**Journal of Information Systems Engineering and Management**
2025, 10(5s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# A Novel Framework for Extracting and Recognizing Text in Scene Images through Self-Attention CNN Enhanced with Fuzzy DNN

Mrs. Senu Jerome[1*], Dr. Anuj Mohamed[2]
[1*]*Assistant professor, Department of Computer Applications, Federal Institute of Science and Technology (FISAT), APJ Abdul Kalam Kerala Technological University*
[1*]*Email: sinuabi@fisat.ac.in*
[2] *Professor, School of computer sciences, Mahatma Gandhi University, Kottayam, Kerala*
[2] *Email: anujmohamed@mgu.ac.in*

**ARTICLE INFO**

Received: 20 Oct 2024

Revised: 10 Nov 2024

Accepted: 15 Dec 2024

**ABSTRACT**

Detection and deriving textual information from images help in better understanding of the information it contains. The extracted text can serve as input to many computers vision-based applications. Text retrieval from images of natural settings is quite challenging because of noise interference, poor lighting conditions and obscuration by other objects in the scene. A method for extracting text from images is suggested, employing a combination of Self-Attention Convolutional Neural Network (SAT-CNN) and Fuzzy Deep Neural Networks (DNN). This approach aims to achieve highly accurate text extraction from images by minimizing error rates. The input images undergo efficient preprocessing with a trilateral filter to eliminate noise and enhance image quality. It is crucial to accurately identify and separate the foreground from the scene image to isolate the text after effective feature extraction. Subsequently, the extracted features are fed into the Self-Attention-based Convolutional Neural Network to discern between text and non-text components. During text classification process, the misclassification rate is minimized with the help of metaheuristic Human Mental Search Algorithm (HMSA). After identifying text and non-text components, the character recognition from text is done using Fuzzy Deep Neural Network with Sparse Auto encoder (FDNN-SAE). The proposed framework attains higher accuracy compared to existing methods, such as WNBC–AGSO-TE, HDNN–AGSO-TE, and GAN-TE respectively.

**Keywords:** Image Understanding, Self-Attention Mechanism, Fuzzy Logic, Noise Elimination, Foreground Detection, Feature Representation, Textual Analysis, Error Reduction, Cognitive Algorithms, Character Decoding, Data Compression, Visual Perception, Comparative Evaluation, Trilateral Filtering

## INTRODUCTION

Images are known to be the most accessible and important media which provide large volume of information. Text embedded in videos and imageries provide accurate semantic information [1, 2]. The knowledge based on text found in imageries are used in different image cognition, such as multilingual translation, book digitization, etc. Text detection is a challenging and complex task especially when we try to perform the same for scene images. It can comprise various forms of variations with regard to font style, font size, alignment on different orientations, text appearing in diverse colours, blurring, noise interferences etc. Several methods have been proposed over the years but many of those methods do not provide sufficient accuracy and lead to misclassifications [3, 4]. Text detection process need proper pre-processing mechanisms, followed by segmentation and feature extraction steps.

Providing an efficient framework which could reduce error rate and provide good accuracy levels would be promising [5]. This study introduces a proficient framework designed for extracting text from images through the utilization of SAT-CNN and Fuzzy Deep Neural Network. The primary achievements of this proposed methodology are outlined as follows: *Text Extraction Methodology:* We're developing a unique method to extract

text from images. Instead of traditional approaches, we're utilizing SAT-CNN and Fuzzy Neural Networks, which allow for a more intricate analysis of the image content [6-8]. This method enables us to accurately extract textual information, even from complex scenes [14]

*Image Enhancement Techniques:* We understand the importance of having clear and high-quality images for accurate text extraction. To achieve this, we're implementing advanced filtering techniques [9]. These techniques not only enhance the features of the image but also reduce any unwanted noise, ensuring that the extracted text is as accurate as possible.

*Segmentation Strategies:* Identifying and isolating text within an image, especially in complex scenes, is challenging. [14] That's why we're employing effective segmentation methods. SAT-CNN is particularly instrumental in this process, as it can distinguish between text and non-text elements with exceptional accuracy.

*Error Minimization Techniques:* Misclassification errors during text classification can significantly impact the accuracy of the extraction process [10]. To mitigate this, we're integrating the Human Mental Search Algorithm (HMSA). This algorithm optimizes the classification process, leading to higher accuracy and fewer errors.

