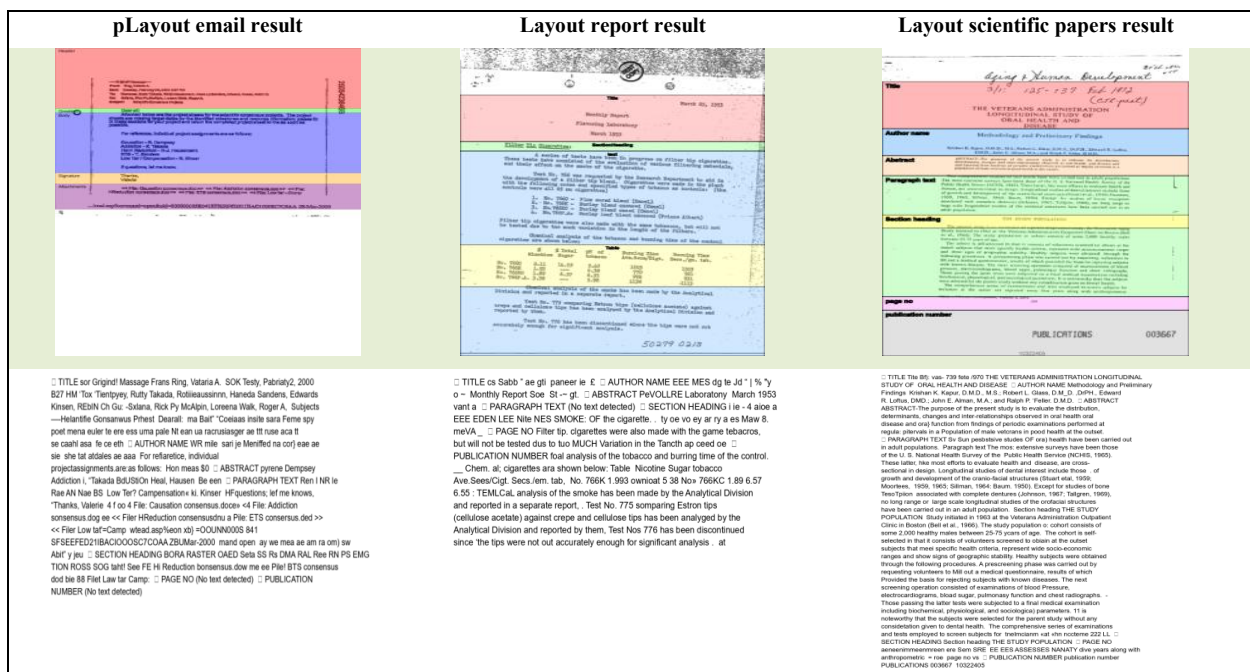


Figure 11. Text extraction results of email, report and scientific paper.

As shown in the **Figure 11** using Cosine CNN-BiLSTM neural network based text feature extraction algorithm Bi-LSTM processes the text features processes in both forward and backward direction via the text feature extraction agent. By using the bi-directional property in turn aids in minimizing the error rate in an extensive manner during the text feature extraction process. The quantitative analysis is provided in the following sub-sections.

4.2 Discussion

In this section a detailed analysis of six different performance metrics is analyzed in discussed in detail.

4.2.1 Performance Analysis of Precision, Recall and Accuracy

First, in this section the performance analysis of three different metrics precision, recall and accuracy analysis is made involved in the process of text feature extraction with quality enhanced compressed documents. Initially, precision is measured as given below.

$$Pre = \frac{TP}{TP+FP} \quad (18)$$

From the above Equation (18), precision '*Pre*', is measured by using the true positive rate '*TP*' (i.e. actual text extracted as text sample) and false positive rate '*FP*' respectively (i.e. actual text extracted as image sample). Next, recall is measured as given below.

$$Rec = \frac{TP}{TP+FN} \quad (19)$$

From the above Equation (19), recall '*Rec*' is measured using the true positive rate '*TP*' (i.e. actual text extracted as text sample) and false negative rate '*FN*' (i.e. image extracted as text sample). Finally, accuracy is mathematically represented as given below.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (20)$$

From the above Equation (20), accuracy '*Acc*' is measured employing the true positive rate '*TP*' (i.e. actual text extracted as text sample) and false positive rate '*FP*' (i.e. actual text extracted as image sample), true negative rate '*TN*' (i.e. image extracted as image sample) and false negative rate '*FN*' (i.e. image extracted as text sample) respectively. **Table 3** given below lists the precision, recall and accuracy results for the proposed LACDNN-FF method and the three existing methods, DeepDoc, DCT-CompSegNet and CNN when applied to the scanned images dataset.

Table 3. Precision, recall, accuracy results of proposed LACDNN-FF, existing DeepDoc, DCT-CompSegNet and CNN.

