International Journal of Computer Sciences and Engineering
Vol.**13**, Issue.**2**, pp.**86-97**, February **2025**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Review Article

An Exploratory Review on Text Detection and Recognition in Images and Videos Using Machine Learning and Deep Learning Techniques

Keerthana Sundar Raj[1*], J. Savitha[2]

[1,2]Dept. of Computer Science, Dr.N.G.P. Arts & Science College, Coimbatore, Tamilnadu, India

*Corresponding Author:

Received: 23/Oct/2024; **Accepted**: 27/Jan/2025; **Published**: 28/Feb/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i2.8697

Abstract: The increasing reliance on images and videos as sources of information has led to a growing demand for automated text detection and recognition systems. This literature review explores the current advancements in text extraction methodologies, focusing on machine learning and deep learning techniques. Various approaches, including text detection, localization, recognition, and tracking, are discussed alongside the challenges posed by environmental conditions, text alignment, font variations, and background noise. The study highlights applications in license plate recognition, industrial automation, vehicle tracking, and self-navigating automobiles, where text extraction plays a crucial role. Furthermore, a comparative analysis of existing machine learning-based and deep learning-based models is conducted, evaluating their effectiveness in different scenarios. This examination also discusses evaluation metrics used to validate model performance and identifies the computational challenges associated with processing high-resolution images and videos in real-time. The findings emphasize the need for robust mathematical models and optimization techniques to improve the efficiency and accuracy of text recognition systems

Keywords: Text Detection, Optical Character Recognition (OCR), Deep Learning, Machine Learning, Text Recognition, Image Processing, Video Analysis, Feature Extraction, Computer Vision and Natural Language Processing (NLP

1. Introduction

The exponential growth of digital content across various platforms, such as social media, surveillance systems, autonomous vehicles, and industrial automation, has significantly increased the need for automated text detection and recognition systems in images and videos. These systems are essential for extracting meaningful textual information from visual data, enabling efficient information retrieval, data indexing, and real-time decision-making across different industries [1]. **Text extraction from images and videos has widespread applications across various domains, significantly enhancing automation, accuracy, and efficiency.** One of the key applications is License Plate Recognition (LPR), where Automatic Number Plate Recognition (ANPR) technology is used in traffic surveillance systems for vehicle identification, law enforcement, toll collection, and traffic monitoring. However, challenges such as varying weather conditions, occlusions, and different font styles make accurate recognition difficult. In industrial automation, automated text recognition plays a crucial role in quality control, inventory management, and automated assembly lines, where it is used to read product labels, serial numbers, and barcodes. This reduces manual errors, improves productivity, and ensures regulatory compliance. Similarly, document digitization is another significant application, where businesses and government organizations utilize OCR-based systems to convert scanned documents, handwritten notes, and printed records into digital text. This conversion enables searchable and editable text storage, facilitating efficient data retrieval, archiving, and automation of document workflows.

Self-navigating vehicles, **including autonomous cars and delivery drones**, rely on text recognition to read **road signs, traffic signals, and lane markings** for navigation. Accurate text detection in such scenarios ensures safe and effective real-time decision-making. In medical imaging and healthcare, text recognition is widely used in digital pathology, radiology reports, and patient record management to extract and process critical medical information. This technology assists in automated diagnostics, medical transcription, and AI-driven decision support systems, improving patient care and healthcare data management. The integration of text recognition across these domains demonstrates its transformative impact on automation, accuracy, and efficiency in various industries, making it a key technology in modern AI-driven systems [2].

Traditional Optical Character Recognition (OCR) methods have been widely utilized for text extraction from images and scanned documents. However, these methods often encounter several challenges that affect their accuracy and reliability. One of the major issues is background clutter, where real-world images contain complex backgrounds, overlapping objects, and noisy elements that make it difficult for OCR algorithms to differentiate text from non-text regions. For instance, street signs with graffiti, shadows, or reflections can significantly hinder readability. Another challenge is the variation in font styles and text alignments, as OCR struggles with non-standard fonts, different character sizes, and diverse text orientations such as curved or slanted text. A common example is storefront signs that use stylized fonts, handwritten texts, or artistic lettering, making it difficult for OCR models to recognize and interpret the text correctly. Furthermore, distortions and deformations caused by camera angles, motion blur, and perspective warping also impact OCR performance. This is often observed in license plate numbers captured from a moving vehicle or documents photographed at an angle, leading to errors in text recognition. Additionally, **lighting conditions and low contrast** pose significant challenges for OCR systems. Poor lighting, uneven illumination, shadows, or low-contrast environments can reduce text clarity, making it harder for OCR to extract accurate information. For example, reading a receipt under dim lighting or extracting text from a faded printed document can result in incomplete or incorrect recognition. These limitations highlight the need for more advanced machine learning and deep learning approaches to improve text recognition under challenging real-world conditions [3]. Recent advancements in machine learning (ML) and deep learning (DL) have significantly improved text detection, localization, recognition, and tracking. Methods such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Transformer-based models have demonstrated superior accuracy in text recognition tasks. **These techniques enable better feature extraction, reduced false positives, and increased robustness in real-world scenarios.** Additionally, the integration of Natural Language Processing (NLP) techniques has further enhanced the accuracy of extracted text by improving contextual understanding. This paper presents a comprehensive literature review on text detection and recognition techniques in images and videos, focusing on the latest ML and DL approaches. The analysis explores various datasets, evaluation metrics, and challenges associated with text extraction. A comparative analysis of **traditional OCR-based methods and deep learning-based models is provided to highlight the strengths and limitations of each approach**. The findings emphasize the need for robust optimization techniques and scalable architectures to improve efficiency and real-time applicability in diverse domains.

2. Literature Review

Adem Akdogan et al.,(2025) [4] introduced ExTTNet, a deep learning-based model designed to autonomously extract product tables from invoices. Automated extraction of tabular data from invoices is a significant challenge in document processing. **Traditional Optical Character Recognition (OCR) systems struggle to accurately identify and structure tabular data** due to variations in invoice layouts, fonts, and image quality. The proposed method employs Tesseract OCR to extract raw text from invoice images, followed by feature extraction techniques to enhance accuracy. A multilayer artificial neural network is used to classify whether each extracted text element belongs to a table or not. The model was trained on an Nvidia RTX 3090 GPU, achieving a high F1 score of 0.92, demonstrating its effectiveness in accurately identifying table structures. Future improvements aim to incorporate non-textual features such as shapes, symbols, and barcodes into the deep learning model, **addressing limitations where OCR alone is insufficient.** Additionally, image preprocessing techniques such as noise reduction and distortion correction are being explored to enhance Tesseract's OCR accuracy, leading to better feature extraction and classification performance. These advancements are expected to improve the robustness of ExTTNet in handling diverse invoice formats.

Advantages: Achieves an impressive F1 score of 0.92, demonstrating strong performance in identifying tabular elements from invoices. The model can be improved by incorporating non-textual features like shapes, symbols, and barcodes, which are currently not utilized in OCR-based approaches. **Disadvantages:** The accuracy of ExTTNet is influenced by the effectiveness of Tesseract OCR, **which may struggle with noisy or distorted images**. Training the model required an Nvidia RTX 3090 GPU and took 162 minutes, making it resource-intensive and potentially limiting its accessibility for smaller organizations.