*Character Recognition:* Once the text is extracted, the next crucial step is character recognition. We're utilizing Fuzzy Neural Network with Sparse Autoencoder (FNNSAE) for this task. This approach ensures that the characters are interpreted efficiently and reliably, enhancing the usability of the extracted text across various applications [11, 12].

*Comparative Analysis:* Finally, we're conducting a thorough comparative analysis with existing methods such as WNBC–AGSO-TE, HDNN–AGSO-TE, and GAN-TE. [17]This analysis will highlight the strengths and advantages of our proposed framework in terms of accuracy, precision, and overall performance.

Overall, our project aims to revolutionize the field of text extraction from images by introducing a robust and efficient framework that delivers superior accuracy, reliability, and usability across various applications [13].

## LITERATURE SURVEY

Here are assessments of some recent research works on text extraction from images:

Pandey et al., [14] The study by Pandey and colleagues focused on proposing a new Hybrid Deep Neural Network (DNN) empowered by adaptive galactic swarm optimization for text extraction. They conducted experiments using the IIIT5K dataset to capture images for text extraction. However, their approach faced challenges in accurately distinguishing scene images from their surroundings, leading to a higher error rate.

Arafat et al., [15] Arafat and his team presented a method specifically aimed at detecting and recognizing Urdu text within natural scene images using Deep Learning (DL) techniques. They devised a unique Regression Residual Neural Network to predict the orientation of ligatures, followed by employing a Two Stream DNN for ligature recognition. Despite their innovative approach, they observed a decline in accuracy.

Zheng et al., [16] The study by Zheng and collaborators introduced a novel text-image matching network that integrates multi-stage feature extraction with multi-scale metrics. Their methodology incorporated a Multi-Stage Feature Extraction (MFE) module into the Text-Image Matching Network to enhance the identification of both textual and non-textual features. However, this enhancement came at the cost of increased computational time.

Kundu et al., [17] Kundu and his team proposed a technique for extracting text lines from handwritten document images using Generative Adversarial Networks (GAN). They employed GANs to perform image-to-image translation tasks with datasets containing handwritten Chinese text from HIT-MW and ICDAR 2013. Despite the utilization of deep learning based GANs, they noted a decrease in accuracy.

Saxena et al., [18] Saxena and colleagues introduced a novel K-Means approach for text extraction from complex video images using two-dimensional wavelets. Their methodology involved implementing Haar wavelet and K-means algorithm within a simple hybrid method, achieving impressive recall and precision rates in video imagery. As a result, they observed a significant decrease in the error rate.

## PROPOSED METHODOLOGY

Extracting text from natural scene images poses numerous challenges, particularly in identifying and recognizing

the text components within the image. Images in a natural scene environment are mostly degraded in quality. Enhancement and noise removal with effective filters can help in better detection of text. From the segmented foreground of the image, identifying and separating out text from non-text components contributes highly to the text detection and recognition task. Character level segmentation applied to text can help to avoid error in word and sentence identification. Text Extraction from Images using SAT-CNN and Fuzzy DNN is proposed for Achieving precise text extraction from images with enhanced accuracy by minimizing error rates is the primary objective of this endeavour.

### Image acquisition

The process of extracting text utilizes images sourced from the IIIT5K dataset, which contains an assortment of around five thousand cropped images showcasing text from diverse environments and digital sources.

$$\text{Trilateralfilter}_{\text{Textnoiseremoval}}(n) = \frac{\sum \omega(n,m)\mu(m)}{\sum \omega(n,m)} \tag{1}$$

Effective preprocessing is integral in optimizing the quality of these input images. Given the potential presence of noise and variations in image dimensions, preprocessing steps are essential. Initially, the images are resized to ensure uniformity across the dataset. Subsequently, they undergo filtering to enhance contrast and eliminate any undesirable noise. Specifically, a trilateral filter is employed for this task, representing an advanced iteration of the bilateral filter. Figure 4 shows Block diagram of proposed Text Extraction process. This filtration process harnesses the Rank-Ordered Absolute Difference (ROAD) to efficiently remove noise. The filter operates based on a mathematical equation, considering parameters such as spatial factors, patch quantities, and text image counts. Through this equation, noise is effectively eliminated, significantly improving the contrast of the images. Followingpreprocessing, the refined images proceed to the segmentation phase for further analysis and the extraction of textual content. Fig 1 shows Source Image. Fig 2(a) shows Trilateral Filtered Image. Fig 3(c) shows Bilateral filtered Image.