Sample s	Precision				Recall				Accuracy			
	LACDNN -FF	Deep Doc	DCT- CompSeg Net	CNN	LACDNN -FF	Deep Doc	DCT- CompSeg Net	CNN	LACDNN -FF	Deep Doc	DCT- Comp SegNet	CNN
500	0.978	0.972	0.968	0.959	0.99	0.981	0.978	0.974	0.974	0.964	0.95	0.938
1000	0.964	0.952	0.945	0.933	0.983	0.962	0.955	0.94	0.97	0.95	0.938	0.933
1500	0.982	0.95	0.946	0.932	0.981	0.95	0.942	0.935	0.965	0.947	0.925	0.921
2000	0.969	0.945	0.935	0.926	0.97	0.943	0.933	0.927	0.962	0.943	0.922	0.92
2500	0.948	0.932	0.93	0.92	0.968	0.94	0.93	0.923	0.959	0.94	0.937	0.93
3000	0.962	0.951	0.943	0.938	0.954	0.95	0.945	0.937	0.957	0.938	0.933	0.928
3500	0.981	0.973	0.967	0.958	0.962	0.953	0.947	0.942	0.973	0.952	0.95	0.946
4000	0.969	0.953	0.95	0.94	0.975	0.964	0.958	0.952	0.967	0.95	0.947	0.943
4500	0.953	0.946	0.937	0.929	0.969	0.952	0.943	0.938	0.962	0.948	0.945	0.941
5000	0.978	0.965	0.955	0.946	0.972	0.96	0.95	0.94	0.97	0.964	0.96	0.953

Figure 12 illustrates the graphical representation of precision, recall and accuracy for the proposed LACDNN-FF and three existing methods, DeepDoc, DCT-CompSegNet and CNN. The results are plotted on the vertical axis and the samples ranging between 500 and 5000 are plotted on the horizontal axis. From the above figure it is inferred that increasing the number of samples does not cause an increase or decrease in the precision, recall and accuracy. This is evident from the results of precision wherein for the first eight iterations the precision rate was lower and was higher for the rest of the two iterations. Similarly, the recall rate for the first six iterations was lower whereas for the rest of the four iterations it was observed to be higher. This in turn confirms the objective that the precision, recall and accuracy results are not affected by the increase in samples. Also, the comparative results are better for the proposed LACDNN-FF method upon comparison to DeepDoc, DCT-CompSegNet and CNN. The reason is that, by applying the Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation algorithm the traditional non-linear activated discriminant analysis of ML was transformed to linear activated discriminant analysis of DL by employing conventional neural network to formulate input vector. Next, Linear Discriminant Analysis (LDA) and Linear Activated Discriminant Analysis results were obtained with the intent of generating class separation results. This in turn not only performed document segmentation but also classified content. With this the overall precision using proposed LACDNN-FF method was observed to be better by 2% compared to DeepDoc, 2% compared to DCT-CompSegNet and 3% compared to CNN and recall rate improved by 2% compared to DeepDoc, 3% compared to DCT-CompSegNet and 3% compared to CNN. Also, by applying the Two Dimension Flexible Fuzzy Filter-based Pre-processing algorithm the maximum brightness difference between background portion and region of interest was formulated thereby emphasizing image quality. This in turn resulted in the accuracy improvement using proposed LACDNN-FF method of 2% compared to DeepDoc, 3% compared to DCT-CompSegNet and 3% compared to CNN respectively.