Hansi Seitaj et al.,(2024) [5] addresses the challenge of manual data extraction from product labels by integrating Computer Vision and Natural Language Processing (NLP) techniques. This research introduces an enhanced Convolutional Recurrent Neural Network (CRNN) model, combining Convolutional Neural Networks (CNNs) for feature extraction and Recurrent Neural Networks (RNNs) for sequential text recognition. This hybrid approach is further strengthened by incorporating the Tesseract OCR engine to improve Optical Character Recognition (OCR) tasks. The CRNN model is trained on encoded product labels and evaluated on a dedicated test set. The research also utilizes the Open Food Facts API for database population and text-only label prediction, ensuring a comprehensive and accurate analysis. A key application of this study is enhancing accessibility for visually impaired individuals by providing them with essential product information, such as ingredients and usage directions. This research highlights the potential of deep learning and OCR in automating label recognition, thereby advancing automated data extraction in critical sectors like healthcare and food safety. In terms of performance, the CRNN model initially achieved a validation accuracy of 47.06% over ten epochs, demonstrating the potential for improvement. **To optimize performance, a Gated Recurrent Unit (GRU) model was tested, showing a**

sharp increase in validation accuracy to 58.82% by the fifth epoch. This suggests that alternative deep learning architectures, such as GRUs and Transformers, could further enhance OCR accuracy. The research also explores the limitations of traditional OCR models, such as LSTM-based methods, which struggle with image variability, and Transformer-based models, which require significant computational resources. This research makes a substantial contribution to the field by demonstrating the effectiveness of a customized OCR system for product label extraction. It also highlights the importance of integrating multiple data sources such as Open Food Facts API and NLP techniques to enhance OCR efficiency. The findings emphasize the need for continued improvements in OCR architectures to tackle complex and diverse datasets, paving the way for future research in multilingual and domain-specific OCR applications.

Advantages: This research introduces a hybrid CNN-RNN approach, which improves text recognition accuracy by combining feature extraction and sequential modeling capabilities. The integration of Tesseract OCR and Open Food Facts API allows for automated data extraction from product labels, benefiting industries such as healthcare, food safety, and accessibility technologies. **Disadvantages: The CRNN model achieves only 47.06% validation accuracy, indicating that further enhancements in model architecture** and training strategies are necessary. While GRU-based models improve accuracy (58.82%), Transformer-based models demand significant computational resources, making real-time deployment challenging.

Houze Liu et al.,(2024) [6] proposed an optimized CNN-based approach utilizing AlexNet and InceptionV3 for pneumonia classification. The classification of pneumonia CT images using Convolutional Neural Networks (CNNs) has become a crucial area of research in medical image analysis. Traditional shallow neural networks often lack the complexity required for accurate pneumonia detection, while deeper networks demand substantial computational resources, making them impractical for deployment on edge devices. The research highlights the use of knowledge extraction technology to enhance the AlexNet model, reducing computational costs while maintaining high diagnostic accuracy. The improved AlexNet_S model achieved better performance than InceptionV3, with notable increases in prediction accuracy (+4.25%), specificity (+7.85%), and sensitivity (+2.32%). Additionally, GPU (Graphics Processing Unit) usage decreased by 51% compared to inceptionV3, making it a more efficient solution for real-world medical applications. **This research underscores the importance of optimizing CNN architectures for medical image classification**, particularly in resource-constrained environments. The proposed AlexNet_S model demonstrates that deep learning-based diagnostic tools can be both computationally efficient and highly accurate. The findings advocate for continued advancements in lightweight yet effective CNN models, potentially improving accessibility to AI-driven diagnostic imaging systems in healthcare.

Advantages: The optimized AlexNet_S model outperforms InceptionV3, achieving notable improvements in accuracy (+4.25%) and specificity (+7.85%), leading to more reliable pneumonia detection. By integrating knowledge extraction techniques, GPU usage is reduced by 51%, making the model more efficient and deployable on low-resource medical devices. **Disadvantages:** The model is specifically optimized for pneumonia classification, and **its performance across other medical conditions remains unexplored.** The accuracy improvements rely heavily on well-annotated CT images, making the approach susceptible to variations in data quality and availability in different clinical settings.

Alexander Rombach et al.,(2024) [7] provides a systematic literature review of 96 approaches published between 2017 and 2023, analyzing state-of-the-art techniques and identifying potential research directions. **Key Information Extraction (KIE) is a crucial component of Document Understanding,** a field that has gained significant momentum due to advancements in Deep Learning. Recent progress in KIE has enabled the automatic processing of visually rich business documents, which often contain complex layouts and high information density. This research categorizes existing methodologies into three primary paradigms for document representation:

- Sequence-based models (e.g., Transformer-based architectures such as BERT)
- Graph-based models (Capturing document structure relationships)
- Grid-based models (Leveraging spatial features)

A key observation is that OCR-independent and autoregressive approaches have been introduced, allowing models to process document text without an external OCR engine. This enhances flexibility and broadens the range of supported downstream tasks. Furthermore, pre-training and fine-tuning strategies play a critical role in developing generalizable deep learning models for KIE. However, this research also highlights limitations in current research trends. Benchmark-driven improvements (e.g., achieving high F1-scores) dominate the field, but the heterogeneous evaluation setups across studies make direct performance comparisons challenging. The authors advocate for shifting focus towards real-world applicability, lightweight models, data-efficient training techniques, and the integration of domain knowledge. Future research should explore more diverse datasets and standardized evaluation frameworks to facilitate a more meaningful comparison of KIE approaches.

Advantages: Deep learning-based KIE models can automatically extract key information from complex business documents, significantly improving efficiency in document-intensive workflows. The development of OCR-independent models enables direct document processing, while lightweight architectures enhance practical deployment on resource-constrained systems. **Disadvantages: Due to heterogeneous evaluation setups, it is difficult to compare different KIE models fairly, leading to challenges in assessing real-world effectiveness.** Most deep learning-based KIE models require extensive labeled data for training, which may not always be available, limiting their adaptability to new document formats and languages.

Tien Do and Thuyen Tran Doan et al.,(2024) [8] propose a transformer-based method leveraging RoBERTa, a pre-trained language model, combined with the LION optimizer to enhance the training process and boost accuracy. **The extraction and recognition of key information from rich text images play a fundamental role in document analysis and natural language processing (NLP) applications**. Two primary tasks in this domain include Line Item Recognition (LIR) and Key Information Localization and Extraction (KILE). LIR focuses on identifying and interpreting structured text data, while KILE involves classifying or extracting essential elements from documents. Traditional approaches to these tasks often rely on sequence-based models, which require large training datasets and extensive computational resources. The proposed model is evaluated on two different benchmarks: DocILE (English) and MCOCR (Vietnamese), demonstrating notable improvements in information extraction tasks. Key findings from the research include:

- A 7.24% accuracy improvement in KILE compared to the baseline on the DocILE dataset.
- A Character Error Rate (CER) of 28.6% on the MCOCR dataset, achieving second place in the competition.
- Improved recognition rates in LIR, making it a reliable solution for document-based text extraction.