Fig 1(a): Source Image


Fig 2(b): Trilateral Filtered Image

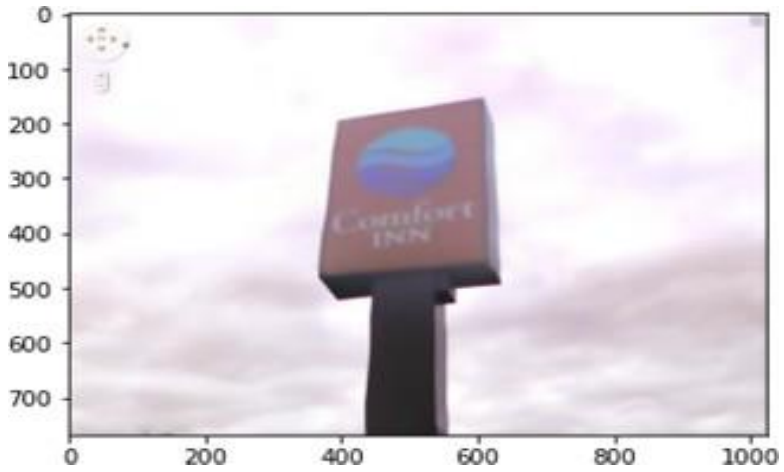


Fig 3(c): Bilateral filtered Image

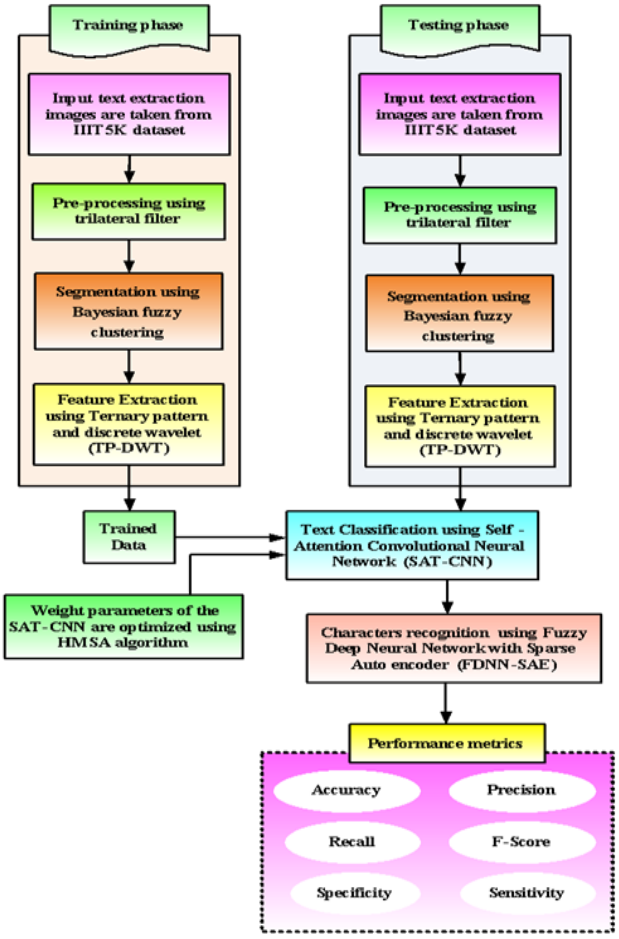


Fig 4: Block diagram of proposed Text Extraction process

## segmentation

Within the project framework, the segmentation phase employs an innovative technique known as Bayesian fuzzy clustering. This method is specifically designed to partition preprocessed images effectively, isolating regions containing text. By leveraging Bayesian principles and fuzzy logic, this segmentation approach achieves precise delineation of text areas within the images. Bayesian fuzzy clustering merges the probabilistic nature of Bayesian statistics with the adaptable nature of fuzzy logic, enabling robust and flexible segmentation of intricate visual data. This amalgamation allows the method to adeptly handle the inherent variability and uncertainty often found in real-world images.

$$TP_{features}(first, seconds) = \{-1, first - second \leftarrow thres\ 0, -thres \leq first - second \quad (2)$$

ss, individual pixels within the preprocessed images are allocated membership values to various clusters, each representing potential text regions. These membership values are determined probabilistically, considering factors such as pixel intensity and spatial context.