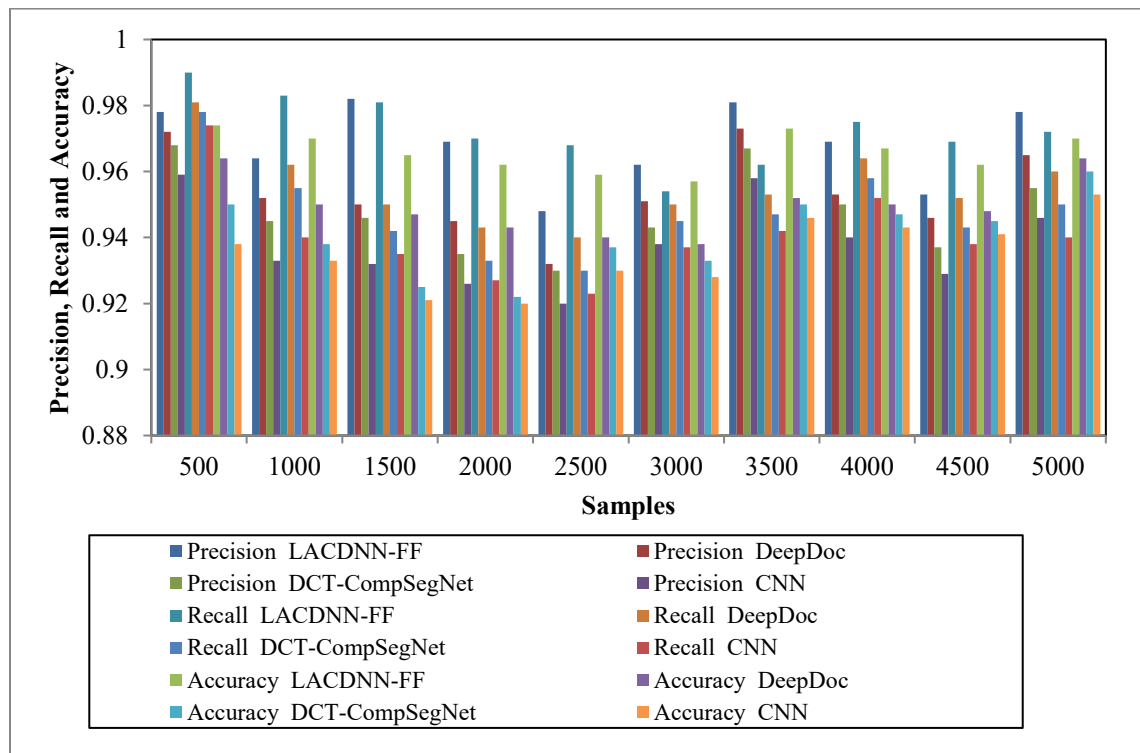


Figure 12. Precision, recall and accuracy versus samples.

4.2.2 Performance Analysis of F-Measure

The F-measure is a performance metric that measures the harmonic mean of precision and recall, imparting a single score that balances both precision and recall values. It is an extensively utilized metric in text feature extraction and deep learning to measure the efficiency of classification models. This is mathematically formulated as given below.

$$F - measure = 2 * \frac{Pre * Rec}{Pre + Rec} \quad (21)$$

From the above Equation (21), the performance metric '*F - measure*' is evaluated by employing the precision '*Pre*' and recall '*Rec*' results. **Table 4** given below lists the results of F-measure using four methods namely proposed LACDNN-FF, conventional DeepDoc, DCT-CompSegNet and CNN.

Table 4. F-measure results of proposed LACDNN-FF, existing DeepDoc, DCT-CompSegNet, and CNN.

Samples	F-measure			
	LACDNN-FF	DeepDoc	DCT-CompSegNet	CNN
500	0.983963	0.976479	0.972974	0.966442
1000	0.973407	0.956974	0.949974	0.936487
1500	0.9815	0.95	0.943996	0.933498
2000	0.9695	0.943999	0.933999	0.9265
2500	0.957896	0.935983	0.93	0.921498
3000	0.957983	0.9505	0.943999	0.9375
3500	0.971407	0.962896	0.956896	0.949933
4000	0.971991	0.958468	0.953983	0.945962
4500	0.960933	0.948991	0.93999	0.933478
5000	0.974991	0.962494	0.952493	0.94299

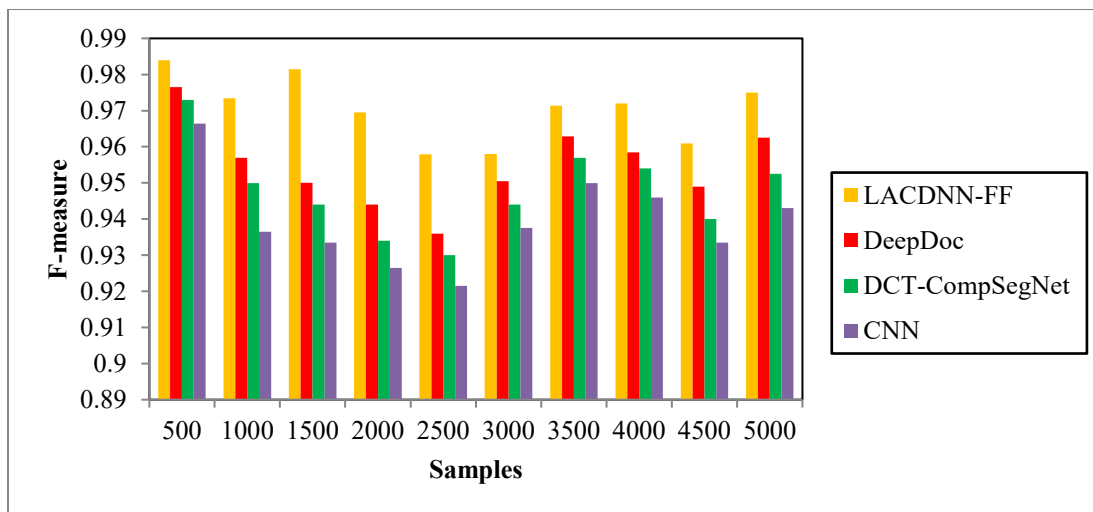


Figure 13. F-measure versus samples.

