

ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
<http://dx.doi.org/10.70102/afts.2026.1836.713>

INTELLIGENT FINANCIAL DOCUMENT DIGITISATION AND CLASSIFICATION VIA A HYBRID ATTENTION-GUIDED DUAL PATH CNN WITH ATOMIC ORBITAL SEARCH OPTIMISATION

Archana Lopes¹, Kolla Bhanu Prakash^{2*}, Yu-Chen Hu³

¹Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India; Assistant Professor, Department of Electronics and Computer Science, Fr. Conceicao Rodrigues College of Engineering, Bandstand, Bandra West, Mumbai, India. e-mail: archana_lopes@frcrce.ac.in,
orcid: <https://orcid.org/0000-0002-8454-9062>

^{2*}Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, Andhra Pradesh, India. e-mail: drkbp@kluniversity.in,
orcid: <https://orcid.org/0000-0002-7955-2777>

³Professor, Department of Computer Science, Tunghai University, Taichung, Taiwan.
e-mail: ychu@thu.edu.tw, orcid: <https://orcid.org/0000-0002-5055-3645>

Received: June 01, 2026; Revised: July 13, 2026; Accepted: August 21, 2026; Published: September 29, 2026

SUMMARY

The financial institution uses different types of documents, including bank statements, cheques, income tax forms, salary slips, and electricity bills. Current approaches deal with either the preprocessing stage, text and feature extraction, or even classification separately and fail to perform efficiently on low-quality, handwritten, and visually similar documents. The proposed method, HAPODCNN, which stands for Hybrid Top-Down Attention Pyramidal Atomic Orbital Dual-Path Convolutional Neural Network, is used for digitization and classification of financial documents in an end-to-end approach. The architecture includes MGSWF-GS as a preprocessing technique for image improvement, JA-ViT-UNet as a text and feature extraction technique, and a bidirectional pyramidal dual-path CNN along with spatial and channel attention for classification. Atomic Orbital Search is employed for model parameter tuning. The experiments were performed on 414 real financial documents in India with 5 categories. First, the 414 original images were partitioned into three sets: 70% of the images for training, 10% for validation, and 20% for testing. The data augmentation process was then performed on the training set only to increase the number of training data from 414 to 1,170. The suggested HAPODCNN yields 98.93% classification accuracy with 97.34% precision, recall, and F1 score, and consumes 18.7 million parameters and 3.42 GFLOPs. The model also reports 98.30% ROC-AUC, $3.2 \pm 0.1\%$ character error rate, $6.5 \pm 0.2\%$ word error rate, and 1.12 s latency. The suggested model ensures 96.1% accuracy even under severe document degradations, which implies its high robustness. As seen from the obtained results, the combination of degradation-aware preprocessing, transformer-based feature extraction, bidirectional attention-based classification, and AOS allows improving recognition and classification accuracy, which makes the suggested framework applicable to automated financial document processing tasks.

Key words: financial document digitization, document classification, vision transformer, convolutional neural network, attention mechanism, atomic orbital search.

INTRODUCTION

There is an increasing trend in financial institutions where automatic document processing is relied on due to the high number of documents that these organizations handle in form of bank statements, tax documents, invoices, payslips, cheques, and utility documents. Manual processing of these documents takes more time and is prone to errors, making automatic recognition significant in improving financial information handling, extraction, and management [2][7]. In similar digitization efforts, the importance of converting paper-based documents into machine-readable formats has been shown in order to efficiently preserve and extract information [1][8]. Besides the financial sector, there have been similar gains shown in the application of natural language processing in the same regard [3].

Financial documents often hold personal, transactional, and institutional data that necessitates extra measures to ensure confidentiality and security in the process of processing such documents. Deep learning methods were therefore explored for detecting and avoiding information leakage in digital documents [4]. Meanwhile, the anonymization of financial documents used neural language models in order to keep confidential data safe [10]. In general, the application of digital technologies is increasingly growing in order to enhance the efficiency and decrease the costs of organizational activities [11]. The successful implementation of technological systems relies on the organizational practices and the capability of organizations to incorporate new digital processes into their operational environment [19].

Document classification is yet another significant element in automated information management. There are clustering algorithms designed to sort a large set of documents, while there are deep learning techniques showing high efficiency in classifying technical documents [6]. The classification of financial transactions has also been assisted by machine learning techniques using text features and character-based representations [7]. Nevertheless, the problem of document classification significantly increases in complexity if several classes share common textual or graphical characteristics. Handwritten and deteriorated documents bring even more complications since degraded characters, overlapped strokes, noise, and structure variations may negatively affect the classification process [9][17].

However, traditional systems face challenges in integrating low-level textual information with the higher-level semantic structure of the document. This is why recent developments have been made in multimodal understanding of financial documents, deep feature text recognition, image enhancement, hybrid recurrent-convolution architecture [16], and visual recognition using convolutional neural networks [18]. Methods have also been developed for structured extraction of tables in heterogeneous documents [20]. Still, an integrated system, which takes care of degradation, feature extraction, semantic understanding, and classification at once, is required. This is the motivation for proposing this paper titled “Hybrid Top-Down Attention Pyramidal Atomic Orbital Dual Path Convolutional Neural Network for Financial Document Extraction and Classification.

The proposed framework offers the following contributions:

1. Proposes JA-ViT-UNet with jump-attention guided by the ViT for combining local features from the encoder and global context information for better financial document text extraction.
2. Proposes HAPODCNN with a bidirectional flow of features at the bottom-up and top-down levels through spatial and channel attention for better financial document classification.
3. Combines MGSWF-GS, JA-ViT-UNet, dual-path classification, and AOS into a complete end-to-end system for varied Indian financial documents.

The rest of the paper is organized as follows. Section 2 provides related work regarding the digitization, recognition, and classification of documents. Section 3 introduces the HAPODCNN model, which includes the pre-processing step, JA-ViT-UNet feature extraction, dual path classifier, AOS algorithm, complexity analysis, and evaluation criteria. Section 4 explains the experimental results and their comparison. Section 5 concludes the paper and its key limitations, whereas Section 6 provides the conclusion and directions for future research.

LITERATURE SURVEY

Previous research works concerning document digitization were mostly devoted to character recognition, archiving, information extraction, and automated document conversion. The efficient OCR technology proved the possibility of historical document digitization using a small amount of training data [1]. Deep learning technologies were used for ancient Devanagari character recognition, whereas CNN transfer-learning approaches improved Devanagari handwriting digitization [17]. In finance, automatic recognition systems were implemented for recognition of financial instruments, and financial transaction classification based on character-level word embeddings was performed [7]. Digitization of corporate annual reports is one more example of the necessity for systematic document processing and cataloging [8]. It proves the relevance of automated document processing but does not solve the problem of financial document classification.