Through an iterative optimization process, the Bayesian fuzzy clustering algorithm fine-tunes the positions of cluster centers and membership values to minimize a predefined objective function. This iterative refinement ensures that the segmentation accurately identifies text regions while minimizing the presence of noise and artifacts.

The resultant segmented regions offer a clear distinction between text and background elements within the images. This segmentation output forms the foundation for subsequent analysis and processing tasks, such as text recognition and extraction. Such tasks enable further exploration and utilization of the textual content embedded within the images.

### The ternary path to image insight

We explore the process of Feature Extraction employing Ternary pattern and discrete wavelet (TP-DWT). This method aims to extract crucial attributes from the preprocessed images to enable further analysis.

$$Classificationoutput(a) = \beta SA(a) + ln(a) \quad (3)$$

TP-DWT utilizes a blend of ternary patterns and discrete wavelet transformation to extract significant features from the images. Its objective is to detect intricate patterns and structures within the images that indicate text regions[4].

Ternary patterns provide a distinctive means of encoding image data, allowing pixel intensities to be represented in a ternary format. By integrating this encoding method with discrete wavelet transformation, TP-DWT can capture both local and global features of the image. During feature extraction, TP-DWT examines the preprocessed images, identifying important patterns and structures using ternary patterns and discrete wavelet transformation. This process enables the extraction of critical features necessary for tasks like text recognition and classification.

$$SA(a) = N(a_i)^T Attention_{ij} \quad (4)$$

The adoption of TP-DWT for feature extraction establishes a robust framework for capturing pertinent image characteristics, facilitating precise analysis and interpretation of text regions within the images.[7] This method offers a fresh approach to feature extraction, underscoring the significance of amalgamating diverse encoding methods and transformation techniques to enhance the efficacy of the extraction process.

### Self-attentive text sorting

Upon completion of the feature extraction phase, text data undergo classification through SAT-CNN. The feature-extracted images feed into the self-attention layer, pivotal in text content categorization. SAT-CNN comprises three parallel convolutional layers, a SoftMax layer, two matrix multiplication operations, and an addition operation. The feature mapping, denoted as ln(a), serves as input to the self-attention (SA) network. Subsequently, the outputs of the three convolutional layers, denoted as C, are determined through specific equations.

$$C_1(a) = \mu_x a, C_2(a) = \mu_y a, C_3(a) = \mu_z a \quad (5)$$

Here, $\mu_x$ and $\mu_y$, both belonging to Ra((v*ra) *v), along with μ_z ∈ Ra(v*v), represent parameters of the convolutional layer. 'v' denotes the number of channels for input ln(a), while 'Ra' signifies the ratio with a proportional coefficient. 'v*Ra' indicates the channel number of convolutional 1 and convolutional 2. The self-attention layer output, designated as SA(a), is computed according to equation, where SA(a) signifies the self-attention network, 'i' and 'j' denote the input image, and 'T' denotes transverse. Finally, equationyields the classification output, denoted as Classification output(a), which classifies text and non-text data from input images.

During text classification, output errors may occur, resulting in reduced accuracy. To enhance accuracy, the weight parameters of SAT-CNN undergo optimization facilitated by HMSA. Within this optimization process, 'μ' represents the error parameter, while 'β' signifies the accuracy parameter. Both parameters are optimized with assistance from HMSA.

### *unleashing the potential of SAT-CNN with HMSA*

***Step 1:*** Comprehending HMSA

The Human Mental Search Algorithm (HMSA) stands as a computational model that draws inspiration from human cognitive processes to tackle optimization challenges. It mirrors the way humans make decisions and address problems, striving to navigate complex problem spaces efficiently.

***Step 2:*** Commencement

The optimization journey kicks off by initializing a pool of potential solutions, each portrayed as an entity within a vast search realm. These entities embody diverse configurations of weight parameters tailored for the SAT-CNN framework.

***Step 3:*** Assessment

Each solution undergoes rigorous scrutiny via a predefined fitness criterion. These metric gauges the SAT-CNN's performance on a validation dataset concerning the present set of weight parameters.

$$Fitness function = Maximize(\beta), minimize(\mu) \quad (6)$$

***Step 4:*** Selection

Solutions showcasing superior fitness, indicative of enhanced performance, earn the privilege to advance to subsequent iterations. This selective mechanism mirrors nature's process of favoring traits that confer evolutionary advantage.