Figure 13 shows the results of F-measure for the proposed LACDNN-FF and three different existing methods, DeepDoc, DCT-CompSegNet and CNN. With the F-measure results generated based on the

harmonic mean of precision and recall, improvement in precision and recall therefore results in the improvement of F-measure also. From the above figure the F-measure results of proposed method were observed to be better than DeepDoc, DCT-CompSegNet and CNN. The reason was by applying the non-linear activation function of ML to linear activation function with deep neural network a transformation matrix was employed. Following which using transformation matrix, employing within class scatter matrix and between classes of scatter matrix via pixel-precise semantic segmentation segments accurately. Finally, the classified regions were then ascertained as headings, paragraphs, tables or images via different classes by LAA. This in turn therefore improved the f-measure results of LACDNN-FF method by 2% compared to DeepDoc, 2% compared to DCT-CompSegNet and 3% compared to CNN respectively.

4.2.3 Performance Analysis of Error Rate

Finally, in this section error rate involved in the process of text feature extraction is analyzed and validated. The error rate refers to the error involved during text feature extraction process. With 5000 different samples involved in the text feature extraction process and each sample possessing a different frequency of features, the error rate plays a predominant role. The error rate is defined as the text features wrongly extracted from the corresponding sample of each category. Here, in our work based on the accuracy, error rate is measure. With true positive, false positive, true negative and false negative of the proposed method being 460, 10, 27 and 3 respectively, the accuracy was observed to be 0.95, hence 0.05 is the error rate involved. Similarly, this is measured for the other three methods with respect to varied samples. **Table 5** below lists the error rate results for the proposed LACDNN-FF and he existing DeepDoc, DCT-CompSegNet and CNN methods.

Table 5. Error rate results of proposed LACDNN-FF, existing DeepDoc and DCT-CompSegNet and CNN.

Samples	Error rate (%)			
	LACDNN-FF	DeepDoc	DCT-CompSegNet	CNN
500	0.026	0.036	0.05	0.062
1000	0.03	0.05	0.062	0.067
1500	0.035	0.053	0.075	0.079
2000	0.038	0.057	0.078	0.08
2500	0.041	0.06	0.063	0.07
3000	0.043	0.062	0.067	0.072
3500	0.027	0.048	0.05	0.054
4000	0.033	0.05	0.053	0.057
4500	0.038	0.052	0.055	0.059
5000	0.03	0.036	0.04	0.047

Finally **Figure 14** shows the error rate results analyzed using the four methods, LACDNN-FF, DeepDoc, DCT-CompSegNet and CNN. From the above graph, although the error rate was found to increase with the increase in the samples, the comparative analysis showed better results using LACDNN-FF method than DeepDoc, DCT-CompSegNet and CNN. The reason behind the minimization of error rate using LACDNN-FF method was due to the application of Cosine CNN-BiLSTM neural network based text feature extraction algorithm. By applying this algorithm, in the input layer, quality enhanced compressed documents along with the accurately and precisely classified regions were provided as input. Next, feature maps were generated and activated using a sigmoid activation function. Finally, in the Bi-LSTM layer, text feature results were extracted via forward hidden state and backward hidden state. Using the forward and backward hidden state to extract text features in turn reduced the error rate of LACDNN-FF method by 31% compared to DeepDoc, 41% compared to DCT-CompSegNet and 47% compared to CNN.

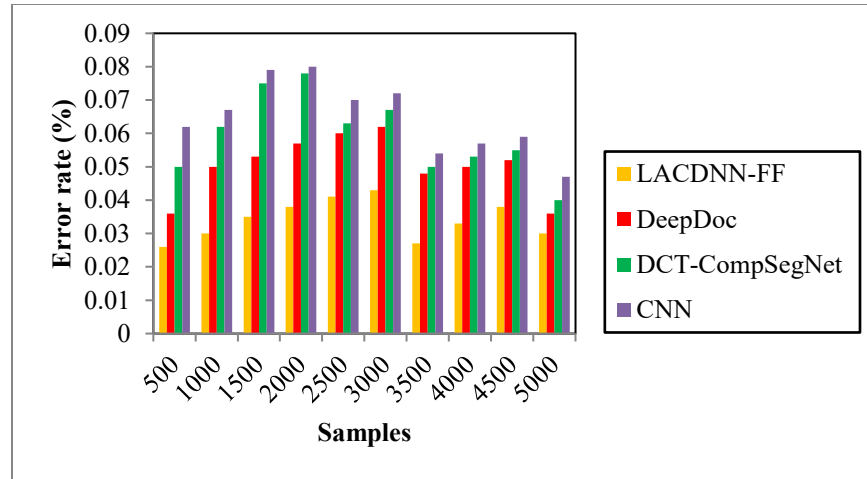


Figure 14. Error rate versus samples.

4.2.4 Performance Analysis of Time Complexity

It is measured as amount of time consumed by algorithm for detecting the image extracted as text sample.

It is mathematically computed as follows,

$$TC = \sum_{i=1}^n DS_i * time(IET) \quad (22)$$

From Equation (22), TC represents an execution time, ' n ' indicates the number of data samples, $time(IET)$ denotes a time consumed to image extract as text samples. It is measured in milliseconds (ms). **Table 6** below lists results of error rate using, LACDNN-FF, DeepDoc, DCT-CompSegNet and CNN.