The research highlights the effectiveness of RoBERTa in multilingual document processing, emphasizing its adaptability to different languages and structured/unstructured document formats. The integration of the LION optimizer further enhances the training efficiency, making it suitable for large-scale text extraction tasks in industries like finance, healthcare, and legal documentation. Despite these advancements, the model still encounters challenges related to handling handwritten and highly complex document structures, suggesting that future work should focus on improving model robustness and reducing computational complexity.

Advantages: The model achieves a 7.24% improvement in the KILE task and competitive performance in LIR, demonstrating superior accuracy compared to existing methods. The RoBERTa-based approach effectively generalizes across different languages, making it adaptable for diverse document analysis applications. **Disadvantages:** The transformer-based approach requires substantial computational resources and training time, which may not be feasible for all applications. **While effective for structured text, the model struggles with noisy, handwritten, or highly unstructured documents,** affecting real-world applicability.

Noura A. Semary et al.,(2024) [9] systematically analyze different feature extraction techniques from a machine learning perspective to guide future research in sentiment analysis. Feature extraction is a critical component of sentiment classification, as it plays a fundamental role in extracting valuable information from text data, directly impacting model performance. The selection of an appropriate feature extraction technique is essential to improve sentiment analysis accuracy and efficiency. The research evaluates six feature extraction methods, including Bag-of-Words (BoW), Word2Vec, N-gram, Term Frequency-Inverse Document Frequency (TF-IDF), Hashing Vectorizer (HV), and GloVe. The experiments are conducted on two benchmark datasets: Twitter US Airlines and Amazon Musical Instrument Reviews. A random forest classifier is trained using 70% of the dataset for training and 30% for testing, and the performance is measured using accuracy and other evaluation metrics. The results demonstrate that TF-IDF achieves the highest performance, with 99% accuracy on the Amazon dataset and 96% accuracy on the Twitter dataset. This highlights the effectiveness of TF-IDF in sentiment analysis, particularly for social media data. This research also emphasizes the importance of preprocessing and the Synthetic Minority Over-sampling Technique (SMOTE) in enhancing classification results. Additionally, **the research underscores the importance of feature selection in machine learning-based sentiment analysis**, as different techniques exhibit varying levels of performance. The findings suggest that TF-IDF is a highly effective method, offering better training efficiency and prediction accuracy compared to other methods. These insights provide practical recommendations for researchers and practitioners developing sentiment analysis models for social media and e-commerce reviews.

Advantages: The research shows that TF-IDF achieves superior accuracy (up to 99%) compared to other feature extraction methods in sentiment analysis. The random forest classifier combined with TF-IDF offers faster training and prediction times, making it suitable for real-time applications. **Disadvantages: The research focuses on only two datasets, which may not represent the performance of feature extraction techniques on other domains** or languages. The effectiveness of feature extraction is heavily reliant on proper preprocessing steps, making it less robust to noisy or unstructured data.

Ivan Malashin et al.,(2024) [10] propose an integrated approach that combines image recognition and natural language processing (NLP) techniques, specifically named entity recognition (NER), for efficient text extraction from scanned medical documents and photographs. Automated information extraction (IE) from medical reports is an essential task in healthcare and financial sectors. **This research utilizes a genetic algorithm (GA) to optimize optical character recognition (OCR) hyperparameters, ensuring maximum text extraction** accuracy. The extracted text is then processed through NER to categorize key entities such as organization identification numbers, tax identification numbers (TINs), license numbers, payment amounts, and payment dates. This adaptive approach dynamically adjusts OCR parameters, refining entity extraction based on manual annotations. Despite the dataset being limited to Russian medical reports, the proposed IE model can be extended to other languages. This research demonstrates that automating document processing can reduce manual labor, enhance efficiency, and improve data accuracy, especially in domains dealing with large volumes of unstructured documents, such as tax authorities and medical organizations. While the research highlights the potential of adaptive models in OCR-based IE, it also acknowledges challenges such as entity recognition inaccuracies and handling complex document formats. Future advancements should focus on refining NLP techniques, integrating machine learning (ML) algorithms for document classification, and improving feature extraction accuracy. These improvements will further streamline operational workflows and enhance document processing automation. **Advantages:** The adaptive OCR model with NER minimizes manual effort, improving data extraction accuracy and reducing errors. The proposed approach can be applied to various domains (e.g., healthcare, taxation, and finance) for processing unstructured documents efficiently. **Disadvantages: The NER model may struggle with complex document formats, leading to misclassification** or missing key entities. This research was limited to Russian medical reports, requiring further adaptation and retraining for use in other languages and document structures.

Guangyun Lu et al.,(2024) [11] propose a novel approach that combines Mask R-CNN, DCGAN, and ALBERT models to enhance fine-grained recognition and contextual understanding in image-text processing. The integration of artificial intelligence (AI) in image and text recognition has significantly evolved, leading to advancements in multimodal fusion models. Their method improves image segmentation, feature extraction, and deep semantic interpretation, making it suitable for applications like content creation, accessibility technologies, and virtual assistants. The Mask R-CNN model is used for high-precision image recognition and segmentation, while DCGAN (Deep Convolutional Generative Adversarial Network) generates nuanced image features, and ALBERT is responsible for deep natural language understanding. Experimental results show an increase in recognition accuracy from 85.3% to 92.5%, along with improved contextual and situational understanding. Furthermore, the ALBERT-DCGAN-Mask R-CNN model enhances image-text alignment accuracy by 12%, achieving an F1 score of 0.85 in graphic understanding tasks, significantly outperforming baseline models with 0.78. It also improves precision in automatic image description (0.87 vs. 0.80 in other models) and boosts recall for intelligent search applications by 15%. **This research highlights that DCGAN effectively processes large-scale unlabeled data, enhancing the overall robustness of the model.** The combination of ALBERT's natural language processing and Mask R-CNN's object detection improves interpretability and reliability in multimodal applications. These findings emphasize the importance of advanced AI models for efficient graphical information sharing and processing, contributing to the broader field of machine vision and natural language processing (NLP). Despite these advancements, the research also identifies key challenges such as high computational resource consumption and reliance on large-scale labeled data, which affect the model's generalization in real-world applications. Future research will focus on reducing dependency on labeled data through transfer learning and weakly supervised learning strategies, improving model efficiency, and enhancing interpretability for better decision-making transparency. These improvements aim to create more scalable and efficient multimodal AI applications. **Advantages:** The model improves image-text alignment accuracy by 12% and increases recognition accuracy from 85.3% to 92.5%, making it superior to traditional techniques. The integration of DCGAN handles large-scale unlabeled data efficiently, enhancing model adaptability and reliability for real-world applications. **Disadvantages:** The deep learning-based approach requires significant computational resources, making it less accessible for low-power or resource-limited systems. **The model relies heavily on labeled datasets, which can be costly and time-consuming to obtain,** limiting its generalization to new domains.