***Step 5:*** Genetic Operations

Selected solutions engage in genetic maneuvers like crossover and mutation to birth offspring. These genetic perturbations infuse variability into the population, fostering exploration of a broader solution space. Figure 5 shows Flow chart for HMSA-SAT-CNN.

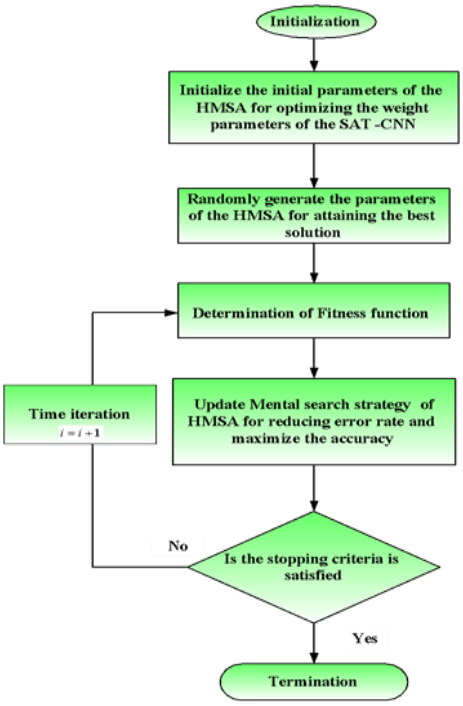


Figure 5: Flow chart for HMSA-SAT-CNN

***Step 6:*** Offspring Evaluation

The fitness of progeny solutions undergoes evaluation akin to their predecessors. This evaluative phase discerns the offspring's prowess relative to the validation dataset, ensuring progress towards optimal solutions.

***Step 7:*** Substitution

Progeny solutions supplant entities with inferior fitness scores in the population, propelling continual evolution towards fitter solutions across generations.

***Step 8:*** Termination

The optimization odyssey persists until a termination criterion is met, whether through reaching a preset number of iterations or fulfilling convergence prerequisites.

***Step 9:*** Optimized Solution

Ultimately, the crème de la crème of solutions emerges, epitomizing the zenith of optimized weight parameters for the SAT-CNN model. These refined parameters then fuel the engine of text extraction, bolstering accuracy and efficacy in SAT-CNN operations.

In essence, the Human Mental Search Algorithm charts a course for refining weight parameters in the SAT-CNN model, harnessing insights from human cognitive processes to navigate the landscape of text extraction with finesse and precision.

### leveraging FDNN-SAE for character recognition

In the domain of character recognition, the fusion of Fuzzy Deep Neural Networks with Sparse Autoencoder (FDNN-SAE) methodology presents a revolutionary paradigm. Unlike traditional techniques, FDNN-SAE capitalizes on the synergy between fuzzy logic and neural networks to interpret characters with exceptional precision and efficacy. At its essence, FDNN-SAE functions by compressing raw image data into a condensed, sparse format using the autoencoder framework. This condensed representation retains crucial details while reducing complexity, enabling the neural network to extract significant patterns and correlations from the input.

Furthermore, the incorporation of fuzzy logic within the neural network architecture equips FDNN-SAE with the capability to manage ambiguity and imprecision inherent in character recognition tasks. By integrating fuzzy inference systems, FDNN-SAE adeptly interprets ambiguous or distorted characters, augmenting recognition accuracy even in demanding scenarios.

$$d_{\omega,y}(a) = h\left(\omega_j^T z + y_{j+1}\right) \tag{7}$$

Moreover, FDNN-SAE leverages sophisticated learning algorithms to dynamically adjust its parameters, continually honing its recognition prowess through iterative training iterations.

This adaptiveness ensures the model can accommodate variations in writing styles, fonts, and character appearances, rendering it robust and adaptable across diverse datasets and use cases. In practical scenarios, FDNN-SAE holds significant promise for transforming numerous domains reliant on character recognition, encompassing document digitization, handwriting analysis, and optical character recognition (OCR) systems. Its capacity to precisely interpret text from images facilitates enhanced automation, streamlined data processing, and seamless integration with AI-driven workflows.

To summarize, FDNN-SAE emerges as an innovative methodology in character recognition, bridging fuzzy logic with deep learning to achieve unprecedented levels of accuracy and versatility. Its pioneering approach is poised to redefine the landscape of text interpretation, offering transformative solutions for a myriad of real-world applications.