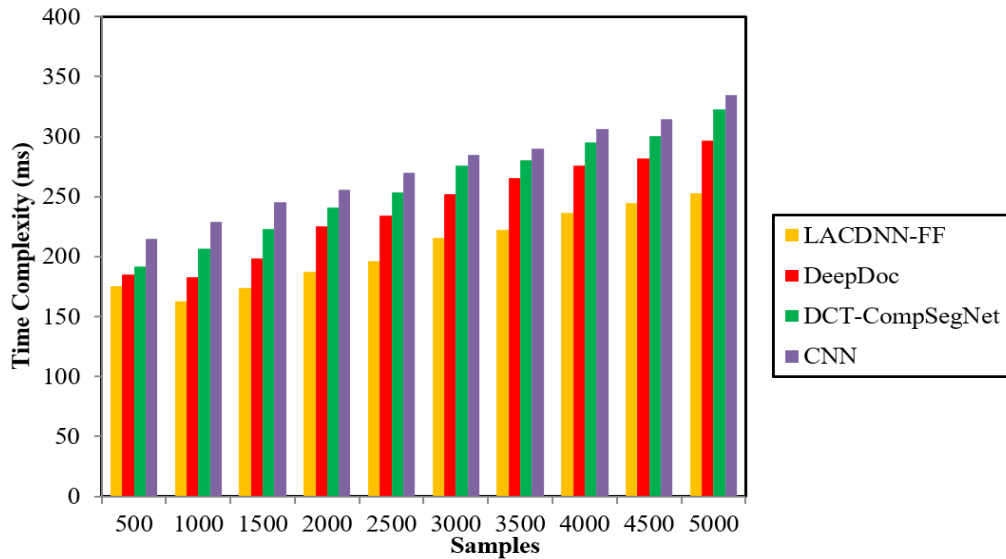


Figure 15. Time complexity versus samples.

Figure 15 illustrates the time complexity results for the proposed LACDNN-FF and the three different existing methods, DeepDoc, DCT-CompSegNet and CNN. From the above figure the time complexity results of proposed method were observed to be better DeepDoc, DCT-CompSegNet and CNN. The reason for the lesser time complexity is to apply Cosine CNN-BiLSTM neural network for extracting the text features in LACDNN-FF. The time complexity is reduced using the LACDNN-FF method by 15% compared to DeepDoc, 21% compared to DCT-CompSegNet and 24% compared to CNN respectively.

Table 6. Time complexity results of proposed LACDNN-FF, existing DeepDoc, DCT-CompSegNet and CNN.

Samples	Time complexity (ms)			
	LACDNN-FF	DeepDoc	DCT-CompSegNet	CNN
500	155	175	192	215
1000	162.5	182.5	206.2	228.7
1500	174	198.5	223	245
2000	186.9	225	241	256
2500	196	234	253.3	269.9
3000	215.6	252	275.6	284.8
3500	222	265.5	280.3	290.3
4000	236.7	275.6	295.2	306.5
4500	244.5	282	300.2	314.6
5000	253	296.6	322.5	334.8

4.2.5 L2 Loss Function

L2 Loss Function is employed to minimize the error. L2 Loss Function is estimated as the sum of the all the squared difference among the true value and the predicted value. This is mathematically evaluated as given below.

$$L2 \text{ Loss Function} = \sum_{i=1}^n (y_{true} - y_{predicted})^2 \quad (23)$$

From the above Equation (23), L2 Loss Function is evaluated. **Table 7** below provides the L2 Loss Function values of proposed method LACDNN-FF, and the state-of-the-art methods, DeepDoc, DCT-CompSegNet and CNN respectively.

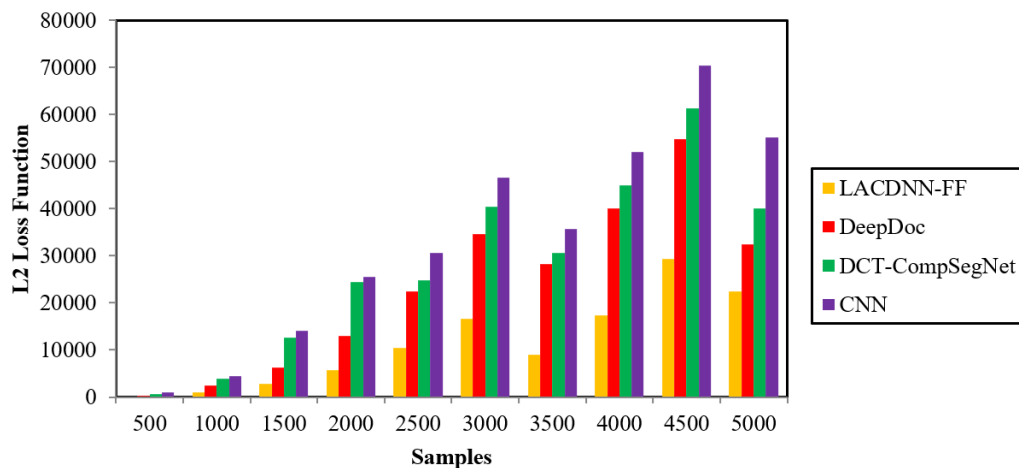


Figure 16. L2 Loss function versus samples.

