A Framework COVID-19 Invoice for Providing Structured InvoiceDocument using Optimized Bert Enabled DCNN Classifier

Prasad Dilip Kulkarni
Assist Prof & Head of Department
Electrical
KCE COEM JALGAON

Dr. Vijay Deshmukh
Associate Professor
Electronics Engineering
SSBT's College of Engineering and Technology, Jalgaon

Dr. Kantilal Pitambar Rane
Professor
Bharati Vidyapeeth College of Engineering Navi Mumbai, India

Abstract—Given Handwritten character recognition is an essential task in various domains, including optical character recognition (OCR), document analysis, and automated data entry. Convolutional Neural Networks (CNNs) have shown remarkable success in addressing this challenge by leveraging their ability to automatically learn meaningful features from raw input data. Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed for analyzing visual data such as images. They have revolutionized the field of computer vision by achieving remarkable performance in tasks such as image classification, object detection, and image segmentation. the context of the COVID-19 pandemic, the efficient processing and management of COVID-19-related invoice documents have become crucial. These documents may contain critical information related to healthcare, supplies, and services provided during the pandemic. Ensuring accurate and efficient handling of such documents is essential for healthcareorganizations, government agencies, and businesses involved in pandemic response. To address these challenges and streamline theinvoice processing workflow, the framework known as Quash HuntOptimization with Bidirectional Encoder Representations from Transformers-enabled Deep Convolutional Neural Network (QHtO-BERT enabled Deep CNN) is proposed that leverages the power of BERT, to capture semantic information from invoice text. Google Optical Character Recognition (OCR) is an advanced technology developed by Google, enabling the retrieval of text from images or scanned documents. The process of text annotation usingNamed Entity Recognition (NER) involves the task of recognizing and categorizing named entities existing within a provided text. TheBERT-based approach significantly improves semantic understanding, while the optimized DCNN ensures feature extraction and classification tasks. The model's accuracy is elevated by harnessing the Hunting and Escaping characteristics within the QHtO algorithm. The performance of the QHtO-BERT- enabled Deep CNN method concerning accuracy, sensitivity, and, specificity stands at 94.7%, 95.92%, and 95.02%, respectively.

Keywords— Invoice documents, BERT enabled Deep CNN, Google OCR, and Named Entity Recognition.

I.INTRODUCTION

Invoice processing holds utmost significance for the financial department of any organization that involves handling and managing invoices received from vendors or customers, ensuring accurate recording of financial transactions, and facilitating timely payments or receipts [1].The COVID-19 pandemic has led to an increased demand for structured invoice documents related to activities and transactions associated with the pandemic response, such as medical supplies procurement, healthcare services, and emergency support.This variation poses difficulty inautomatically extracting organized data from unstructured documents during the process of invoice recognition and data entry automation [2]. Information extraction constitutes a correlated task which intented to autonomously retrieve targeted data from intricate documents. The existing literature concentrates on two main areas: extracting information from unstructured plain text [3] [4] and extracting from scanned documents [5]. The process of extracting information from plain text is not well-suited to handle the underlying complexity of document structure. This complexity goes beyond merely extracting text-related information and necessitates the inclusion of the original layout information [6].

To overcome this obstacle, prior text extraction tools, such as OCR, [7] are commonly employed. OCR is a tool for converting various types of texts, including scanned converting study documents, PDF files, or digital cameraphotos into editable and searchable data. OCR software recognizes the text content present in these documents and converts them into machine-readable text[7]. To address this issue, the development of innovative technologies and frameworks has become crucial. Leveraging advancements in Natural Language Processing (NLP) [9], businesses can now deploy sophisticated systems capable of recognizing and extracting structured information from unstructured invoicedocuments automatically. To tackle the challenges associated with unstructured invoices a novel framework was designed that utilizes an optimized BERT model, along with a Deep CNN classifier.