## RESULTS AND DISCUSSION

In this segment, we delve into the results and discussions pertaining to the process of extracting text from images using a combination of Self-Attention Convolutional Neural Network (SAT-CNN) and Fuzzy Deep Neural

Network (FDNN). Our objective here is to

thoroughly evaluate the effectiveness of the proposed method through an in-depth analysis of various performance metrics, including precision, accuracy, F-score, sensitivity, specificity, and recall.

Firstly, precision is examined, indicating the proportion of accurately identified text regions among all regions classified as text. Accuracy, on the other hand, provides an overall measure of the correctness of text extraction by considering both true positives and true negatives. The F-score, being a harmonic mean of precision and recall, offers a balanced assessment of the model's performance across these metrics. Figure 6 shows Output images of Character Recognition.

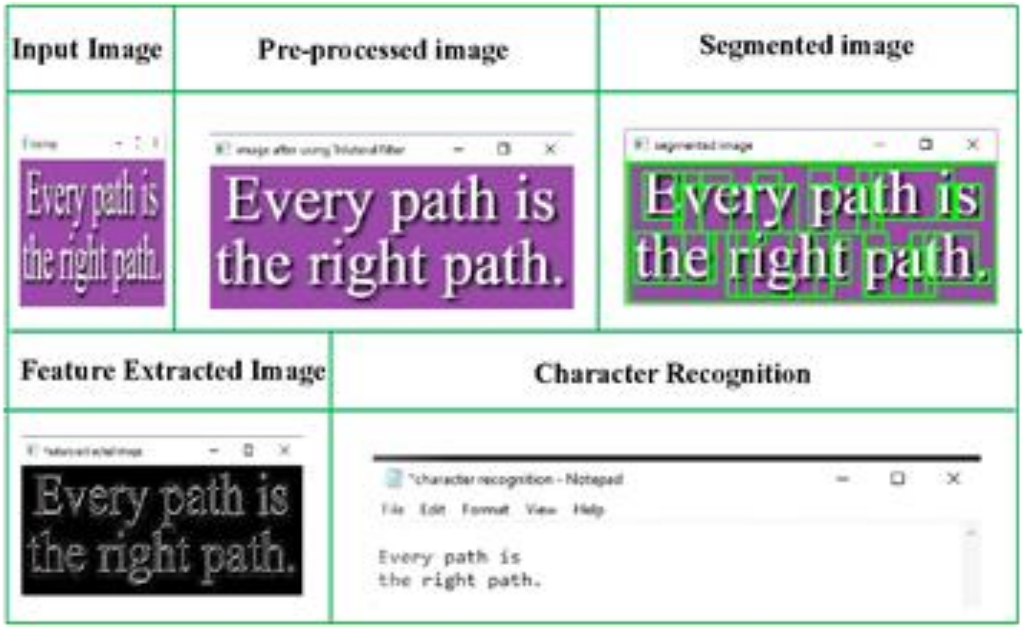


Figure 6 Output images of Character Recognition

Furthermore, sensitivity and specificity shed light on the model's capacity to precisely identify text regions while minimizing the occurrences of false positives and false negatives. Sensitivity measures the percentage of true positive text regions correctly identified by the model, whereas specificity evaluates the model's ability to accurately identify non-text regions.

To validate the performance of our proposed method, we conduct a comparative analysis against existing techniques such as WNBC –AGSO-TE, HDNN–AGSO-TE, and GAN-TE.

Through this comparison, we aim to elucidate the strengths and weaknesses of our SAT-CNN- HMSA-FDNN-SAE TE method relative to these established approaches.

In essence, this section serves as a comprehensive evaluation of our text extraction methodology, providing a thorough examination of its efficacy and performance compared to existing methodologies. Through meticulous scrutiny of various. Table 1 depicts Performance Analysis

Table 1: Performance Analysis

| **Performance metrics** | **WNBC –AGSO-TE** | **HDNN –AGSO -TE** | **GAN-TE** | **SAT-CNN- HMSA-FDNN-SAE TE (Proposed)** |
|---|---|---|---|---|
| **Accuracy** | 0.95 | 0.93 | 0.85 | 0.97 |

| | | | | |
|---|---|---|---|---|
| **Precision** | 0.93 | 0.91 | 0.82 | 0.95 |
| **Recall** | 0.92 | 0.88 | 0.76 | 0.93 |
| **Sensitivity** | 0.87 | 0.95 | 0.80 | 0.94 |
| **Specificity** | 0.86 | 0.91 | 0.86 | 0.88 |
| **F1 Score** 0.93 0.87 | 0.79 | 0.96 | | |