Robert West et al.,(2024) [12] investigated the use of ML for extracting information from reports of randomized trials on smoking cessation interventions. This research, as part of the Human Behaviour-Change Project, aimed to assess the effectiveness of ML algorithms in processing study reports and predicting smoking cessation outcomes. The researchers manually annotated 512 reports, identifying 70 key entities related to intervention content, delivery, population, setting, outcome, and study methodology using the Behaviour Change Intervention Ontology. A named-entity recognition system, built using the 'FLAIR' framework, was employed to train ML algorithms for automatic information extraction. Additionally, a deep-learning model based on Long Short-Term Memory (LSTM) layers was developed to predict smoking cessation outcomes. However, the ML algorithm had limited success, with an F1 score of 0.42 on average, compared to a human annotator's score of 0.75. **The algorithm also struggled to assign entities to study arms (intervention vs. control) and failed to outperform** traditional linear regression or mean outcome-based prediction models. The research highlights key challenges in ML-based information extraction and prediction, including inconsistencies in study reporting and the need for standardized research documentation. **The authors suggest that future advancements should focus on developing semantically aware and interpretable ML** architectures that integrate domain knowledge with predictive modeling. Standardization in study reports, improved annotation tools, and novel ML methodologies could enhance the effectiveness of ML in behavioral science research. **Advantages:** ML can process large datasets of research reports efficiently, reducing manual effort and enabling quicker data analysis. With advancements in ML and ontological integration, predictive models could be refined to provide more accurate and actionable insights for behavioral interventions. **Disadvantages: The ML algorithm performed significantly worse than human annotators, indicating challenges** in correctly identifying and categorizing study entities. The lack of standardized reporting formats hindered ML effectiveness, highlighting the need for structured and machine-readable study documentation.

M Mahfi Nurandi Karsana et al.,(2023) [13] explores the performance of different ML models for single-label and multi-label text classification, comparing Artificial Neural Networks (ANN), Naïve Bayes, and Support Vector Machines (SVM). Machine Learning (ML) has become an essential tool in various applications, with text classification being one of its key domains. ANN, inspired by biological neural networks, has been widely used for classification tasks due to its ability to approximate complex functions. **The research uses a simple ANN architecture to evaluate its effectiveness in comparison to SVM and Naïve Bayes**. Performance is measured using the F1-macro score, with experiments conducted using K-Fold Cross-Validation on multiple datasets, including 20 Newsgroups (large single-label dataset) and BBC (smaller single-label dataset). The results indicate that SVM achieves the highest F1-macro score (0.82) in single-label classification, while ANN follows closely with 0.79. However, on a smaller dataset (BBC), all models perform exceptionally well, with Naïve Bayes and SVM scoring 0.97, and ANN scoring 0.96. The research also highlights the challenges of multi-label classification, where performance drops significantly. ANN demonstrates the best performance with an F1-macro score of 0.48, outperforming SVM (0.44) and Naïve Bayes (0.34). Additionally, the Precision-Recall tradeoff is analyzed, showing that SVM has high recall (0.72) but low precision (0.33), indicating a tendency for false positives. In contrast, ANN and Naïve Bayes exhibit a more balanced Precision-Recall performance. **The research emphasizes the potential of ANN in text classification, particularly in multi-label scenarios,** where it outperforms traditional models. The findings suggest that while SVM excels in single-label classification, ANN is a strong contender for complex classification tasks, showcasing its adaptability and robustness in text-based ML applications. **Advantages:** ANN outperforms both SVM and Naïve Bayes in multi-label text classification, making it a suitable choice for complex classification problems. Unlike SVM, which has a high recall but low precision, ANN maintains a more balanced trade-off, reducing false positives and improving classification reliability. **Disadvantages: ANN models, even simple architectures, require more computational resources compared to traditional models like Naïve Bayes,** making them less efficient for large-scale deployment. While ANN performs well, SVM achieves higher F1-macro scores in single-label classification, making it a better choice for tasks where precision is a priority.

Gagandeep Kaur et al.,(2023) [14] proposes a consumer review summarization model using Natural Language Processing (NLP) and Long Short-Term Memory (LSTM) networks. In the era of exponential textual content generation, platforms such as social media, messaging apps, e-commerce, and news websites continuously produce vast amounts of unstructured text data. **Analyzing this data can provide valuable insights into consumer sentiment, helping businesses understand** public perception and improve their products or services. The research introduces a hybrid approach for sentiment analysis, consisting of three main stages:

- Pre-processing – Eliminates undesirable data from raw text reviews.
- Feature Extraction – Utilizes a hybrid feature vector (HFV) combining review-related and aspect-related features to enhance sentiment classification.
- Sentiment Classification – Implements LSTM-based deep learning for improved sentiment detection.

The experimental evaluation of the proposed model is conducted on three different research datasets, achieving high performance with an average precision of 94.46%, recall of 91.63%, and F1-score of 92.81%. **This research is further highlights that hybrid feature extraction significantly enhances sentiment analysis performance** compared to traditional methods. The research also addresses various research challenges in sentiment analysis, such as handling raw online reviews, feature selection, and robust classification methods. The authors compare their model's results with state-of-the-art sentiment analysis techniques, demonstrating improvements in F1-score, accuracy, precision, recall, and AUC (Area under the Curve). Overall, this research contributes to the field by enhancing sentiment classification through hybrid feature extraction and deep learning, paving the way for more accurate, reliable, and scalable consumer sentiment analysis models. **Advantages:** The proposed hybrid feature extraction and LSTM-based classification improve precision (94.46%), recall (91.63%), and F1-score (92.81%), outperforming traditional approaches. The model can be applied to business intelligence, brand reputation analysis, and e-commerce platforms, helping companies make data-driven decisions. **Disadvantages:** The model requires high-quality, labeled datasets for training, which can be time-consuming and resource-intensive to acquire. **The model's effectiveness is currently limited to specific languages, requiring further enhancements** to support multilingual sentiment analysis.

Kitti Szabó Nagy et al.,(2023) [15] introduces a novel feature extraction method, TwIdw (Term Weight–Inverse Document Weight), which extends the traditional TfIdf approach by incorporating word depth within documents instead of term frequency. **Fake news classification has become a critical challenge in the field of Natural Language Processing (NLP).** Researchers have explored various approaches, including feature-based methods, deep learning models, and hybrid techniques, to improve the accuracy of detecting misleading information. Traditional methods like Term Frequency-Inverse Document Frequency (TfIdf) have been widely used for feature extraction, but they often fail to capture deeper semantic and contextual relationships in text. **This method is designed to enhance feature representation and improve fake news classification performance.** The effectiveness of TwIdw is evaluated using three different datasets, comparing it against basic TfIdf in combination with Random Forest and Feedforward Neural Networks (FNN). Key findings from the research include:

- TwIdw significantly outperforms basic TfIdf in terms of precision, recall, and F1-score.
- Feedforward Neural Networks (FNN) performs better than Random Forest, particularly when applied to the KaiDMML dataset, with an accuracy improvement of up to 3.9%.