Figure 16 demonstrates the results of L2 Loss Function for the proposed LACDNN-FF and three different existing methods, DeepDoc, DCT-CompSegNet and CNN. From the above figure the L2 Loss Function results of proposed method were observed to be better than other methods. The reason for lesser L2 Loss Function because of the use of Cosine CNN-BiLSTM neural network for extracting the text features in LACDNN-FF. L2 Loss Function is minimized using LACDNN-FF method by 53% compared to DeepDoc, 65% compared to DCT-CompSegNet and 71% compared to CNN.

Table 7. L2 loss function results of proposed LACDNN-FF, existing DeepDoc, DCT-CompSegNet and CNN.

Samples	L2 Loss function			
	LACDNN-FF	DeepDoc	DCT-CompSegNet	CNN
500	169	324	625	961
1000	900	2500	3844	4489
1500	2756.25	6320.25	12656.25	14042.25
2000	5776	12996	24336	25600
2500	10506.25	22500	24806.25	30625
3000	16641	34596	40401	46656
3500	8930.25	28224	30625	35721
4000	17424	40000	44944	51984
4500	29241	54756	61256.25	70490.25
5000	22500	32400	40000	55225

4.2.6 Confusion Matrix

A confusion matrix is a valuable tool in classification for evaluating the performance of the proposed LACDNN-FF method. It summarizes the prediction results by comparing the actual (true) labels with the predicted labels generated by the model.

Table 8. Confusion metrics using proposed LACDNN-FF method.

Predicted value	Actual positive	Actual negative	Total
Positive	TP = 4600	FP = 100	4700
Negative	FN = 30	TN = 270	300
Total	4630	370	5000

As shown in **Table 8**, the confusion matrix of the proposed LACDNN-FF method is presented using 5000 samples. The analysis of the results shows that the LACDNN-FF method achieved 4600 true positives (TP) and 270 true negatives (TN), with 100 false positives (FP) and 30 false negatives (FN).

4.2.7 ANOVA with Pairwise-t-Test

ANOVA (Analysis of Variance) with Pairwise-t-test is a statistical test employed to evaluate whether there is a major dissimilarity among multiple groups within a dataset. In the context of quality improved compressed image documents, ANOVA is applied with Pairwise-t-test to assess whether the mean values of a dependent variable (accuracy) differ significantly across different groups defined by a categorical variable (various image compression methods). Several methods such as proposed LACDNN-FF, existing DeepDoc, DCT-CompSegNet and CNN are used for performing the ANOVA with Pairwise-t-test by using accuracy parameter.

Figures 17 and **18** demonstrates the ANOVA with Pairwise-t-test results. In these figures, the x-axis denotes the models such as proposed LACDNN-FF and the existing DeepDoc, DCT-CompSegNet and

CNN, and the y-axis indicates the performance metrics (accuracy). The Pairwise-t-test is employed as a post-hoc test to determine which specific techniques (Proposed versus DeepDoc) differ significantly from one another. The P-value is derived from the F-statistic. A P-value is lower than 0.05 (the significance level) indicates statistically significant difference in the accurate feature extraction. Finally, the proposed LACDNN-FF method performs significantly better than existing methods.