The framework's primary objective is to provide structuredoutputs, delivering critical invoice fields such as names of people, organizations, locations, dates, and more. The firststep involves collecting invoices that pertain to COVID-19- related activities and transactions. Following the data collection, the content extraction process utilizes Google OCR. Then the annotation of text using NER is performed. The collected data undergoes preprocessing before being introduced into the optimized BERT enabled Deep CNN classifier, where optimization is performed through the utilization of the QHtO algorithm. Subsequently, the trained model's evaluation will be carried out using a validation dataset to assess its performance in accurately extracting structured information from invoice documents. The main contribution of the research is outlined as follows:

- ✓ *Quash Hunt Optimization (QHtO):*The QHtO algorithm is formulated by integrating the hunting behavior of hawks [10] along with the escaping characteristics of coati [11] which can handle complex optimization landscapes, Strives to achieve equilibrium between exploration and exploitation in the pursuit of the optimal solution, making it suitable for a diverse array of optimization problems.
- ✓ *Quash Hunt Optimization with BERT enabled Deep CNN (QHtO-BERT enabled CNN):*BERT is a potent pre-trained language model renowned for its comprehension of text context. By integrating BERT with a deep CNN, the framework can handle both text-based and image-based data. The CNN component can process images, such as scanned invoice documents, while BERT can process textual data present in invoices, such as descriptions and numbers. The hybrid approach allows the framework to handle diverse optimization tasks; QHtO contributes its exploratory abilities, enabling efficient and effective exploration of the search space, which is crucial in finding promising regions.

The document is categorized into several sections; Section IIdepicts the steps and difficulties involved in extracting data from invoice documents. In section III, this study includes the methodology, architecture, and data preprocessing procedures used to create our suggested framework. The findings and conclusions of the method are presented in Section IV. The conclusion of the research and the subsequentactions are detailed in Section V.

II. LITERATURE REVIEW

Manuel Carbonel *et al.*[12] employed a neural model designed for text localization and named entity recognition, effectively harnessing shared features to tackle interconnected tasks concurrently. In the case of unstructured documents, the model's capability to leverage utilizing common attributes to concurrently address interlinked tasks is limited due to the primary focus on handling a single independent task.

Arsen Yeghiazaryan *et al.* [13] introduced a Tokengrid method to extract data from unstructured documents. The authors employed two alternative 1D approaches for Line Item detection in invoices. The network architecture has a reduced number of parameters, resulting in quicker inference times during the model's evaluation phase. However, building a robust Tokengrid model demands a large and diverse datasetof annotated unstructured documents

Halil Arslan [14] deployed an application for Invoice Processing Based on Key Field Extraction. The author utilized a YOLOv5-based deep learning method for preprocessing which can handle a large volume of invoices efficiently and consistently, making it scalable to accommodate growing business needs. However, tuning the model to handle various invoices formats and layouts can be challenging and may require continuous optimization.

Hongbin Sun *et al.* [15] utilizes Spatial Dual-Modality Graph Reasoning for Key Information Extraction. This model demonstrates efficacy and resilience in managing intricate document layouts. but the ability to generalize to unseen or different document layouts and formats could be limited.

Zhanzhan Cheng *et al.* [6] employed amethod for Information extraction. Utilizing multi-modal features, such as visual, layout, and textual attributes, has the potential to enhance information extraction performance. Simultaneously, the information extraction process can serve as supervision foroptimizing text comprehension. However, generalizing across a wide range of font styles might be challenging, leading to potential errors in information extraction.

Problem Statement: Limited capability to handle interconnected tasks in unstructured documents due to a primary focus on a single independent task.Requires a large and diverse dataset of annotated unstructured documents to build a robust Tokengrid model. Difficulty in tuning the model for various invoice formats and layouts, requiring continuous optimization. Difficulty in generalizing across a wide range offont styles, leading to potential errors in information extraction.