- TwIdw improves classification accuracy for all datasets when combined with FNN, while the Random Forest method only shows improvements with the KaiDMML dataset (1%).
- Experimental results indicate that feature selection, dataset choice, and classification algorithms play a crucial role in determining the model's performance.

The research concludes that TwIdw is a promising technique for fake news detection, offering better accuracy and generalizability than traditional feature extraction methods. However, this research acknowledges limitations, particularly the limited availability of manually annotated unstructured text datasets. Future research directions include further optimizing TwIdw, expanding the dataset, and applying the technique to a broader range of NLP applications. **Advantages:** The TwIdw method enhances traditional TfIdf by incorporating word depth, leading to better classification accuracy and stronger contextual understanding. The technique improves precision, recall, and F1-score when used with Feedforward Neural Networks, making it adaptable for real-world fake news classification. **Disadvantages: The approach relies on labeled datasets, which are scarce (limited) and time-consuming to create, limiting its scalability**. While FNN benefits from TwIdw, Random Forest shows only minor improvements, suggesting that the method's effectiveness depends on the chosen classification model.

Xiujuan Wang et al.,(2023) [16] explores a DL-based Chinese text information extraction model for analyzing coastal biodiversity data in China. Text information extraction using Deep Learning (DL) has gained significant attention due to its ability to handle large volumes of unstructured data with improved accuracy and efficiency. Traditional text recognition and information extraction methods often rely on syntax analysis and **rule-based algorithms, which may struggle with complex linguistic structures and semantic variations in large datasets**. Nearly 500 species of coastal plants and seaweeds were collected, and their textual descriptions, including species morphology, habitat distribution, and resource value, were extracted. The authors propose an extraction model utilizing Long Short-Term Memory (LSTM) and Multiscale Fully Connected Neural Networks (L-MFCNN) for short text classification. Key findings from this research include:

- DL-based models outperform traditional text extraction methods in terms of accuracy, precision, recall, and F1-score.
- The neural network model achieves a high recognition accuracy of 96.69%, demonstrating its effectiveness in classifying and identifying species categories.
- Increasing the amount of training data enhances model convergence and reduces reconstruction errors, leading to faster and more accurate extractions.
- Compared to conventional text recognition algorithms, the proposed DL-based approach significantly improves extraction efficiency and reduces failure rates.

There research highlights the adaptability of DL for various text information extraction tasks, proving its scalability and robustness. Furthermore, the authors predict that as big data grows, DL architectures will evolve, becoming more complex and integral to future innovations in biodiversity research and environmental monitoring. **Advantages:** The DL-based model achieves 96.69% accuracy, significantly outperforming traditional text extraction methods. The model's performance improves with increased training data, making it highly adaptable for big data applications in biodiversity research. **Disadvantages: Training deep learning models requires significant computational resources, which can be expensive and time-consuming**. The model's performance heavily depends on large, high-quality labeled datasets, which may not always be available or easy to obtain.

Sunil Kumar Dasari et al.,(2023) [17] introduces an optimal architecture fusion neural network (FNN), which integrates CNN and RNN for text identification and recognition. Text recognition is a crucial task in computer vision and natural language processing (NLP), widely used in automatic translation, navigation systems, and aiding visually impaired individuals. Traditional text recognition methods rely on rule-based approaches, while recent advancements have shifted towards deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The CNN layers extract textual features from images, while the RNN layers handle sequence prediction and classification. The model is trained using the Devanagari MLT-19 and RRC-MLT-2019 datasets and evaluated on three key metrics:

- Script word identification
- Character Recognition Rate (CRR)
- Word Recognition Rate (WRR)

The results indicate that the FNN achieves 98.67% script identification accuracy, 92.93% CRR, and 84.65% WRR, demonstrating superior performance compared to existing deep-learning-based text recognition models. The comparative analysis highlights a marginal improvement over previous methodologies, making FNN an effective approach for scene text recognition. However, **the research also identifies limitations the model performs well under controlled conditions but struggles with distorted or low-quality images**. Additionally, dataset availability and diversity remain a challenge, suggesting that future research should focus on improving robustness against real-world image distortions. **Advantages:** The FNN model achieves 98.67% script identification accuracy, demonstrating significant improvements over previous methods. The integration of CNN for feature extraction and RNN for sequence prediction improves text recognition efficiency and accuracy. **Disadvantages: The model struggles with text recognition in distorted, noisy, or naturally degraded images, limiting its real-world applicability**. The effectiveness of the model depends on well-structured datasets, making it less adaptable to diverse, unstructured text images.

Cong Tran et al.,(2022) [18] propose a Graph Convolutional Network (GCN)-based approach for information extraction from Vietnamese invoices. Information extraction (IE) plays a critical role in automatically retrieving structured information from unstructured or semi-structured documents. Traditional IE methods focus on tagging words and storing them as key-value pairs, which are later processed in databases. However, extracting key information from invoices and receipts presents additional challenges due to the layout-based structure of such documents. **In this context, advanced machine learning techniques, particularly graph-based neural networks, have been introduced to improve extraction accuracy**. This research integrates Optical Character Recognition (OCR) techniques (VietOCR) with GCN models to improve the accuracy of text recognition and structured data extraction. The method was evaluated on 731 invoices collected from Vietnamese supermarkets, where text detection, recognition, and feature processing were performed before classification. The results demonstrate that the proposed model achieves high accuracy (99.50%), recall (98.52%), precision (98.52%), and F1-score (98.52%), proving its effectiveness in extracting invoice details. The research also highlights the importance of automatic invoice direction detection for precise adjustments, making the approach more scalable for diverse invoice formats. Despite being primarily trained on G7 store invoices, the method is expected to generalize to other invoice templates, making it valuable for future large-scale applications in financial and retail sectors. While the research presents a robust and high-performing system, it also acknowledges certain limitations, such as specialization to a specific dataset and potential challenges in scaling to highly diverse invoice formats. This research suggests that future work should extend the model's adaptability to a broader range of invoices and enhance data augmentation techniques to improve robustness. **Advantages:** The proposed GCN-based model achieves 99.50% accuracy, significantly improving OCR-based invoice processing. The model includes automatic invoice direction detection, allowing precise adjustments and making it scalable for larger datasets and diverse invoice templates. **Disadvantages: The model was trained primarily on G7 store invoices, making it less adaptable to different invoice formats** without further training. The integration of VietOCR and GCN models requires substantial computational power, which may limit its deployment on low-resource devices.