Information security and understanding of the contents have also been considered. Methods of preventing loss of information using deep learning can detect the presence of confidential information in multi-page digital documents [4]. In addition, neural language models have been used to anonymize financial documents by using contextual representation [10]. Automation of text analysis has extended beyond the field of finance by proving the ability of NLP systems to find important documentation among unstructured ones [3]. Text clustering has been enhanced by employing an optimization-based approach, which clusters large numbers of documents based on their content, whereas technical document classification using a deep learning approach has also been studied [5][6].

Several visual and multimodal document-processing models have attempted to overcome these drawbacks. A hybrid approach for the processing of unstructured financial documents based on robotic process automation and multimodal learning by capturing both textual and layout information was introduced [12]. Deep feature networks were used for recognition of texts in ID cards, while dual discriminator generative adversarial networks were utilized to improve image binarization for difficult visual situations [15]. Hybrid deep models that incorporate bi-directional gated recurrent units together with 2D CNNs have shown the value of the joint modeling of sequential and local visual features [16]. Also, CNN-based recognition has been used effectively in visually structured objects like hand-drawn electrical and electronic devices [18]. Moreover, an end-to-end model of table structure recognition can be utilized for recognizing structured tables in heterogeneous documents [20]. However, these approaches usually target each individual phase in the pipeline, such as preprocessing, detection, recognition, or extraction of the information [13].

Modern computational research is characterized by an increased focus on integrating modeling and optimization and making efficient use of technology. In technology-oriented research, it has been demonstrated that novel systems can make a contribution to decreasing costs and increasing efficiency [11]. Computational modeling has also been utilized in representing complex interactions between entities, thus proving the added benefits of structured computational modeling [14]. In organizational research, research on multinational organizations has revealed the need to choose and configure interconnected practices in line with operational needs [19]. While these works do not specifically examine financial document recognition, they still provide additional context for designing digital systems.

Some of the problems found in the literature include the following. Methods related to OCR and handwriting recognition concentrate more on text extraction, and methods that are security-based mostly consider confidentiality and anonymization. Classification and clustering techniques have largely considered textual features, but those related to multimodal, CNN, GAN, and extraction mostly consider solving each component of the process independently. Thus, there is a need for an architecture that can improve the quality of the degraded financial documents, extract their discriminatory local and global features, differentiate visually similar categories of the documents, and tune the classification parameters in one go. HAPODCNN solves all these problems by considering MGSWF-GS preprocessing, JA-ViT-UNet feature extraction, pyramidal dual-path classification, attention, and Atomic Orbital Search optimization in an end-to-end financial-document processing architecture.

THE PROPOSED METHOD

In the framework, financial document image data has been used with extra pre-processing to increase the quality of information content. In this research work, an innovative U-shaped network model with jump attention vision transformer (JAVIT) is proposed, in which features of the U-shaped network, attention network, and vision transformer have been integrated for enhanced text extraction from document images. Additional Atomic Orbital Search-based Optimization has been incorporated into the HAPODCNN to enhance the financial document classifier.

Input Acquisition

The input financial documents are collected from the financial image document dataset. This dataset contains personal financial and identification documents for individuals in India. The dataset comprises images of various financial documents, including bank statements, salary slips, income tax returns, utility bills, and cheques.

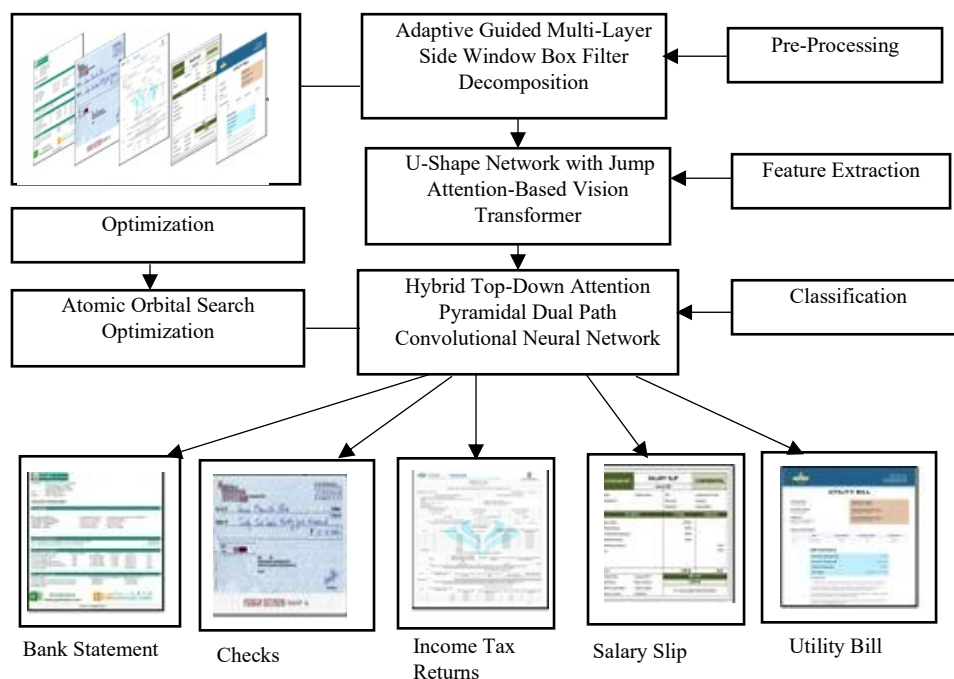


Figure 1. Proposed model architecture

Figure 1 shows the entire workflow of the proposed HAPODCNN framework for the classification of financial documents. It begins with a set of images of financial documents including bank statements, checks, income tax returns, salary slips, and utility bills. It has been introduced as a preprocessing stage in the first block by making use of an adaptive guided multilevel side-window box filter to increase image quality, retain the information contained, and separate it from noise. Specifically, extract features from preprocessed images through a U-shape network with an attention-jumper vision transformer to detect the structures and distance-based relations. The proposed HAPODCNN is utilized in this case to obtain a better multi-level feature representation and decision process. Besides, tune the parameters of the neural network with the help of Atomic Orbital Search optimization. Ultimately, the system classifies all documents into five classes (bank statements, checks, income tax returns, salary slips, and utilities), which can then be used in automatic document management systems.

Adaptive Guided Multi-layer Gaussian Side Window Box Filter (MGSWF) with Strengthening Decomposition-Based Pre-Processing

The MGSWF preprocessing step divides financial document images into smooth and detail-conserving elements such that delicate text, edges, writing, and structural details are preserved before feature extraction. In contrast to traditional approaches to decomposition that deal with images in a uniform

way, MGSWF takes into account image complexity by applying Gaussian smoothing and side-window filtering. The Gaussian-smoothed form at scale s can be obtained as (Equation 1)

$$L_s(x, y) = I(x, y) * G_s(x, y) \quad (1)$$

where $I(x, y)$ is the input image, $G_s(x, y)$ is the Gaussian kernel, and $*$ denotes convolution. To preserve edge information, side-window filtering calculates the local response only from pixels located on the same side of an edge:

$$S(p) = \frac{1}{|W_s(p)|} \sum_{q \in W_s(p)} I(q) \quad (2)$$

(Equation 2) where $W_s(p)$ denotes the selected side-window region around pixel p . The image is then represented through multiple smooth and detail layers, with each detail component expressed as (Equation 3)

$$D_i(x, y) = L_{i-1}(x, y) - L_i(x, y) \quad (3)$$

These detail sub-layers preserve text boundaries, thin lines, and local variations in intensity. MGSWF technique is used mostly on images that have high complexity, whereas MGSWF-GS works on the decomposition process through guided strengthening of images with low complexity. The process of strengthening helps in improving the visibility of the degraded document areas for feature extraction and classification tasks.