The limitations that are mentioned above can be overcome using the proposed methods are as follows, The proposedmethod aims to create an efficient system for accurate information extraction from unstructured invoice documents, which can handle multiple tasks simultaneously by utilizing anoptimized BERT-enabled Deep CNN. The proposed method doesn't rely solely on dataset size but instead uses NER and preprocessing to handle unstructured invoice documentsefficiently. The proposed method utilizes an optimized BERT-enabled Deep CNN, which may provide more flexibility in handling different invoice formats and layouts. The proposed method's use of an optimized BERT-enabled Deep CNN may help improve text comprehension and reduce errors.

A. Challenges

- ➢ The process of annotating invoice documents is performed manually, which introduces the risk of errors and reliance on the annotator's knowledge and expertisein the data labeling task [16].
- ➢ In practice, invoices may contain noise, such as scanned artifacts or low-quality images, which can impact the accuracy of key field extraction. The application needs to handle and preprocess such noisy data effectively [14].
- ➢ Tokengrid may face challenges in comprehending complex contextual relationships within unstructured documents. This limitation can result in reduced accuracy when extracting information that heavily relies on broadercontext [13].
- ➢ Extracting information from documents with abundant visual content relies heavily on the complexity of the document's visual elements, including fonts, formatting,

and layout. If the document contains intricate visual structures, the font style embeddings may struggle to accurately capture and represent the information [6].

- While IGL-CNN considers spatial relationships, it may lack a deeper contextual understanding of the document content, leading to potential errors in key information extraction, especially when dealing with ambiguous or complex text structures [15].

III. A FRAMEWORK FOR PROVIDING STRUCTUREDINVOICE DOCUMENTS USING OPTIMIZED BERT-ENABLED DEEP CNN

The main objective of developing a novel framework for unstructured invoice documents using optimized BERT-enabled Deep CNN is to create an efficient and effective system that can accurately process and extract relevant information from unstructured invoice documents. The initial step involves collecting unstructured invoice documents related to COVID-19 from real-time sources. Next, the text is extracted from the collected data using Google OCR, the extracted text is annotated using NER to identify and categorized named entities such as names of people, organizations, locations, and dates. After annotation, the collected data undergo preprocessing to remove noise, standardize formats and handle inconsistencies. TF-IDF is a popular statistical technique employed in natural language processing and information retrieval which quantifies the significance of a term in a document to a larger collection of documents. The preprocessed data was then fed into an optimized BERT enabled deep CNN to make accurate and nuanced predictions. To optimize the classifier this research utilizes the characteristics of Golden Hawks and Coati. The final step involves evaluating the trained model using the validation dataset to measure its performance in correctly extracting structured information from invoice documents. The schematic representation of the optimized BERT-enabled Deep CNN for providing structured invoice documents is shown in Figure 1.

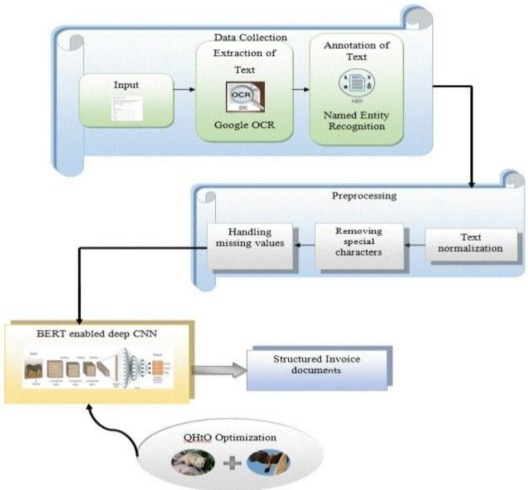


Figure: 1 Schematic representation of Optimized BERT enabled deep CNN

3.1 Input

The input for the proposed framework consists of 100 images or scanned copies of invoices related to COVID-19. Each image contains 16 labels, and the label has a dimension of $[16 \times 25]$. The total label dimension for all the images is $[100 \times 16 \times 25]$.

3.2 Extraction of Text Using Google OCR

The primary function of OCR [16] is to recognize and extract the textual content from the input documents so that the text can be processed, searched, edited, and analyzed by computers and software applications. In this proposed framework OCR is commonly used to convert physical invoices into digital formats, making them more accessible and editable.