Qing Kuang et al.,(2021) [19] explore local and global feature extraction methods in facial recognition using a deep learning-based approach. Facial recognition has become a core application of artificial intelligence, leveraging deep learning to enhance feature extraction and classification. Their study introduces a training classification method that integrates a local pattern and Global Local Quantized Pattern (GLQP) representation with a Convolutional Neural Network (CNN) model. The research employs local quantization techniques to preprocess data, which is then filtered and input into a four-layer deep CNN model. The trained network achieves an accuracy of 92.2% on the test dataset, demonstrating the effectiveness of deep learning in automatic face feature extraction. This research highlights that deep learning overcomes limitations in shallow feature learning, leading to higher recognition efficiency and generalization compared to traditional methods. Moreover, the research emphasizes that deep learning models enhance feature extraction capabilities by providing rich hierarchical structures that capture high-resolution features. The local and global feature extraction combination further improves the robustness and precision of facial recognition systems. These findings suggest that deep learning offers superior visualization, automatic feature learning, and higher classification accuracy, making it a promising approach for biometric authentication and security applications. **Advantages:** The proposed deep learning model achieves 92.2% accuracy, outperforming traditional facial recognition methods. The CNN model automatically extracts local and global features, reducing the need for manual feature engineering. **Disadvantages:** The deep learning model requires high computational power due to its four-layer CNN architecture, making it less suitable for low-resource devices. **The performance of the model is highly reliant on large, high-quality datasets, which may limit its effectiveness** in real-world scenarios with insufficient training data.

Chaitra Yuvaraj Lokkondra et al.,(2021) [20] this researcher focus images and videos serve as direct sources of information, and the need for intelligent analysis of such data is increasing rapidly. Text extraction and recognition from images and videos have gained significant attention in recent years due to their applicability in various domains. Researchers have focused on text identification, making it a key area of study in machine vision. The process of extracting text from images and videos involves multiple stages, including text detection, localization, recognition, and tracking. **The research presents a dataset for text identification, describes various text formats found in images and videos, and explores existing machine** learning and deep learning techniques used for training models. Furthermore, it discusses different evaluation metrics used to validate model performance. Challenges such as capturing techniques, lighting conditions, environmental factors, text alignment, font variations, and background noise significantly impact the accuracy of text extraction. Text extraction plays a crucial role in applications like license plate recognition, vehicle tracking, industrial automation, and self-navigating automobiles. The extracted text can be used for analysis, documentation, preservation, and evaluation. Despite advancements, extracting text from images and videos remains complex due to diverse text formats, varying dimensions, low-quality images, and occlusions. This research highlights the need for robust mathematical models and evaluation techniques to improve accuracy and efficiency in text recognition systems. **Advantages:** Text extraction enables automatic data retrieval from images and videos, reducing manual efforts and enhancing efficiency in various industries such as transportation, healthcare, and document digitization. The technology is applicable in multiple domains, including license plate recognition, industrial automation, and self-navigating vehicles, contributing to advancements in smart systems and automation. **Disadvantages:** Factors like lighting conditions, environmental variations, different font styles, and background clutter make text extraction challenging, affecting accuracy. **Processing high-resolution images and videos for text extraction requires significant computational resources**, making it difficult to implement in real-time applications with limited hardware.

Nadeesha Perera et al.,(2020) [21] review the current methodologies for Named Entity Recognition (NER) and Relation Detection (RD), which are essential for extracting relationships between biomedical entities such as proteins, drugs, genes, and diseases. The rapid growth of scientific literature in biomedical, health, and clinical sciences has led to an increasing demand for automatic information extraction techniques. Traditional approaches fail to efficiently archive and retrieve valuable knowledge, leaving vast amounts of critical information buried in textual data. To address this challenge, Natural Language Processing (NLP) and text mining methods are employed to automatically extract, classify, and structure key biomedical data. These techniques are used to build structured networks of biomedical knowledge, making large-scale data more accessible for further analysis and decision-making. **A key aspect of this research is the integration of deep learning approaches into NER and RD tasks.** This research categorizes various methods based on their characteristics and highlights recent advancements in deep learning, which have significantly improved the accuracy and efficiency of entity recognition and relation extraction. These deep learning methods offer a new perspective on information retrieval and are expected to drive future developments in biomedical text mining. While deep learning has enhanced NER and RD performance, challenges remain in handling complex sentence structures, domain-specific terminology, and interpretability of models. The paper suggests that continued improvements in artificial intelligence and NLP will further enhance automated biomedical knowledge extraction and analysis. **Advantages:** The research highlights deep learning methods that significantly improve the accuracy of Named Entity Recognition (NER) and Relation Detection (RD), making biomedical data more structured and usable. By extracting relationships between proteins, drugs, genes, and diseases, the methods enable the creation of knowledge graphs and structured databases, improving data management and accessibility for researchers. **Disadvantages: Biomedical NLP requires specialized training datasets and domain knowledge, making it difficult to generalize models across different biomedical subfields**. Deep learning-based approaches require high computational resources and large annotated datasets, which limit their scalability in real-world applications.

Philomina Simon et al.,(2020) [22] propose a hybrid texture classification approach that leverages CNN-based feature extraction combined with Support Vector Machine (SVM) classification. Texture classification is a fundamental task in computer vision and pattern recognition, widely applied in industrial automation, defect detection, and visual inspection. Traditional approaches rely on handcrafted texture features or texture descriptors, which have proven effective but often require domain expertise and manual feature selection. Recent advances in deep learning have introduced Convolutional Neural Networks (CNNs) as powerful feature extractors, significantly improving classification accuracy. This research investigates the performance of CNN features extracted from pretrained models such as DenseNet201, ResNet50, ResNet101, InceptionV3, AlexNet, VGG19, and InceptionResNetV3. **The classification process uses cross-entropy loss to estimate errors during training.** Experimental evaluations on gray and color texture datasets, including KTH-TIPS, CURET, and flower datasets, demonstrate high accuracy ranging from 85% to 95%, with reduced computation time. The findings highlight that CNN-based feature extraction outperforms traditional handcrafted features, providing rich and meaningful representations for texture classification. Additionally, the study confirms that SVM is an effective classifier, achieving optimal performance when combined with CNN-extracted features. The research also emphasizes that deep learning requires a large amount of training data and is memory-intensive, posing challenges for practical deployment in resource-constrained environments. Despite these limitations, the proposed method demonstrates superior performance in various datasets, establishing CNN-SVM as a robust approach for industrial and real-world texture classification applications. Future research may focus on reducing memory consumption, optimizing training efficiency, and incorporating transfer learning techniques to address data limitations.

Advantages: The hybrid CNN-SVM approach achieves 85%-95% accuracy across multiple datasets, demonstrating superior performance over traditional handcrafted feature methods. Unlike end-to-end deep learning models, using CNN only for feature extraction and SVM for classification optimizes computational efficiency, making the approach more practical for real-time industrial applications. **Disadvantages: Deep learning-based feature extraction requires significant computational resources, making it challenging for deployment in low-power or edge devices**. CNN-based models require extensive training datasets to generalize effectively, which can be difficult to obtain in specialized industrial settings.

Vaibhav Goel et al,(2019) [23] propose a Convolutional Neural Network (CNN)-based technique for efficient text extraction from natural scene images. With the rapid proliferation of camera-embedded smartphones, computer vision tasks such as text extraction from natural scene images have gained significant attention. Unlike structured documents, natural scene images contain text with random backgrounds, diverse color schemes, and varying font styles, making text detection and extraction challenging. Traditional methods for optical character recognition (OCR) often struggle with such unstructured environments, necessitating advanced deep learning approaches. Their method combines Open Source Computer Vision Library (OpenCV) with CNNs in a two-stage pipeline that eliminates unnecessary intermediate steps, thereby improving time complexity and accuracy. **The model specifically focuses on horizontally aligned text, employing the EAST (Efficient and Accurate**

Scene Text Detector) for text detection, followed by another CNN model for recognition. Proper preprocessing of the images before feeding them into the network significantly improves performance. The proposed technique has several real-world applications, including document retrieval, vehicle number plate detection, road sign recognition for smart cars, intelligent driving assistance, video content summarization, and aiding visually impaired individuals. Despite its effectiveness, this research highlights that no single algorithm can generalize across all types of natural images, as different techniques have unique strengths and limitations. This research suggests further improvements in handling multi-oriented and curved text to enhance the robustness of text extraction models.