U-Shape Network With Jump Attention-Based Vision Transformer-Based Feature Extraction

The proposed method integrates ideas from U-Net, attention, and vision transformers to extract text from the document image. The U-Net has an encoder (a contracting path) and a decoder (an expansive path) and is extensively used in segmentation tasks and can be modified for feature extraction tasks. In this case, the encoder extracts visual features, whereas the decoder locates text features. The pre-processed document image is fed as the input to the network. Figure 2 shows the structure diagram of the U-Net with jump attention vision transformer feature extraction.

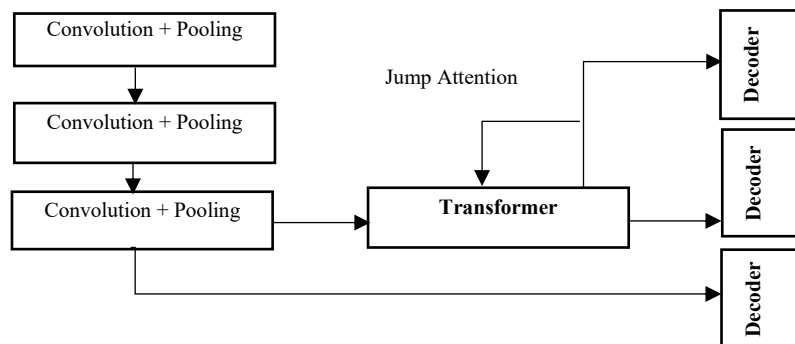


Figure 2: Architecture of U-shape network with jump attention-based vision transformer-based feature extraction

As the name suggests, the U-shaped encoder of the network progressively decreases the spatial dimensions of the document image by employing convolutions and pooling operations, as well as learning hierarchical features such as edges, textures, text strokes, and the overall visual appearance of the image. However, at the bridge, where the image features are broken down into patches and processed via a transformer-based attention mechanism, long-range dependencies are learned. Subsequently, the decoder increases the dimensions of the encoded feature map by applying upsampling operations along with skip connections with the encoder features.

Hybrid Top-Down Attention Pyramidal Dual Path Convolutional Neural Network

Hybrid top-down attention pyramidal dual-path convolutional neural network is used for financial

document classification, where low-level features and high-level semantic information are both utilized for document classification. This neural network has feature paths working from bottom-up and top-down approaches in different pyramid levels. The Spatial Attention (SA) helps in highlighting the informative parts of images, which are done through the use of (3×3) deconvolutional layer and sigmoid activation function to output the spatial attention masks. Such spatial attention masks help in highlighting the informative structural parts like text, signature, tables, and document borders. On the other hand, the Channel Attention (CA) helps in determining the most informative channels through the use of Global Average Pooling and fully connected layers with ReLU and sigmoid activations.

To further enhance the classification accuracy, Atomic Orbital Search (AOS) Optimization is performed on the neural network. AOS is a population-based optimization method which is motivated by the movement of electrons near an atomic nucleus and some concepts from quantum physics. In this method, the candidate values for the neural network parameters are assessed based on a fitness function and then the neural network parameters are optimized towards higher-performance values. The combination of AOS with the pyramidal dual path CNN will enhance both the parameter optimization and classification accuracy.

The pseudocode for AOS optimization is listed below.

Pseudocode for AOS Optimization

In the search space, determine the initial positions of solution candidates A_i by n candidates.

For the initial solution, candidates evaluate the fitness function F_i

Determine the Binding Energy (BE) (BE^P) and Binding State (BS) (BS^P) of the atom

In the atom, determine the candidate by Lowest Energy (LE) (LE^P) level

While Iteration < maximum number of iterations

 Number of imaginary layers m (generated)

 Imaginary layers (created)

 Sort the solution candidates in descending or ascending order by probability density function in the imaginary layers and distribute the solution candidates

for $P = 1 : m$

 Determine BE^P and BS^P of the P^{th} layer, and in the P^{th} layer determine the candidate with LE^P

for $i = 1 : w$

 Generate L, B, X, M

 Determine WR

if $L \geq WR$

if $F_i^P \geq BE^P$

$$A_{i+1}^P = A_i^P + \frac{L_i \times (B_i \times LE - M_i \times BS)}{P}$$

else if $F_i^P < BE^P$

$$A_{i+1}^P = A_i^P + B_i \times (X_i \times LE^P - M_i \times BS^P)$$

end

else if $L < WR$

$$A_{i+1}^P = A_i^P + M_i$$

end

end

end

Update (BE) and (BS) of atom

Update candidate with LE level in an atom

end while

end procedure

The AOS algorithm pseudocode starts by initializing the candidates with fitness values inside the search space. It calculates the binding energy, state of atoms and the lowest energy candidate solution. In subsequent iterations, candidate solutions are positioned in virtual layers based on probability densities and are updated based on generated search parameters. This is repeated until the maximum iteration is attained.

Theoretical Computational Complexity Analysis

The complexity of HAPODCNN depends on the processing, feature extraction, and classification modules of HAPODCNN. The complexity of MGSWF-GS is $O(LHWck^2)$, while the encoder-decoder of JA-ViT-UNet results in $O(HWC^2)$ complexity per stage. The complexity of multi-head self-attention is $O(N^2d)$, whereas the jump-attention gate is $O(NH'W'C)$ per skip connection. On the other hand, the dual-path pyramidal classifier deals with four scales with complexity of $O(H_sW_sC_s)$. Overall, HAPODCNN has 3.42 GFLOPs and 18.7 million parameters, which consists of around 11M in the JA-ViT-UNet, 7M for the classifier, and 0.7M for the preprocessing and attention modules.

Performance Evaluation Metrics

Performance of the proposed HAPODCNN model will be based on metrics such as accuracy, precision, recall, specificity, F1 Score, CER, WER, MAE, MSE, RMSE, and ROC-AUC, similar to those used for assessment of models presented in the manuscript.