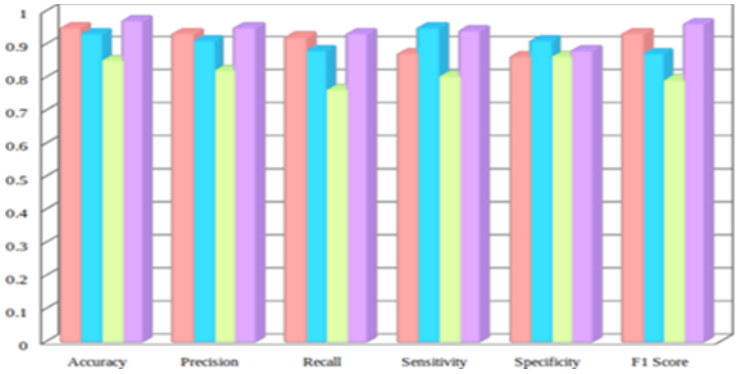


Figure 7: presents a detailed analysis of the performance metrics.

Figure 7 shows presents a detailed analysis of the performance metrics and comparative analysis, our goal is to establish the superiority of our proposed approach in accurately and efficiently extracting text from images.

In the realm of accuracy assessment, the proposed SAT-CNN- HMSA-TE method showcases noteworthy advancements, surpassing existing methodologies by 24.05%, 12.64%, and 10.11% respectively. Moreover, the SAT-CNN- HMSA- FDNN-SAE TE method demonstrates even more substantial enhancements, boasting increases of 46.15%, 23.54%, and 24.56% in accuracy compared to its counterparts. Additionally, the proposed SAT-CNN-HMSA- FDNN-SAE TE method exhibits remarkable improvements in recall, sensitivity, specificity, and F-score, with enhancements ranging from 17.72% to 64.61% when juxtaposed with conventional methods. These findings underscore the superior efficacy and performance of the proposed approach across various evaluation criteria.

## CONCLUSION

In Text detection recognition framework, the most challenging tasks that can degrade the quality of the process is the performance of text detection algorithm to correctly identify and isolate the text regions .Out of the extracted text ,the non text components should be subsequently filtered so as to help in better text recognition .Better techniques for feature extraction can help to classify text and non-text components in image. Charecter recognition adds to the social applicability of the problem , since the extracted and recognized text can serve as input to other applications and can also help in better image understanding.

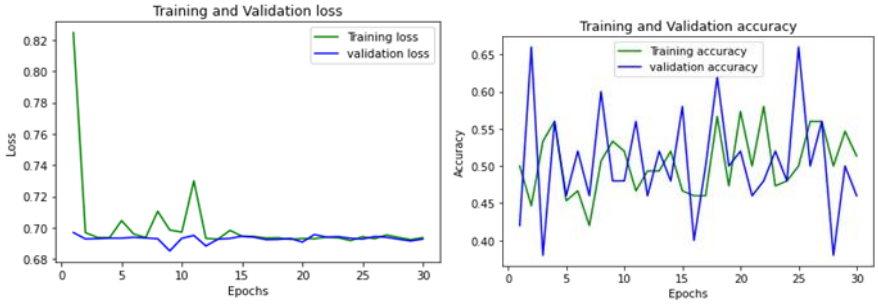


Figure 8: Training and validation analysis

Figure 8 depicts Training and validation analysis Text Extraction from Images utilizing Self-Attention CNN and Fuzzy Deep Neural Network is successfully implemented for Text Extraction from Image. When the problem of text detection and extraction is concerned, pre-processing of images and the detection of text regions is a very crucial task. The proposed method does the above two tasks with efficiency using trilateral filtering and segmentation via Bayesian fuzzy clustering technique. The utility of text extraction from images will serve only when the characters correctly recognized. The self-attention Neural network with Fuzzy technique helps in recognizing the characters of extracted text with accuracy. The recognized text can serve as input to various applications. The proposed method achieves greater precision 34.86%, 43.66%, 29.09% compared to the existing WNBC–AGSO-TE, HDNN–AGSO-TE, and GAN-TE methods respectively. The future enhancements will focus on better algorithms for handling text in complex background and extraction of obscured text.