Advantages: The model eliminates unnecessary intermediate steps, improving speed and accuracy in text extraction from natural images. The technique is useful in document processing, autonomous driving, smart navigation, and assistive technologies for the visually impaired. **Disadvantages:** The model performs well only on horizontally aligned text, struggling with multi-oriented or curved text. **The CNN model requires proper preprocessing of images before training, making it less effective in low-quality or noisy images**.

Ruishuang Wang et al.,(2019) [24] explore the effectiveness of Convolutional Neural Networks (CNN) and Convolutional Recurrent Neural Networks (CRNN) for text feature extraction and compare their performance with TF-IDF and Word2Vec. Deep learning has significantly enhanced natural language processing (NLP), particularly in text classification tasks. Traditional methods such as TF-IDF and Word2Vec have been widely used for feature representation, but they often fail to capture deep semantic relationships in textual data. This research establishes CNN and CRNN-based feature extraction models to extract high-level textual features and evaluates their effectiveness on a Chinese academic paper dataset. These extracted features are then classified using Support Vector Machine (SVM) and Random Forest classifiers. Experimental results indicate that the CNN and CRNN models outperform traditional feature extraction techniques, achieving higher classification accuracy. Furthermore, SVM and Random Forest classifiers exhibit better classification performance compared to standalone neural networks, demonstrating their effectiveness in combining deep learning-based feature extraction with traditional machine learning classifiers. The findings suggest that CNN and CRNN effectively capture semantic representations of text, leading to superior classification accuracy**. This highlights the potential of hybrid approaches combining deep learning-based feature extraction with traditional classifiers for text classification tasks in NLP**. Future research could focus on optimizing computational efficiency, reducing training time, and exploring attention mechanisms to further enhance performance. **Advantages:** CNN and CRNN extract high-level semantic features from text, providing more accurate and meaningful text representations compared to traditional methods like TF-IDF and Word2Vec. The combination of deep learning-based feature extraction (CNN & CRNN) with the SVM and Random Forest classifiers leads to superior classification performance and also outperforming standalone neural networks. **Disadvantages:** CNN and CRNN require significant processing power and memory, making them computationally expensive for large-scale datasets. **The performance of deep learning models depends on large amounts of labeled training data,** which may not always be available for specialized domains like Chinese academic paper classification.

Zhi Tian et al.,(2016) [25] proposed a Connectionist Text Proposal Network (CTPN) to improve the accuracy of text localization. Unlike traditional bottom-up approaches that require multiple post-processing steps, the CTPN offers an end-to-end trainable model that directly detects text lines in convolutional feature maps. The model introduces a vertical anchor mechanism, which predicts the location and text/non-text classification of each fixed-width proposal. By incorporating a Recurrent Neural Network (RNN) layer into the convolutional network, the CTPN can naturally connect sequential text proposals, enabling it to explore contextual information effectively. **This makes it particularly robust in detecting complex, multi-scale, and multi-language text.** The model was evaluated on the ICDAR 2013 and 2015 benchmarks, achieving F-measures of 0.88 and 0.61, respectively, surpassing previous state-of-the-art methods. The computational efficiency of 0.14s per image, achieved using the deep VGG16 model, demonstrates its real-time applicability. The results highlight the CTPN's advantages in accurately localizing text in challenging environments without requiring additional post-processing. **Advantages:** Unlike traditional multi-step text detection methods, CTPN integrates proposal generation and classification into a single framework, improving efficiency and accuracy. The incorporation of an RNN layer helps in connecting sequential text proposals, making the model highly effective for detecting ambiguous and multi-scale text. **Disadvantages:** The CTPN leverages the VGG16 model, which adds computational complexity and reliance on pretrained weights, making fine-tuning challenging. While achieving strong results on ICDAR benchmarks, **the model's performance may degrade in real-world scenarios involving extreme variations in lighting, distortions, and occlusions**.

3. Conclusion

This literature review has explored various traditional and modern techniques, emphasizing the transition from Optical Character Recognition (OCR) methods to deep learning-based approaches. While traditional OCR methods have been widely used, their limitations in handling background clutter, font variations, distortions, and lighting conditions necessitate more advanced methodologies. The integration of machine learning and deep learning techniques, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Transformer-based models, has demonstrated superior accuracy and efficiency in text recognition tasks. **These models improve feature extraction, contextual**

understanding, and robustness against real-world challenges, making them suitable for applications in license plate recognition, industrial automation, self-navigating vehicles, and medical imaging. Despite these advancements, several challenges remain, including computational resource constraints, dataset availability, and real-time processing limitations. **Future research should focus on optimizing deep learning architectures, developing lightweight models, and improving text recognition in complex environments**. Additionally, enhancing evaluation metrics and standardizing datasets will further refine model performance and applicability. In conclusion, text detection and recognition continue to be an evolving field with promising advancements. **The ongoing development of AI-driven solutions and optimization techniques will play a crucial role** in making these technologies more accessible, efficient, and adaptable across various domains.

References

[1] U. Pal, A. Halder, P. Shivakumara, and M. Blumenstein, "A Comprehensive Review on Text Detection and Recognition in Scene Images," *Artificial Intelligence Advances*, Vol.**4**, No.**2**, pp.**1–20, 2024.**

[2] Y. Zhang et al., "Text Detection and Recognition Based on a Lensless Imaging System," *arXiv preprint arXiv:* Vol.**22**, Issue.**10,** pp.**4244, 2022.** https://arxiv.org/abs/2210.04244.

[3] Z. Cheng, J. Lu, Y. Niu, S. Pu, F. Wu, and S. Zhou, "You Only Recognize Once: Towards Fast Video Text Spotting," *arXiv preprint arXiv:1903.03299*, **2019**. https://arxiv.org/abs/1903.03299.

[4] Adem Akdoğan, Murat Kurt,"ExTTNet: A Deep Learning Algorithm for Extracting Table Texts from Invoice Images",Computer Vision and Pattern Recognition, **2025**, https://doi.org/10.48550/arXiv.2402.02246

[5] Hansi Seitaj and Vinayak Elangovan, "Information Extraction from Product Labels: A Machine Vision Approach", International Journal of Artificial Intelligence and Applications (IJAIA), March Vol.**15**, No.**2**, **2024.**

[6] Houze Liu, Iris Li, Yaxin Liang, Dan Sun, YiningYang, Haowei Yang,"Research on Deep Learning Model of Feature Extraction Based on Convolutional Neural Network," 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), Shenyang, China, pp.**810-816, 2024.** doi: 10.1109/ICIPCA61593.2024.10709168.