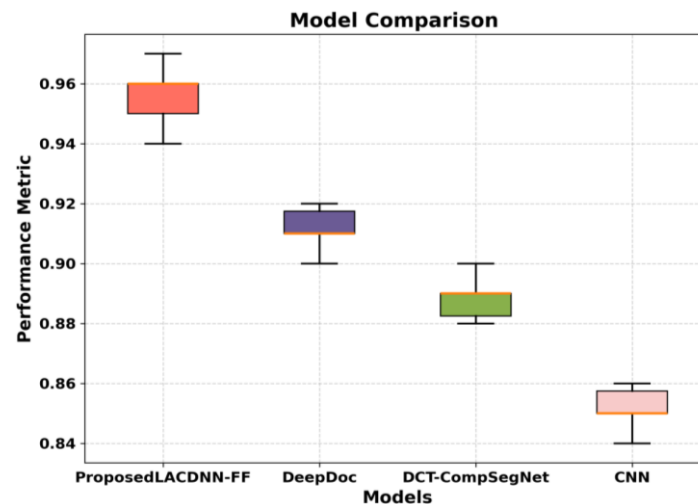


Figure 17. ANOVA with Pairwise-t-test results for proposed LACDNN-FF, existing DeepDoc, DCT-CompSegNet and CNN.

```

=== ANOVA Test ===
F-statistic: 309.263158
p-value: 0.000000
Result: Significant difference between models

=== Pairwise T-tests (Post-hoc) ===
ProposedLACDNN-FF vs DeepDoc: t=12.1034, p=0.000000
ProposedLACDNN-FF vs DCT-CompSegNet: t=17.8920, p=0.000000
ProposedLACDNN-FF vs CNN: t=27.8904, p=0.000000
DeepDoc vs DCT-CompSegNet: t=6.6670, p=0.000003
DeepDoc vs CNN: t=18.1827, p=0.000000
DCT-CompSegNet vs CNN: t=11.5157, p=0.000000

```

Figure 18. ANOVA with Pairwise-t-test results with F-statistic and p-value.

5. Conclusion and Future Work

Text feature extraction from quality enhanced compressed images is a demanding task owing to the diverse categories of entities involved like, text, image, table and so on. In this paper, we propose a deep learning method Linear Activated Cosine Deep Neural Network based text feature extraction from Fuzzy Filter (LACDNN-FF) quality improved compressed image documents. LACDNN-FF adopts a Two Dimension Flexible Fuzzy Filter-based Pre-processing via image enhancement agent to enhance quality and therefore

eliminating error extensively. To analyze layouts, Linear Activated Discriminant Analysis with Deep Neural Network-based Segmentation is applied via layout analysis agent, therefore improving precision and accuracy in complex categories. LACDNN-FF then efficiently uses the Cosine CNN-BiLSTM neural networks based Text Feature Extraction via text feature extraction agent to extract text feature vectors of input samples, thereby better capturing global information. Experiments on the scanned images dataset show that the LACDNN-FF method outperforms all baselines in terms of precision, recall, accuracy, F-measure and error rate. In future work, the proposed method will be extended by applying novel deep learning and optimization techniques for quality improved compressed image documents. Also, several metrics such as specificity and space complexity will be considered for feature extraction performance.

Conflicts of Interest

The authors confirm that there is no conflict of interest to declare for this publication.

Acknowledgments

None.

AI Disclosure

The author(s) declare that no assistance is taken from generative AI to write this article.

References

- Al-Khafaji, M., & Ramaha, N.T.A. (2025). Hybrid deep learning architecture for scalable and high-quality image compression. *Scientific Reports*, 15(1), 22926. <https://doi.org/10.1038/s41598-025-06481-0>
- Bisen, T., Javed, M., Kirtania, S., & Nagabhushan, P. (2023a). DWT-CompCNN: deep image classification network for high throughput JPEG 2000 compressed documents. *Pattern Analysis and Applications*, 26(4), 1641-1655. <https://doi.org/10.1007/s10044-023-01190-8>
- Bisen, T., Javed, M., Nagabhushan, P., & Watanabe, O. (2023b). Segmentation-less extraction of text and non-text regions from jpeg 2000 compressed document images through partial and intelligent decompression. *IEEE Access*, 11, 20673-20687. <https://doi.org/10.1109/ACCESS.2023.3249961>
- Choudhari, R.M., & Dakhane, D.M. (2018). Information extraction among symbolically compressed image in database. *International Journal of Creative Research Thoughts*, 6(1), 746-753.
- Du, J., Wang, D., Lin, B., He, L., Huang, L.-C., Wang, J., Manion, F.J., Li, Y., Cossrow, N., & Yao, L. (2025). Use of deep learning-based NLP models for full-text data elements extraction for systematic literature review tasks. *Scientific Reports*, 15(1), 19379. <https://doi.org/10.1038/s41598-025-03979-5>
- Golla, S., Sujatha, B., & Sumalatha, L. (2024). TIE-text information extraction from natural scene images using SVM. *Measurement: Sensors*, 33, 101018. <https://doi.org/10.1016/j.measen.2023.101018>
- Horadi, K.V., Pujari, D.J., & Bhat, N.P. (2022). Deep learning based model for table detection content and layout analysis in compressed document images-a comprehensive approach. *Journal of Theoretical and Applied Information Technology*, 100(14), 5212-5222.
- Ji, X., Yang, X., Yue, Z., Yang, H., & Zheng, B. (2025). Deep learning image compression method based on efficient channel-time attention module. *Scientific Reports*, 15(1), 15678. <https://doi.org/10.1038/s41598-025-00566-6>
- Joodi, M.A., Saleh, M.H., & Kadhim, D.J. (2023). A new proposed hybrid learning approach with features for extraction of image classification. *Journal of Robotics*, 2023(1), 9961421. <https://doi.org/10.1155/2023/9961421>
- Li, C. (2021). Research on methods of English text detection and recognition based on neural network detection model. *Scientific Programming*, 2021(1), 6406856. <https://doi.org/10.1155/2021/6406856>