Accuracy refers to the percentage of correctly classified samples (Equation 4):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (4)$$

Precision measures the proportion of correctly predicted positive samples (Equation 5):

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (5)$$

Recall or sensitivity measures the ability to correctly identify positive samples (Equation 6):

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (6)$$

Specificity measures the proportion of correctly identified negative samples (Equation 7):

$$Specificity = \frac{TN}{TN+FP} \times 100 \quad (7)$$

F1-score provides the harmonic mean of precision and recall (Equation 8):

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

For OCR/text extraction, the paper evaluates CER and WER (Equations 9 and 10).

$$CER = \frac{S_c + D_c + I_c}{N_c} \times 100 \quad (9)$$

$$WER = \frac{S_w + D_w + I_w}{N_w} \times 100 \quad (10)$$

where S , D , and I denote substitutions, deletions, and insertions, while N represents the total number of ground-truth characters or words.

The error measures used in the paper can be expressed as (Equations 11 – 13):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

EXPERIMENTAL RESULTS

Table 1. Ablation study on component contribution

Components	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Avg. Time(s)
Baseline — U-shape network only	88.5	87.0	86.5	86.8	1.80
Baseline and Pre-processing (MGSWF-GS)	91.2	90.1	90.5	90.3	1.95
Baseline and Jump Attention-based Vision Transformer	94.6	93.7	94.0	93.9	2.40
Baseline Jump Attention-based Vision Transformer (replace U-shape-only) + Top-down Attention Pyramidal Dual-Path CNN (classification head; no AOS)	96.1	95.4	95.6	95.5	2.20
Baseline + JA-ViT-UNet only (no Pyramidal CNN, no AOS)	93.2	92.1	92.6	92.3	2.25
Baseline + Pyramidal Dual-Path CNN only (no JA-ViT-UNet, no AOS)	92.4	91.6	91.9	91.7	2.05
MGSWF-GS + JA-ViT-UNet + Pyramidal CNN (no AOS)	96.8	96.0	96.2	96.1	2.10
Full model with AOS optimization (HAPODCNN)	98.93	97.34	97.34	97.34	1.12

Experimental investigation is done on the performance of HAPODCNN based on the measures of accuracy, precision, recall, F1-score, and computational time in comparison with MLP, RNN, RNN-CRF, and flairNER. An ablation study is performed on the role of the MGSWF-GS preprocessing,

JA-ViT-UNet, top-down attention pyramidal dual-path CNN, and AOS. Table 1 shows the performance results of each combination of modules in the measures of accuracy, precision, recall, F1-score, and average computational time. The baseline performance for U-shape reaches 88.5%, whereas preprocessing enhances recall by almost 4%. Accuracy reaches 94.6% thanks to JA-ViT-UNet, and the pyramidal dual-path CNN enhances multiscale discrimination. The final AOS-based HAPODCNN reaches an accuracy of 98.93% along with a precision, recall, and F1-score of 97.34%.

Implementation Details

The proposed model is implemented through the open-source TensorFlow framework on the Intel Core i5-4570T processor working at 2.90 GHz. In order to optimize parameters, use the Atomic Orbital Search optimization technique with a momentum factor equal to 0.6. The initial learning rate is chosen to be 0.01, and the model is trained for 200 epochs. The 414 initial original records were randomly partitioned following a strict 70% training, 10% validation, and 20% testing split. Consequently, the original dataset comprises exactly 290 training images, 41 validation images, and 83 test images before any augmentation. The validation and test sets remained isolated from the model training process and were used for hyperparameters tuning and performance assessment, correspondingly. The data augmentation technique was employed only to the final training set due to concerns of information leakage across training, validation, and test samples.

Dataset Descriptions

In the financial document images dataset [21], there are a total of 414 unique real-world documents collected from India and grouped into five classes, including bank statement, salary slip, ITR form, utility bill, and cheque. The class imbalance issue is handled using augmentation with $\pm 15^\circ$ rotation, horizontal flip, $\pm 20\%$ change in brightness and contrast, and Gaussian noise, but limited to the training set only. The total number of unique original documents was first divided into 70%, 10%, and 20% for the training, validation, and test sets, respectively. Data augmentation, including rotation within $\pm 15^\circ$, horizontal flip, brightness and contrast within $\pm 20\%$, and adding Gaussian noise, was applied only to the training set. The number of images in the training dataset reached 1,170 after applying data augmentation. None of the augmented validation and test documents were included in the training set. Table 2 below shows the different classes of documents, their description, and their image counts.

Table 2. Distribution of augmented training images across financial document classes

Document types	Descriptions	No. of augmented training images
Bank statement	A bank statement is a monthly record summarizing financial transactions within an account, including deposits, withdrawals, and balances	250
Check	A check is a written order instructing a bank to pay a specific amount from the account holder's funds to another party	300
ITR	Income tax returns are annual documents submitted to the government detailing an individual's earnings and taxes owed or refunded	120
Salary slips	A salary slip is a concise record provided by employers showing an employee's earnings, deductions, and net pay for a specific period.	240
Utility	A utility document is a bill or statement indicating charges for essential services like electricity, water, gas, or other amenities	260

Classification Outcomes of the Proposed Method for Classified Financial Document Images

Table 3 below demonstrates the classification results from the HAPODCNN model. The HAPODCNN achieves higher recall, precision, specificity, and F1-score on financial document classification (bank statement, check, ITR form, salary slip, and utility). Table 3 demonstrates the comparison of the classified results.

Table 3. Comparison of classified outcomes

The model is very accurate with respect to accuracy, recall, F1 Score, specificity, and precision in the case of bank statements. The model works effectively with high precision and slightly lower recall in case of checks. The precision in the case of ITR forms is high, while the recall is slightly less than precision. The model is very accurate, precise, and has a high F1 score in the case of salary slips, while the recall is slightly low. The model is effective for all metrics in the case of utility, while precision and recall are slightly less than accuracy.

Pre-processing

 BANDHAN BANK LIMITED BALANCE SHEET AS AT 31 MARCH, 2018			
	Schedule	As at 31 March 2017 (₹)	As at 31 March 2018 (₹)
Capital & Liabilities			
Capital	1	19,951,410,340	9,510,500,000
Reserves & Surplus	2	22,393,616,302	5,760,406
Provisions	3	320,887,400,485	-
Borrowings	4	30,334,476,135	-
Other liabilities and provisions	5	14,438,139,170	232,41,350
Total		199,137,320,851	9,548,676,796
Assets			
Cash and balances with Reserve Banks of India	6	8,103,885,512	101,068
Balance with Banks and Money at call and short notice	7	23,633,129,412	4,313,022,224
Investments	8	37,580,339,328	-
Advances	9	114,373,461,133	-
Other Assets	10	5,372,311,261	1,988,120,881
Other Assets	11	1,097,195,135	1,70,176,475
Total		199,137,320,851	9,548,676,796
Contingent liabilities			
Provision for collection	12	49,218,264	-
Significant Accounting Policies and Notes to Accounts			
Disclosures referred to above form an integral part of the Balance Sheet			
As per our report in even date			
Dr. P. K. Battacharya & Associates LLP Chartered Accountants Pan Registration Number : 101049R030004	For Bandhan Bank Limited		
Dr. / Per Anil Kakra Partner Membership Number : 94533	Dr. A. K. Lahiri Chairman	Sd/- Chandni Shekhar Ghosh Managing Director & CEO	Sd/- C.H. Deka Director
	Sd/- Bijoy Bandopadhyay Company Secretary	Sd/- Bijoy Bandopadhyay Chief Financial Officer	