[7] Alexander Rombach, & Peter Fettke, "Deep Learning based Key Information Extraction from Business Documents: Systematic Literature Review", 10.48550/arXiv.2408.06345, **2024.**

[8] Tien Do, Thuyen Tran Doan, Khiem Le, Thua Nguyen, Duy-Dinh Le, and Thanh Duc Ngo, "Key Information Extraction and Recognition from Rich Text Images", Vietnam Journal of Computer Science, Vol.**11**, No.**4**, pp.**569-594, 2024**. https://doi.org/10.1142/S2196888824500131.

[9] Noura A. Semary, Wesam Ahmed, Khalid Amin, Paweł Pławiak,Mohamed Hammad, "Enhancing machine learning-based sentiment analysis through feature extraction techniques", PLoS ONE, Vol.**19**, Issue.**2**, pp.**e0294968, 2024.** https://doi.org/10.1371/journal.pone.0294968.

[10] Ivan Malashin, Igor Masich, Vadim Tynchenko, Andrei Gantimurov , Vladimir Nelyub and Aleksei Borodulin, "Image Text Extraction and Natural Language Processing of Unstructured Data from Medical Reports", Machine Learning and Knowledge Extraction., 6, pp.**1361–1377, 2024.** https://doi.org/10.3390/make6020064.

[11] Guangyun Lu, Zhiping Ni, Ling Wei, Junwei Cheng, Wei Huang, "Graphic association learning: Multimodal feature extraction and fusion of image and text using artificial intelligence techniques", Heliyon, Vol.**10**, Issue.**18**, pp.**e37167**, **2024**. ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2024.e37167.

[12] Robert West, Francesca Bonin, James Thomas, Alison J. Wright, Pol Mac Aonghusa , Martin Gleize, Yufang Hou, Alison O'Mara-Eves ,Janna Hastings, Marie Johnston, Susan Michie, "Using machine learning to extract information and predict outcomes from reports of randomised trials of smoking cessation interventions in the Human Behaviour", Wellcome Open Res, 8:452, **2024**. https://doi.org/10.12688/wellcomeopenres.20000.2

[13] M Mahfi Nurandi Karsana, Kemas Muslim L, Widi Astuti. Single-Label and Multi-Label Text Classification using ANN and Comparison with Naïve Bayes and SVM. JURNAL MEDIA INFORMATIKA BUDIDARMA. 7. 857. 10.30865/mib.v7i2.6024, **2023.**

[14] Gagandeep Kaur, and Amit Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis.", *J Big Data* **10**, 5, **2023**. https://doi.org/10.1186/s40537-022-00680-6.

[15] Kitti Szabó Nagy , and Jozef Kapusta, "A Novel Method for Feature Extraction from Unstructured Texts", Applied Science, 13, 6438, **2023**. https://doi.org/10.3390/app13116438.

[16] Xiujuan Wang, Xuerong Li, "Deep Learning in Chinese Text Information Extraction Model for Coastal Biodiversity", International Journal on Semantic Web and Information Systems (IJSWIS), Vol.**19**, Issue.**1**, pp.**1-15, 2023**. https://doi.org/10.4018/IJSWIS.331756.

[17] Sunil Kumar Dasari, Shilpa Mehta, "Text detection and recognition through deep learning-based fusion neural network", IAES International Journal of Artificial Intelligence (IJ-AI), Vol.**12**, No.**3**, pp.**1396-1406, 2023.** ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i3.pp1396-1406.

[18] An Cong Tran, Lai Thi Ho, and Hai Thanh Nguyen,"Information Extraction from Invoices by using a Graph Convolutional Neural Network: A Case Study of Vietnamese Stores", IEIE Transactions on Smart Processing and Computing, vol.**11**, no.**5, 2022.** https://doi.org/10.5573/IEIESPC.2022.11.5.316.

[19] Qing Kuang, "Face Image Feature Extraction based on Deep Learning Algorithm", Journal of Physics: Conference Series, 1852. 032040, **2021**. doi://10.1088/1742-6596/1852/3/032040.

[20] Lokkondra, C.Y., Ramegowda, D., Thimmaiah, G.M., Vijaya, A.P.B., Shivananjappa, M.H., "ETDR: An exploratory view of text detection and recognition in images and video", Revue d'Intelligence Artificielle, Vol.**35**, No.**5**, pp.**383-393, 2021.** https://doi.org/10.18280/ria.350504.

[21] Nadeesha Perera, Matthias Dehmer and Frank Emmert-Streib(2020), "Named Entity Recognition and Relation Detection for Biomedical Information Extraction", Frontiers in Cell and Developmental Biology, 8:673. doi: 10.3389/fcell.2020.00673.

[22] Philomina Simon, Uma V, "Deep Learning based Feature Extraction for Texture Classification", Procedia Computer Science, Vol.**171**, pp.**1680-1687, 2020.** ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.04.180.

[23] Vaibhav Goel, Vaibhav Kumar, Amandeep Singh Jaggi, Preeti Nagrath, "Text Extraction from Natural Scene Images using OpenCV and CNN", I.J. Information Technology and Computer Science, 9, pp.**48-54, 2019.** DOI: 10.5815/ijitcs.2019.09.06.

[24] Ruishuang Wang, Zhao Li, Jian Cao, and Tong Chen. "Chinese Text Feature Extraction and Classification Based on Deep Learning", In Proceedings of the 3rd International Conference on Computer Science and Application Engineering (CSAE '19), Association for Computing Machinery, New York, NY, USA, Article 89, pp.**1–5, 2019.** https://doi.org/10.1145/3331453.3361636

[25] Zhi Tian, Weilin Huang, Tong He, Pan He, and Yu Qiao,"Detecting Text in Natural Image with Connectionist Text Proposal Network", European Conference on Computer Vision, 9912. pp.**56-72, 2016.** 10.1007/978-3-319-46484-8_4.

AUTHORS PROFILE

Keerthana Sundar Raj received her **M.Sc., (Computer Science)** Degree in the year 2021 in Providence College for Women affiliated to Bharathiar University. She is pursuing her **Ph.D** Degree (Part Time) in Dr. N.G.P Arts and Science College, Coimbatore. She is working as **Assistant Professor** in Department of Artificial Intelligence and Machine Learning at Nehru Arts and Science of College, Coimbatore, Tamil Nadu, India. Her current research of interests includes Data Mining, Machine Learning and Deep Learning.

Dr. J. Savitha received her **Ph.D** Degree from Karpagam University, Coimbatore in the year 2017. She received her **M.Phil** Degree from Annamalai University, in the year 2009.She received her **M.Sc.,** Degree from Annamalai University, in the year 2006. She is working as **Professor**, in Department of Computer Science, Dr.N.G.P. Arts & Science College, Coimbatore, Tamilnadu, India. She has above **18 years** of experience in academic field. She has published **6 books,** more than **30 papers** in International Journals, National & International Conferences so far. Her areas of interest include Image Processing, Cyber Security, Artificial Intelligence, Machine Learning, Networks and Web Development.