3.3 Text Annotation using NER

NER holds utmost importance and serves as a critical task within information extraction.A natural language processing technique called NER involves locating and assigning categorizations to named entities [17] within a document, classifying them into specific groups like individuals' names, locations, seat no, date, flight no, village, district, pin code, and so on.

3.4 Preprocessing

Preprocessing refers to the initial steps taken to prepare raw data before it can be used for analysis or input into a machine learning model. In this research, preprocessing is done manually which includes text normalization; removing special characteristics, and handling missing values.

3.4.1 Text normalization: Text normalization [18] is a crucial step in natural language processing which involves transforming text into a canonical or standardized form. The purpose of text normalization is to bring different variations of words or phrases to a common representation, reducing redundancy and facilitating more accurate analysis and comparison of text.

3.4.2 Removing special characteristics: Removing special characteristics refers to the process of eliminating non-alphanumeric characters, symbols, and other non-textual elements from a piece of text during text normalization. By removing special characters, the text becomes cleaner and more focused on the essential content, making it easier to process and analyze.

3.4.3 Handling missing values: Handling missing values [19] is a crucial step in data preprocessing, applicable to various data analysis tasks.Missing values occur when certain data points are not available or are incomplete in a dataset.

IV. RESULT AND DISCUSSION

In this research, the QHtO-BERT enabled Deep CNN is developed for data extraction from invoice documents lacking structured formatting and the goal is to build an intelligent system that can accurately and efficiently extract relevant information from invoice documents. The method's results and discussion are outlined as follows.

4.1 Experimental Setup

The QHtO-BERT enabled Deep CNN was executed using the dataset on a Windows 10 machine with 8GB of RAM, utilizing MATLAB software.

4.2 Dataset Description

The dataset used for theQHtO-BERT enabled Deep CNN consists of a collection of documents relevant to COVID-19 which is designed to train and evaluate the QHtO-BERT enabled Deep CNN model that combines the QHtO-BERT and Deep CNN for named entity recognition. The dataset comprises 100 numbers of samples, each containing text data and corresponding 16 labels.

4.3 Performance Metrics

The parameters for measuring the efficacy of information extraction in unstructured invoice documents are discussed as follows,

4.3.1 Accuracy

Accuracy is characterized as the proportion of accurately classified instances to the overall number of instances within the dataset and it is mathematically assessed as follows:

$$accuracy = \frac{P_{TP} + P_{TN}}{P_{TP} + P_{TN} + P_{FP} + P_{FN}} \tag{13}$$

4.3.2 Sensitivity

Sensitivity measures the percentage of positive occurrences that the model properly classifies as positive. Mathematically sensitivity is expressed as,

$$sensitivity = \frac{P_{TP}}{P_{TP} + P_{FN}} \tag{14}$$

4.3.3 Specificity

Specificity represents the percentage of real negative events that the model accurately classifies as negative and is evaluated as,

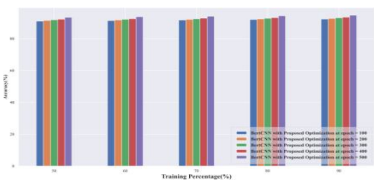
$$specificity = \frac{P_{TN}}{P_{TN} + P_{FP}} \tag{15}$$

4.4 Performance analysis

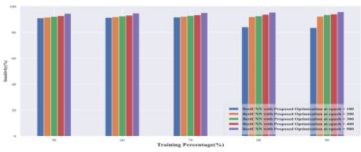
The effectiveness ofQHtO-BERT enabled DCNN model is computed using the dataset and the TP metric. For epoch sizes of 100, 200, 300, 400, and 500, the TP values 50, 60, 70, 80, and 90 are taken into consideration.