(a) Bank Statement

		SALARY SLIP		EXCEL DATA PRO	
info@socialdatapara.com call: 9847 9847 9847		MAY	2018		
Employee Name:		Jacob Smith			
Employee Designation:		Clearing Staff			
Department:		General Affairs			
Total Days		31	Leave Taken	5	
Allowed Leave:		2	Days Worked:	28	
Description		Earning	Description	Deduction	
Basic Pay		\$5,000.00		Tax	
HRA		\$1,000.00		Insurance	
Core Allowance		\$300.00		Advance	
Medical Allowance		\$350.00	Loan Repayment	\$500.00	
Total		\$6,550.00	Total	\$1,850.00	
Net Pay		\$4,700.00	Four Thousand Seven Hundred Dollars		
Payment Details:					
Date of Transfer		01-06-2018	Account Holder Name		Jacob Smith
Bank Name		ABC BANK	Account Number		9876543210
Prepared By				Approved By	

(d) Salary Slip

Figure 3 (a – e): Pre-processed outcomes

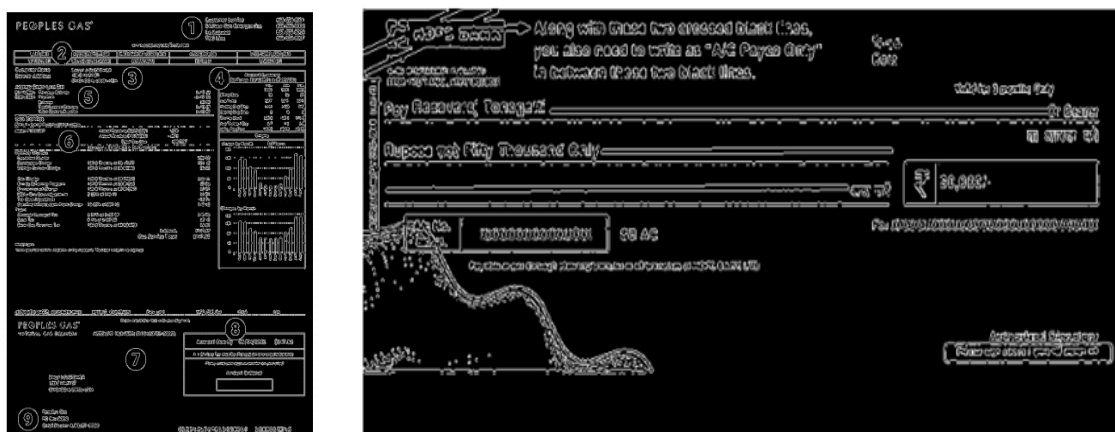
888

Bay Bridge of Bucks

Statement of Income

Address: 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900,

(c) Salary slip



(e) Bank check

OCR/Text Extraction Quality Analysis

Technical Institute Bijeljina, Archives for Technical Sciences. Year XVIII – N⁰ 36

Table 4. Text extraction quality analysis

Method	CER (%) ↓	WER (%) ↓
MLP (baseline OCR)	12.5 ± 0.8	21.3 ± 1.5
RNN	10.8 ± 0.6	19.4 ± 1.2
RNN-CRF	9.7 ± 0.5	18.1 ± 1.1
flairNER	11.4 ± 0.7	20.7 ± 1.3
LayoutLMv3	5.8 ± 0.3	10.6 ± 0.7
HAPODCNN	3.2 ± 0.1	6.5 ± 0.2

Performance Comparison With Existing Financial Document Classification Approaches

All baseline models as well as proposed models were compared based on the same split of the initial documents into 70% training, 10% validation, and 20% testing dataset. Data augmentation was used only for the training dataset, whereas the validation and testing datasets comprised exclusively of the initial, non-augmented data for a fair and leakage-free comparison. Publicly available pretrained weights of LayoutLMv3, DocFormer, and Donut were fine-tuned on the same training dataset. Grid search was used for hyperparameter tuning with the help of the validation dataset, and early stopping with a patience of 10 epochs was utilized for all models. Table 5 shows the final chosen hyperparameters for each model, such as optimizer, learning rate, batch size, epochs, and dropout. Figure 5(a-e) depicts the classification results of bank statements, ITR forms, salary slips, utility bills, and cheques.

Predicted : Bank Statement

Date	Description	Withdrawals	Deposits
01/06/19	Opening Balance		5,134.22
01/06/19	Salary		5,134.22
02/06/19	ATM Withdrawal	5,134.22	
03/06/19	ATM Withdrawal	5,134.22	
04/06/19	ATM Withdrawal	5,134.22	
05/06/19	ATM Withdrawal	5,134.22	
06/06/19	ATM Withdrawal	5,134.22	
07/06/19	ATM Withdrawal	5,134.22	
08/06/19	ATM Withdrawal	5,134.22	
09/06/19	ATM Withdrawal	5,134.22	
10/06/19	ATM Withdrawal	5,134.22	
11/06/19	ATM Withdrawal	5,134.22	
12/06/19	ATM Withdrawal	5,134.22	
13/06/19	ATM Withdrawal	5,134.22	
14/06/19	ATM Withdrawal	5,134.22	
15/06/19	ATM Withdrawal	5,134.22	
16/06/19	ATM Withdrawal	5,134.22	
17/06/19	ATM Withdrawal	5,134.22	
18/06/19	ATM Withdrawal	5,134.22	
19/06/19	ATM Withdrawal	5,134.22	
20/06/19	ATM Withdrawal	5,134.22	
21/06/19	ATM Withdrawal	5,134.22	
22/06/19	ATM Withdrawal	5,134.22	
23/06/19	ATM Withdrawal	5,134.22	
24/06/19	ATM Withdrawal	5,134.22	
25/06/19	ATM Withdrawal	5,134.22	
26/06/19	ATM Withdrawal	5,134.22	
27/06/19	ATM Withdrawal	5,134.22	
28/06/19	ATM Withdrawal	5,134.22	
29/06/19	ATM Withdrawal	5,134.22	
30/06/19	ATM Withdrawal	5,134.22	

(a) Bank statement

Predicted : ITR Form

(b) ITR forms

Predicted : Salary Slip

Particulars	Amount
Gross Salary	1,00,00.00
Basic Salary	80,00.00
Dearness Allowance	10,00.00
House Rent Allowance	5,00.00
Medical Allowance	2,00.00
Conveyance Allowance	1,00.00
Other Allowances	2,00.00
Total Allowances	17,00.00
Gross Salary	97,00.00
Deductions	10,00.00
Net Salary	87,00.00