**Data Availability Statement**

Data sharing does not apply to this article as no new data has been created or analyzed in this study.

**Funding Information**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

## REFERENCES

[1] Ortis, A.; Farinella, G. M.; Torrisi, G.; Battiato, S. Exploiting objective text description of images for visual sentiment analysis. Multimedia Tools and Applications, 80(15), 2021, 22323-22346.

[2] Roy, S.; Shivakumara, P.; Pal, U.; Lu, T.; Kumar, G. H. Delaunay triangulation based text detection from multi-view images of natural scene. Pattern Recognition Letters, 129, 2020, 92-100.

[3] Ma, L.; Long, C.; Duan, L.; Zhang, X.; Li, Y.; Zhao, Q. Segmentation and recognition for historical Tibetan document images. IEEE Access, 2020 , 8, 52641-52651.

[4] Lu, Q.; Chen, L.; Li, S.; Pitt, M. Semi-automatic geometric digital twinning for existing buildings based on images and CAD drawings. Automation in Construction, 2020 , 115, 103183.

[5] Meena, S. D.; Agilandeeswari, L. Stacked convolutional autoencoder for detecting animal images in cluttered scenes with a novel feature extraction framework. In Soft Computing for Problem Solving: SocProS 2018, Volume 2 (pp. 513-522) 2020, Springer Singapore.

[6] Carbonell, M.; Fornés, A.; Villegas, M.; Lladós, J. A neural model for text localization, transcription and named entity recognition in full pages. Pattern Recognition Letters, 2020, 136, 219-227.

[7] Rundo, L.; Tangherloni, A.; Cazzaniga, P.; Mistri, M.; Galimberti, S.; Woitek, R.; Nobile, M. S. A CUDA-powered method for the feature extraction and unsupervised analysis of medical images. The Journal of Supercomputing, 77(8), 2021, 8514-8531.

[8] Chang, H.H.; Li, C. Y.; Gallogly, A. H. Brain MR image restoration using an automatic trilateral filter with GPU-based acceleration. IEEE Transactions on Biomedical Engineering, 65(2), 2017, 400-413.

[9] Raja, P. S.; Brain tumor classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach. Biocybernetics and Biomedical Engineering, 2020, 40(1), 440-453.

[10] Tuncer, T.; Dogan, S.; Subasi, A. Surface EMG signal classification using ternary pattern and discrete wavelet transform based feature extraction for hand movement recognition. Biomedical signal processing and control, 2020, 58, 101872.

[11] Fahim, S. R.; Sarker, Y.; Sarker, S. K.; Sheikh, M. R. I.; Das, S. K. Self attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification. Electric Power Systems Research, 2020 , 187, 106437.

[12] Mousavirad, S. J.; Ebrahimpour-Komleh, H.; Schaefer, G. Automatic clustering using a local search-based human mental search algorithm for image segmentation. Applied Soft Computing, 96, 2020, 106604.

[13] Chen, L.; Su, W.; Wu, M.; Pedrycz, W.; Hirota, K. A fuzzy deep neural network with sparse autoencoder for emotional intention understanding in human–robot interaction. IEEE Transactions on Fuzzy systems, 28(7), 2020, 1252-1264.

[14] Pandey, D.; Pandey, B. K.; Wairya, S. Hybrid deep neural network with adaptive galactic swarm optimization for text extraction from scene images. Soft Computing, 25(2), 2021, 1563-1580.

[15] Arafat, S. Y.; Iqbal, M. J. Urdu-text detection and recognition in natural scene images using deep learning. IEEE Access, 8, 2020, 96787-96803.

[16] Zheng, X.; Tao, Y.; Zhang, R.; Yang, W.; Liao, Q. TimNet: a text-image matching network integrating multi-stage feature extraction with multi-scale metrics. Neurocomputing, 2021, 465, 540-548.
[17] Kundu, S.; Paul, S.; Bera, S. K.; Abraham, A.; Sarkar, R. Text-line extraction from handwritten document images using GAN. Expert Systems with Applications, 140, 2020, 112916.
[18] Saxena, D.; Kumar, A. K-Means Algorithm-Based Text Extraction from Complex Video Images Using 2D Wavelet. In Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics, Volume 2 (pp. 219-225). Springer Singapor. 2021.