4.4.1 Performance analysis for QHtO-BERT enabled Deep CNN with TP

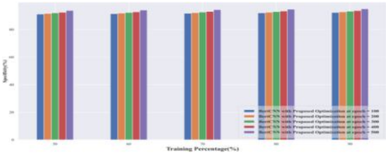
Figure 6 shows the performance outcomes for the QHtO-BERT enabled Deep CNN for data extraction using different epoch settings. The accuracy measured in terms of TP 90 utilizing QHtO-BERT enabled deep CNN with epochs 100, 200, 300, 400, and 500 was 92.30%, 92.7%, 93.1%, 93.5%, and 94.7%, as shown in Figure 4(a). The QHtO-BERT enabled Deep CNN achieves values of 83.7%, 92.47%, 93.62%, 94.2%, and 95.92% for the epochs 100, 200, 300, 400, and 500 when assessing the sensitivity for TP 90, as shown in figure 4(b). The QHtO-BERT enabled DCNN achieves values of 92.36%, 92.81%, 93.25%, 93.69%, and 95.02%, respectively, with different epochs of specificity over TP 90 with the epochs 100, 200, 300, 400, and 500, as shown in figure 4(c).



a) accuracy



b) sensitivity



c) specificity

II. CONCLUSION

In this research, the framework combines the power of BERT's contextual understanding with the deep convolutional neural network's image processing capabilities to efficiently extract structure information from COVID-19 re-invoice documents. The optimized BERT model significantly improved the performance of the deep CNN classifier, enhancing its ability to handle complex and varied invoice layouts. The framework demonstrated robustness in handling different invoice formats, even when dealing with scanned images of this study invoices. Furthermore, the comparative analysis with conventional methods showcased the superiorityof QHtO-BERT enabled Deep CNN framework, outperforming traditional approaches. The achieved values for accuracy, sensitivity, and specificity were 94.7%, 95.92%, and 95.02%, respectively. Subsequent efforts might center on expanding the framework's capabilities to handle multilingual invoices, handling unclear data, and exploring potential integration with blockchain technology for enhanced security and transparency in invoice processing workflows.

References

[1] Carbonell, M., Fornés, A., Villegas, M. and Lladós, J., 2020. A neural model for text localization, transcription, and named entity recognition in full pages.Pattern Recognition Letters, 136, pp.219-227.
[2] Yeghiazaryan, A., Khechoyan, K., Nalbandyan, G. and Muradyan, S., 2022. Tokengrid: Toward More Efficient Data Extraction From Unstructured Documents. IEEE Access, 10, pp.39261-39268.
[3] Arslan, H., 2022. End to End Invoice Processing Application Based on Key Fields Extraction. IEEE Access, 10, pp.78398-78413.
[4] Sun, H., Kuang, Z., Yue, X., Lin, C. and Zhang, W., 2021. Spatial dual-modality graph reasoning for key information extraction. arXiv preprint arXiv:2103.14470.
[5] Cheng, Z., Zhang, P., Li, C., Liang, Q., Xu, Y., Li, P., Pu, S., Niu, Y. and Wu, F., 2022. Trie++: Towards end-to-end information extraction from visually rich documents. arXiv preprint arXiv:2207.06744..
[6] Sang, E.F. and De Meulder, F., 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. arXiv preprint cs/0306050.
[7] Pogorilyy, S.D. and Kramov, A.A., 2020. Assessment of Text Coherence by Constructing the Graph of Semantic, Lexical, and Grammatical Consistancy of Phrases of Sentences.Cybernetics and Systems Analysis, 56, pp.893-899.
[8] Jaume, G., Ekenel, H.K. and Thiran, J.P., 2019, September. Funsd: A dataset for form understanding in noisy scanned documents. In 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW) (Vol. 2, pp. 1-6).IEEE.
[9] Palm, R.B., Winther, O. and Laws, F., 2017, November. Cloudscan-a configuration-free invoice analysis system using recurrent neural networks.In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 1, pp. 406-413).IEEE.
[10] Park, J., Lee, E., Kim, Y., Kang, I., Koo, H.I. and Cho, N.I., 2020. Multi-lingual optical character recognition system using the reinforcement learning of character segmenter.IEEE Access, 8, pp.174437-174448.
[11] Park, J., Lee, E., Kim, Y., Kang, I., Koo, H.I. and Cho, N.I., 2020. Multi-lingual optical character recognition system using the reinforcement learning of character segmenter.IEEE Access, 8, pp.174437-174448.
[12] Liu, W., Zhang, Y. and Wan, B., 2016. Unstructured document recognition on business invoice. Machine Learning, Stanford iTunes University, Stanford, CA, USA, Technical report.
[13] Li, J., Sun, A., Han, J. and Li, C., 2020. A survey on deep learning for named entity recognition. IEEE Transactions on Knowledge and Data Engineering, 34(1), pp.50-70.
[14] Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M. and Chen, H., 2019. Harris hawks optimization: Algorithm and applications. Future generation computer systems, 97, pp.849-872.