(c) Salary slip

Predicted : Utility

Particulars	Amount
Gas Consumption	1,00,00.00
Basic Charge	5,00.00
Service Charge	2,00.00
Other Charges	1,00.00
Total Charges	18,00.00
Payment	18,00.00

(d) Utility

Predicted : Check

(e) Bank check

Figure 5(a-e). Classification outcomes

Table 5: Hyperparameter settings

Model	Optimizer	LR	Batch Size	Epochs	Dropout
MLP	Adam	0.001	64	100	0.4
RNN	Adam	0.0005	32	120	0.3
RNN-CRF	Adam	0.01	32	120	0.4
flairNER	Adam	0.001	32	150	0.5
LayoutLMv3	AdamW	5e-5	16	20	0.1
DocFormer	AdamW	3e-5	16	20	0.1
Donut	AdamW	2e-5	8	15	0.1

This section presents the performance analysis of the proposed HAPODCNN architecture. The proposed system evaluates various parameters in comparison with the existing ones for the purpose of classifying financial documents, including the precision, recall, F1 score, and computation time. Furthermore, the proposed approach is evaluated through the use of several measures (MAE, MSE, RMSE, and accuracy) in varying conditions of optimization, learning rate, and batch size.

Accuracy Analysis of the Proposed Method With and Without Optimization

Figure 6, there is a comparison of the performance accuracy of the HAPODCNN model before and after optimizing using the Atomic Orbital Search technique (AOS). HAPODCNN+AOS optimization gives an accuracy of around 98.93%, which is marginally better compared to that of HAPODCNN. This shows that the AOS optimization technique plays an important role in improving the proposed financial document classification model.

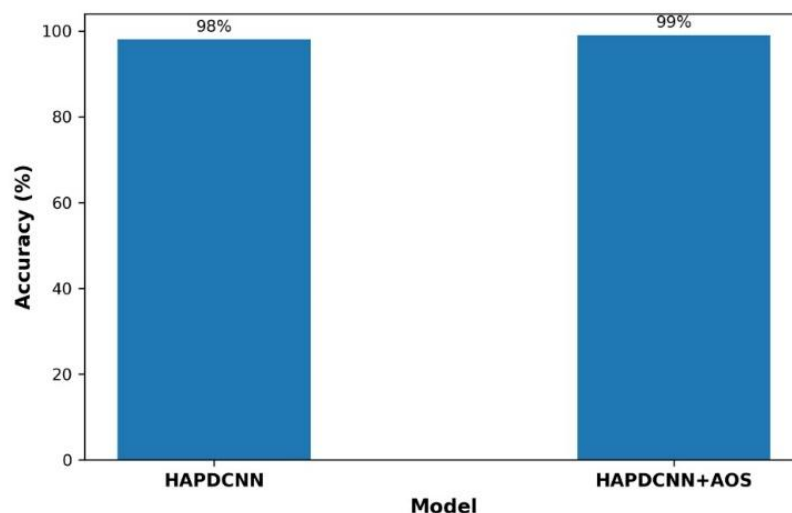


Figure 6. Accuracy analysis with and without optimization

The accuracy of the unoptimized model was evaluated at its starting configuration. This baseline accuracy represents the model's natural ability to classify images based on the features it has learned during training. Without optimization, though, the model may struggle to learn complex patterns in the image dataset, as well as noisy data and fine details. This may lead to less accurate results — especially in difficult situations. However, the proposed method had a higher accuracy without optimization. The inclusion of AOS improves the quality of HAPODCNN. AOS optimizes the weight parameter of HAPODCNN. Therefore, HAPODCNN achieves higher accuracy, and the proposed method successfully classifies financial document images.

Accuracy Analysis for Learning Rate and Batch Size

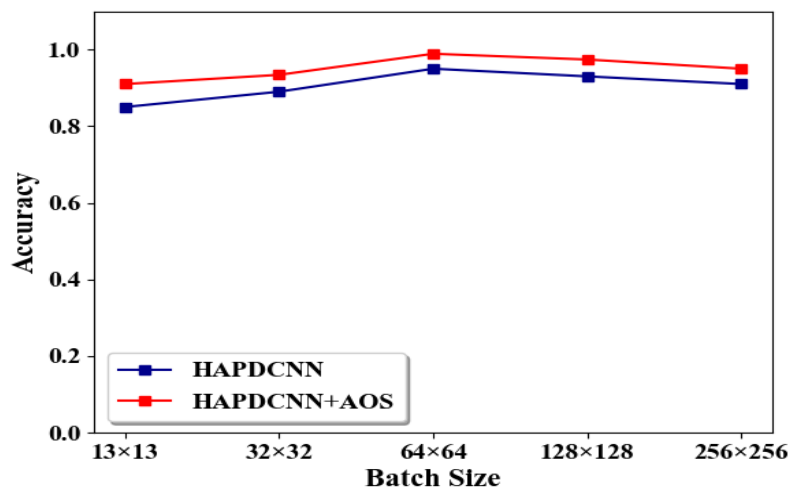


Figure 7. Accuracy analysis with respect to batch size

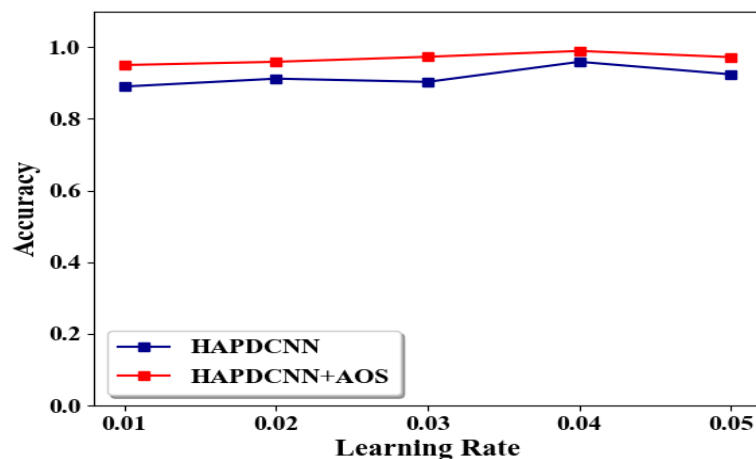


Figure 8. Accuracy analysis with respect to learning rate

In this part, the performance of the model during learning using different learning rates and batch sizes is analyzed. The research aims to evaluate the effect of adjusting the hyperparameters on the classification accuracy of the proposed framework for analyzing financial documents. The batch size allows comparison of the number of training samples used during each iteration, which is very useful for analyzing convergence. Figure 7 depicts the impact of varying minibatch sizes (16,32,64,128,256) on the classification accuracy of HAPDCNN, alongside spatial patch dimensions. Hyperparameter tuning indicates that a minibatch size of 64 yields optimal trade-off between gradient stability and classification accuracy across both optimized and unoptimized variants.