- Li, D.-L., Lee, S.-K., & Liu, Y.-T. (2025). Printed document layout analysis and optical character recognition system based on deep learning. *Scientific Reports*, 15(1), 23761. <https://doi.org/10.1038/s41598-025-07439-y>
- Liang, Y., Zhao, M., Liu, X., Jiang, J., Lu, G., & Jia, T. (2024). An adaptive image compression algorithm based on joint clustering algorithm and deep learning. *IET Image Processing*, 18(3), 829-837. <https://doi.org/10.1049/ipr2.13021>
- Liu, H., Wang, H., Bai, J., Lu, Y., & Long, S. (2022). DeepSSR: a deep learning system for structured recognition of text images from unstructured paper-based medical reports. *Annals of Translational Medicine*, 10(13), 740. <https://doi.org/10.21037/atm-21-6672>
- Mahadevkar, S.V., Patil, S., Kotecha, K., Soong, L.W., & Choudhury, T. (2024). Exploring AI-driven approaches for unstructured document analysis and future horizons. *Journal of Big Data*, 11(1), 92. <https://doi.org/10.1186/s40537-024-00948-z>
- Malashin, I., Masich, I., Tynchenko, V., Gantimurov, A., Nelyub, V., & Borodulin, A. (2024). Image text extraction and natural language processing of unstructured data from medical reports. *Machine Learning and Knowledge Extraction*, 6(2), 1361-1377. <https://doi.org/10.3390/make6020064>
- Natei, K.N., Viradiya, J., & Sasikumar, S. (2018). Extracting text from image document and displaying its related information. *Journal of Engineering Research and Application*, 8(5), 27-33.
- Park, J., Seo, W., & Yun, T.S. (2025). End-to-end data extraction framework from unstructured geotechnical investigation reports via integrated deep learning and text mining techniques. *Developments in the Built Environment*, 23, 100733. <https://doi.org/10.1016/j.dibe.2025.100733>
- Rajan, R., & Devasena, M.S.G. (2025). Deep learning based optimization model for document layout and text recognition. *Ain Shams Engineering Journal*, 16(10), 103587. <https://doi.org/10.1016/j.asej.2025.103587>
- Rajesh, B., Javed, M., & Nagabhushan, P. (2020). Automatic tracing and extraction of text-line and word segments directly in JPEG compressed document images. *IET Image Processing*, 14(9), 1909-1919. <https://doi.org/10.1049/iet-ipr.2019.1437>
- Rajesh, B., Zaman, S.M., Javed, M., & Lin, M. (2024). DCT-CompSegNet: fast layout segmentation in DCT compressed JPEG document images using deep feature learning. *Multimedia Tools and Applications*, 83(25), 66201-66221. <https://doi.org/10.1007/s11042-024-18204-0>
- Rautaray, J., Panigrahi, S., Nayak, A.K., Sahu, P., & Mishra, K. (2025). Deep learning based feature extraction technique for single document summarization using hybrid optimization technique. *IEEE Access*, 13, 24515-24529. <https://doi.org/10.1109/ACCESS.2025.3538169>
- Rombach, A.M., & Fettke, P. (2025). Deep learning based key information extraction from business documents: systematic literature review. *ACM Computing Surveys*, 58(2), 1-37. <https://doi.org/10.1145/3749369>
- Sharma, M., Saini, H.C., Dutta, R., Chaudhary, S., Agarwal, S., & Rajawat, V.S. (2024). Enhancing low-resolution text recognition with advanced deep learning models. In *Proceedings of the 6th International Conference on Information Management & Machine Intelligence* (pp. 1-6). Association for Computing Machinery. New York, USA. <https://doi.org/10.1145/3745812.3745856>
- Sinhuja, M., Padubidri, C.G., Jayachandra, G.S., Teja, M.C., & Kumar, G.S.P. (2024). Extraction of text from images using deep learning. *Procedia Computer Science*, 235, 789-798. <https://doi.org/10.1016/j.procs.2024.04.075>
- Suvradeep (2026). Scanned images dataset for OCR and VLM finetuning. *Kaggle*. <https://www.kaggle.com/datasets/suvroo/scanned-images-dataset-for-ocr-and-vlm-finetuning>
- Yang, W., & Yu, M. (2025). Automatic extraction of PDF table data based on deep learning. In *Proceedings of the 2025 4th International Conference on Big Data, Information and Computer Network* (pp. 222-226). Association for Computing Machinery. New York, USA. <https://doi.org/10.1145/3727353.3727391>

Zhou, L., Liu, C., Yadav, A., Azam, S., & Karim, A. (2024). An image quality assessment method based on edge extraction and singular value for blurriness. *Machine Vision and Applications*, 35(3), 37. <https://doi.org/10.1007/s00138-024-01522-6>

Original content of this work is copyright © Ram Arti Publishers. Uses under the Creative Commons Attribution 4.0 International (CC BY 4.0) license at <https://creativecommons.org/licenses/by/4.0/>

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.