[15] Osama, S., Ali, A.A. and Shaban, H., 2023. A hybrid of Information gain and a Coati Optimization Algorithm for gene selection in microarray gene expression data classification.Kafrelsheikh Journal of Information Sciences, 4(1), pp.1-16.
[16] Baviskar, D., Ahirrao, S. and Kotecha, K., 2021. Multi-layout unstructured invoice documents dataset: A dataset for template-free invoice processing and its evaluation using AI approaches. IEEE Access, 9, pp.101494-101512.
[17] Ingólfsdóttir, S.L., Guðjónsson, Á.A. and Loftsson, H., 2020, September. Named entity recognition for icelandic: Annotated corpus and models. In International Conference on Statistical Language and Speech Processing (pp. 46-57). Cham: Springer International Publishing.
[18] Sunkara, M., Shivade, C., Bodapati, S. and Kirchhoff, K., 2021, June. Neural inverse text normalization.In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 7573-7577).IEEE.
[19] Khan, S.I. and Hoque, A.S.M.L., 2020. SICE: an improved missing data imputation technique. Journal of big Data, 7(1), pp.1-21.
[20] Caragea, D., Chen, M., Cojoianu, T., Dobri, M., Glandt, K. and Mihaila, G., 2020, December. Identifying FinTech innovations using BERT. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 1117-1126). IEEE.
[21] Ouni, S., Fkih, F. and Omri, M.N., 2022. BERT-and CNN-based TOBEAT approach for unwelcome tweets detection. Social Network Analysis and Mining, 12(1), p.144.
[22] Abas, A.R., Elhenawy, I., Zidan, M. and Othman, M., 2022. BERT-CNN: A Deep Learning Model for Detecting Emotions from Text. Computers, Materials & Continua, 71(2).
[23] Shi, P., Ng, P., Wang, Z., Zhu, H., Li, A.H., Wang, J., dos Santos, C.N. and Xiang, B., 2021, May. Learning contextual representations for semantic parsing with generation-augmented pre-training.In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 15, pp. 13806-13814).
[24] Sun, Y., Zheng, Y., Hao, C. and Qiu, H., 2021. NSP-BERT: A Prompt-based Zero-Shot Learner Through an Original Pre-training Task--Next Sentence Prediction. arXiv e-prints, pp.arXiv-2109.
[25] Qu, C., He, W., Peng, X. and Peng, X., 2020. Harris hawks optimization with information exchange. Applied mathematical modelling, 84, pp.52-75.
[26] Rama Krishna, K., Mohan, K.G. and Mahalakshmi, R., A NOVEL APPROACH TO PREDICT DENGUE DISEASES IN PATIENTS USING COATI OPTIMIZATION-BASED SUPPORT VECTOR MACHINE.
[27] Li, Y., Huang, Z., Yan, J., Zhou, Y., Ye, F. and Liu, X., 2021. GFTE: graph-based financial table extraction. In Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part II (pp. 644-658). Springer International Publishing.
[28] Mulahuwaish, A., Gyorick, K., Ghafoor, K.Z., Maghdid, H.S. and Rawat, D.B., 2020. Efficient classification model of web news documents using machine learning algorithms for accurate information. Computers & Security, 98, p.102006.
[29] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L., 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, pp.1-74.