The variation of accuracy of the suggested approach for various learning rates is illustrated in figure 8. The value of learning rate α regulates how responsive the model is to the predicted error each time the weights of the model are adjusted. It can be seen from the experimental results that optimization improves the model performance concerning accuracy, while not optimizing provides no significant improvement in mutation performance.

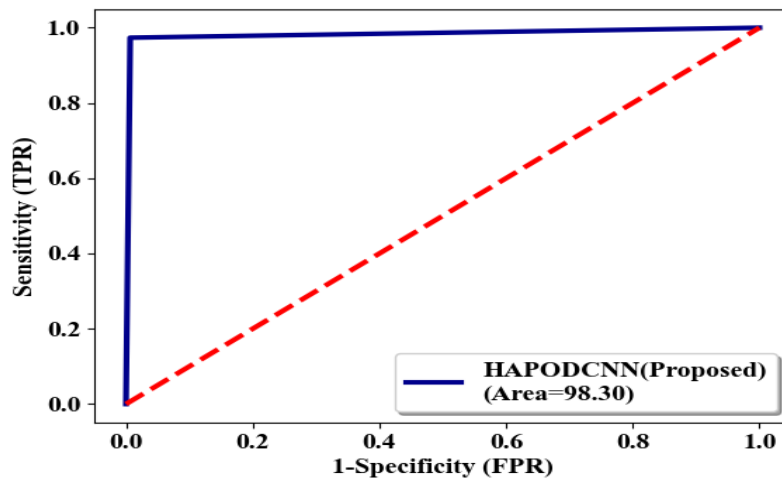


Figure 9. 1-Specificity analysis

Figure 9 below shows the receiver operating characteristic (ROC) curve that illustrates the sensitivity versus 1–specificity relationship for the HAPODCNN model. The area under the curve is 98.30%. Such an area close to unity shows good discrimination power.

Actual Label	Bank stmt	247	1	0	2	0
	Check	2	293	1	3	1
	ITR	0	1	116	1	2
	Salary slip	2	1	1	234	2
	Utility	0	2	2	1	255
		Predicted Label				
		Bank stmt	Check	ITR	Salary slip	Utility

Figure 10: Confusion matrix

The confusion matrix for the HAPODCNN classification is shown in figure 10. Figures 11(a&b) highlights the training pattern for 200 epochs, where figure 11(a) shows the training and testing accuracy curve while figure 11(b) shows the loss curve. The training and testing accuracy curves almost converge together, whereas both loss curves decrease and stabilize, which means there is very little overfitting. Despite the fact that there are only 414 distinct documents in the original dataset, only the training set is augmented up to 1,170 images. In addition to this and to the 10% validation set, early stopping, AOS optimization, and attention mechanism techniques enable robust generalization without adding augmented samples to the validation/test sets. This pattern further ensures the consistency of results in terms of accuracy of 98.93%.

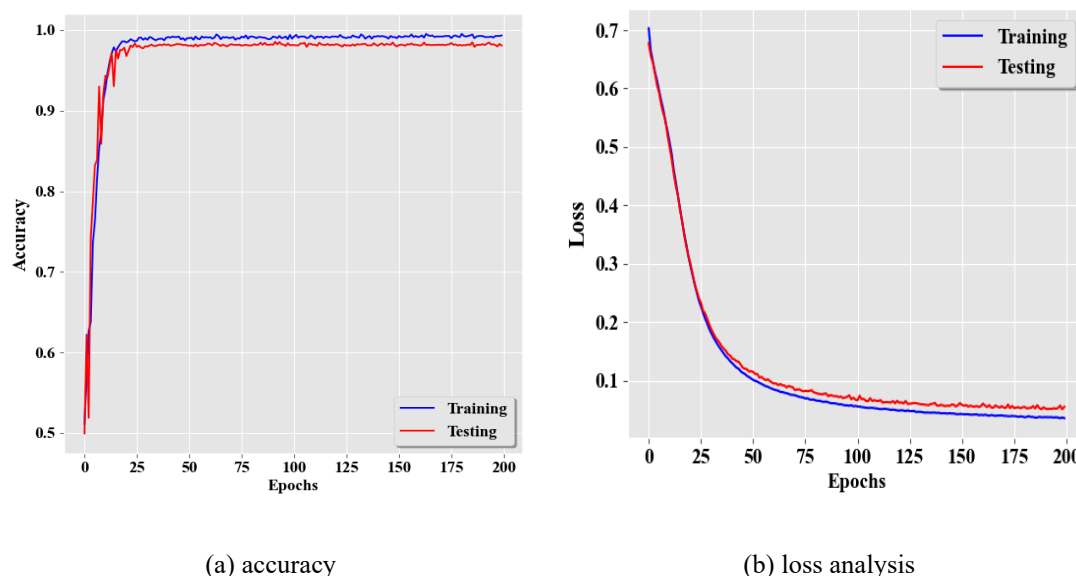


Figure 11(a&b). Training and testing

DISCUSSION

As seen from the experimental results, HAPODCNN is performing quite well in comparison with the state-of-the-art models for document classification, such as LayoutLMv3, DocFormer, and Donut. The proposed framework achieves 98.93% classification accuracy and remains above 96% even when there are changes in layout, fonts, lighting conditions, and Gaussian noise. Moreover, lower MAE, MSE, RMSE, CER, and WER scores show better classification and text extraction. As can be seen from the ablation study, the proposed framework uses the contributions of all three components mentioned. Tests for robustness reveal an accuracy of 96.1% under the most severe degradations. Training and testing curves stay very close for 200 epochs, indicating that overfitting is not prevalent. Nevertheless, the paper confines itself to only one database of financial documents, and further validation from outside sources is needed. Hence, future research will have to examine the performance of HAPODCNN on TOBACCO-3482, RVL-CDIP, and NIST-SD19, make HAPODCNN multilingual, and optimize its deployment efficiency. Further research will include model compression, quantization, federated learning, explainable AI, secured APIs, and privacy-aware enterprise integration.

CONCLUSION

In this research work, an HAPODCNN-based intelligent system is proposed to digitize and classify financial documents through MGSWF-GS preprocessing, JA-ViT-UNet-based feature extraction, a hybrid top-down attention pyramidal dual-path CNN model, and an atomic orbital search optimization method. This framework was designed in such a way that some of the major problems associated with financial document processing, like image degradation, handwritten faded text, similar layout designs of the document, and the need to consider both local and global information together, can be addressed efficiently. The proposed model was tested using 414 different financial documents from India, where each document was split into three categories: 70% for the training set, 10% for the validation set, and 20% for the testing set, with only the training set being augmented to get 1,170 images. The complete system showed a classification accuracy of 98.93% with 97.34% precision, recall, and F1-score, along with a 98.30% ROC-AUC value. In addition, character error rate and word error rate were found to be $3.2 \pm 0.1\%$ and $6.5 \pm 0.2\%$, respectively. Moreover, the model optimization involved an average execution time of 1.12 s and had 96.1% accuracy even when faced with severe document degradation, showing robustness and stability. The ablation studies showed that preprocessing, transformation-based feature extraction, pyramid classifier, and AOS optimization are all important towards achieving final performance levels. In general, from the above results, it can be concluded that the architecture that was designed is able to offer a possible solution for automated financial document processing through its

capability to recognize and classify. Future research is needed to validate the framework through different datasets such as TOBACCO-3482, RVL-CDIP, and NIST-SD19. Other future research should involve multilingualism, model compression, quantization, federated learning, explainable AI, secure APIs, and privacy-aware enterprise applications.

Compliance with Ethical Standards

Funding: The authors declare that they do not have any funding or grants for the manuscript.

Conflict of Interest: The authors declare that they do not have any conflicts of interest that influence the work reported in this paper.

Data Availability: The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethical Approval: No animals were involved in this study. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

Dataset link: <https://www.kaggle.com/datasets/mehaksingal/personal-financial-dataset-for-india>

REFERENCES

- [1] Martínek J, Lenc L, Král P. Building an efficient OCR system for historical documents with little training data. *Neural Computing and Applications*. 2020 Dec;32(23):17209-27. <https://doi.org/10.1007/s00521-020-04910-x>
- [2] Pan W, Chen L, Zhang R. Automatic recognition of financial instruments based on anisotropic partial differential equations. *Advances in Mathematical Physics*. 2021;2021(1):6529859. <https://doi.org/10.1155/2021/6529859>
- [3] Lindvall C, Deng CY, Moseley E, Agaronnik N, El-Jawahri A, Paasche-Orlow MK, Lakin JR, Volandes A, Tulskey JA, ACP-PEACE Investigators. Natural language processing to identify advance care planning documentation in a multisite pragmatic clinical trial. *Journal of pain and symptom management*. 2022 Jan 1;63(1):e29-36. <https://doi.org/10.1016/j.jpainsymman.2021.06.025>
- [4] Guha A, Samanta D, Banerjee A, Agarwal D. A deep learning model for information loss prevention from multi-page digital documents. *IEEE Access*. 2021 May 28;9:80451-65. <https://doi.org/10.1109/ACCESS.2021.3084841>
- [5] Abualigah L, Almotairi KH, Al-qaness MA, Ewees AA, Yousri D, Abd Elaziz M, Nadimi-Shahraki MH. Efficient text document clustering approach using multi-search Arithmetic Optimization Algorithm. *Knowledge-Based Systems*. 2022 Jul 19;248:108833. <https://doi.org/10.1016/j.knosys.2022.108833>
- [6] Jiang S, Hu J, Magee CL, Luo J. Deep learning for technical document classification. *IEEE Transactions on Engineering Management*. 2022 Mar 8;71:1163-79. <https://doi.org/10.1109/TEM.2022.3152216>
- [7] Jørgensen RK, Igel C. Machine learning for financial transaction classification across companies using character-level word embeddings of text fields. *Intelligent Systems in Accounting, Finance and Management*. 2021 Jul;28(3):159-72. <https://doi.org/10.1002/isaf.1500>
- [8] Visser A, Cheung J, Marshall E. Digitizing and cataloging a corporate annual reports collection: A Canadian perspective. *Journal of Business & finance liBrarianship*. 2022 Jul 3;27(3):176-94. <https://doi.org/10.1080/08963568.2022.2072460>
- [9] Narang SR, Kumar M, Jindal MK. DeepNetDevanagari: a deep learning model for Devanagari ancient character recognition. *Multimedia Tools and Applications*. 2021 May;80(13):20671-86. <https://doi.org/10.1007/s11042-021-10775-6>
- [10] Biesner D, Ramamurthy R, Stenzel R, Lübbering M, Hillebrand L, Ladi A, Pielka M, Loitz R, Bauckhage C, Sifa R. Anonymization of German financial documents using neural network-based language models with contextual word representations. *International Journal of Data Science and Analytics*. 2022 Mar;13(2):151-61. <https://doi.org/10.1007/s41060-021-00285-x>
- [11] Chlahawi MO. Application of blockchain technology to reduce costs. *International Academic Journal of Social Sciences*. 2024;11(1):26-38. <https://doi.org/10.9756/IAJSS/V11I1/IAJSS1104>
- [12] Cho S, Moon J, Bae J, Kang J, Lee S. A framework for understanding unstructured financial documents using RPA and multimodal approach. *Electronics*. 2023 Feb 13;12(4):939. <https://doi.org/10.3390/electronics12040939>

- [13] Van Hoai DP, Duong HT, Hoang VT. Text recognition for Vietnamese identity card based on deep features network: DP Van Hoai et al. International Journal on Document Analysis and Recognition (IJDAR). 2021 Jun;24(1):123-31. <https://doi.org/10.1007/s10032-021-00363-7>
- [14] Mirabi M, Abdullah MMAB. Building a Framework for Analyzing the Impact of International Trade Agreements on Local Communities Using Social Network Models. IAJH. 2025 Sep. 30;12(3):16-23. <https://doi.org/10.71086/IAJH/V12I3/IAJH1210>
- [15] De R, Chakraborty A, Sarkar R. Document image binarization using dual discriminator generative adversarial networks. IEEE Signal Processing Letters. 2020 Jun 22;27:1090-4. <https://doi.org/10.1109/LSP.2020.3003828>
- [16] Liu F, Zheng J, Zheng L, Chen C. Combining attention-based bidirectional gated recurrent neural network and two-dimensional convolutional neural network for document-level sentiment classification. Neurocomputing. 2020 Jan 2;371:39-50. <https://doi.org/10.1016/j.neucom.2019.09.012>
- [17] Pande SD, Jadhav PP, Joshi R, Sawant AD, Muddebihalkar V, Rathod S, Gurav MN, Das S. Digitization of handwritten Devanagari text using CNN transfer learning—A better customer service support. Neuroscience Informatics. 2022 Sep 1;2(3):100016. <https://doi.org/10.1016/j.neuri.2021.100016>
- [18] Dey M, Mia SM, Sarkar N, Bhattacharya A, Roy S, Malakar S, Sarkar R. A two-stage CNN-based hand-drawn electrical and electronic circuit component recognition system. Neural Computing and Applications. 2021 Oct;33(20):13367-90. <https://doi.org/10.1007/s00521-021-05964-1>
- [19] Nematov O, Uralov X. A Comparative Study of HRM Practices in Multinational Firms Through a Configurational Model of Best Practices for Talent Retention. IAJOBHRM. 2026 Mar. 30;13(1): 36-44. <https://doi.org/10.71086/IAJOBHRM/V13I1/IAJOBHRM1305>
- [20] Kashinath T, Jain T, Agrawal Y, Anand T, Singh S. End-to-end table structure recognition and extraction in heterogeneous documents. Applied Soft Computing. 2022 Jul 1;123:108942. <https://doi.org/10.1016/j.asoc.2022.108942>