Intelligent Character Recognition System for Account Payable by using SVM and RBF Kernel

Muhammad Umer Farooq[1]*, Abdul Karim Kazi[1], Mustafa Latif[2], Shoaib Alauddin[1], Kisa-e-Zehra[1] and Mirza Adnan Baig[1]

[1]Department of Computer Science and Information Technology, NED University of Engineering and Technology, Karachi, 75270, Pakistan.

[2]Department of Software Engineering, NED University of Engineering and Technology, Karachi, 75270, Pakistan.

*Corresponding Author: Email: umer@neduet.edu.pk

Summary
Intelligent Character Recognition System for Account Payable (ICRS AP) Automation represents the process of capturing text from scanned invoices and extracting the key fields from invoices and storing the captured fields into properly structured document format. ICRS plays a very critical role in invoice data streamlining, we are interested in data like Vendor Name, Purchase Order Number, Due Date, Total Amount, Payee Name, etc. As companies attempt to cut costs and upgrade their processes, accounts payable (A/P) is an example of a paper-intensive procedure. Invoice processing is a possible candidate for digitization. Most of the companies dealing with an enormous number of invoices, these manual invoice matching procedures start to show their limitations. Receiving a paper invoice and matching it to a purchase order (PO) and general ledger (GL) code can be difficult for businesses. Lack of automation leads to more serious company issues such as accruals for financial close, excessive labor costs, and a lack of insight into corporate expenditures. The proposed system offers tighter control on their invoice processing to make a better and more appropriate decision. AP automation solutions provide tighter controls, quicker clearances, smart payments, and real-time access to transactional data, allowing financial managers to make better and wiser decisions for the bottom line of their organizations. An Intelligent Character Recognition System for AP Automation is a process of extricating fields like Vendor Name, Purchase Order Number, Due Date, Total Amount, Payee Name, etc. based on their x-axis and y-axis position coordinates.

Keywords:
Account Payable Automation; Intelligent Character Recognition; Invoice processing; Smart payments

1. Introduction

The Intelligent character recognition for Account Payable Automation is designed specially to capture, and extract account information and can serve many applications and services for account payable invoices. ICR system work with OCR extracted data to automate the process of capturing the data from forms and eliminates the way of keystrokes for manually entering data. An Accounts Payable Clerk is responsible for processing this paperwork making an average of $39,950 per year, or $19.21 per hour, according to Glassdoor [1]. OCR scanned invoices is a process of extricating text from scanner invoices receipts. Then again, extricating key texts from receipts and invoices and saving the texts to organized reports can serve a variety of applications and services, for example, effective storage, quick indexing, and record examination. ICRS plays a very critical role in streamlining document-intensive processes and office automation in many financial, account, and taxation fields. However, the breakthrough in the performance was boosted by recent deep learning advancements in terms of accuracy and processing time. The OCR development becomes much more mature for many practical problems such as handwritten character recognition, license plate number recognition, visiting card data extraction, and many more. But still, invoice OCR required high accuracy requirements as compared to general OCR tasks for many other applications. Therefore, in the existing invoice processing system, there is still a considerable demand for human resources. There is an urgent need to develop a system to correctly classify the required fields without any further human intervention with OCR extracted results.

OCR extracts text from invoices, but for extracting key parameters this (extracted text from the invoice (ETFI) is not enough [2]. The extricating of key fields is challenging since there is no defined format for invoices. Most methods for KIPE use textual information from a detected bounding box using character level sequence tagging [3], inspired by Name Entity Recognition (NER) architecture [4]. By integrating spatial and visual elements, we will be able to get a complete overview and accurately extract key fields from the invoice. The aforementioned issue can be solved with our end-to-end sequential design approach. The proposed approach may be used to extract key information from any

Manuscript received November 5, 2022
Manuscript revised November 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.11.31

document with an unknown structure.

There has been a rise of interest in recent years in the subject of process automation in an attempt to reduce the need for human engagement in the operation. The automated method usually begins by transforming physical or digital invoice documentation into a machine-readable format, followed by translating the invoice images into a textual discoverable document. The method of processing the documents varies depending on the automation solution, but the final result is the registration of the desired field values into the organization's Enterprise Resource Planning systems (ERPs)

Dataset challenges

Unstructured dataset is an asset to any organization as they have a huge variety of key information stored in them. If an organization extracts these key insights and uses them for the decision-making process, it can significantly increase its operational efficiency [5]. However, manually invoice processing and extricating key fields from various complex invoice layouts is a time is taken and error-prone process. Hence, developing a machine learning ml-enabled tool for automatic key fields information collected from various invoice layouts is a promising and recently, research focus. However, the automatic extraction of key insights from unstructured document research faces certain key challenges [6,7]. Obtaining a high-quality standard and well-annotated dataset from unstructured invoices collection is one of the most basic and critical challenging tasks. A good quality dataset is the foremost essential entity to train a machine learning model. The model's resilience and accuracy are determined by the quality of the training dataset. Therefore, it is always necessary to include variations in the training data so that the model learns to recognize the unknown data [8]. In this study, standard datasets or publicly available datasets confront several obstacles, which are mentioned below:

- Datasets that are publicly available are unlabeled. As a result, manually labeling or annotating the dataset takes time and effort [6,7].
- Poor-quality, fuzzy, skewed, and reduced document photos dominate publicly accessible databases, resulting in poor text extraction [9].
- Datasets made accessible to the public are outdated, as they contain ancient or obsolete document formats [10].
- Datasets that are publicly available are domain- and task-specific. The dataset offered in [11-12] is, for example, utilized in the healthcare area, while the dataset proposed in [13-14] is used in the legal contract analysis sector. They're also utilized for particular tasks like extracting information from scholarly papers [15] or extracting patient details from a clinical dataset [16].

A few research projects offered a custom dataset as a solution to these problems. Custom datasets, on the other hand, are private and subject to confidentiality concerns [6,17]. Documents with comparable layouts or formats are included in the custom dataset. The ability of key extraction operations to be generalized is limited when comparable layout documents are provided.

Dataset and Annotations

The dataset contains 20 different types of multi-page scanned PDF invoices with more than 100 total invoices, which are used to train and test the machine learning model. These given invoices are annotated by using the defined model layout. Each type of invoice has different fields to capture such as vendor name, Payee Name, Purchase Order Number, Due Date, Total Amount, etc. The annotated field in the dataset mainly consists of digits, English characters, and a combination of both. The dataset was divided into two parts: a training set and a test set. The "train" set consists of 70 invoices with all different variants of layouts made available to the participants along with their annotations. The "test" set consists of 30 invoices with all different types of vendor layouts. The dataset is confidential and only used for training the machine learning model with limited rights. For invoices key fields detection and OCR tasks, each field in the available data is annotated with a text position coordinate (bbox) and the bbox transcript of each text Location is labeled as rectangles with four vertices that are arranged clockwise from the top. Annotations for each invoice field are stored in a JSON file format with the same invoice name, whereas in information extraction task, each image in the dataset is also annotated by using the model layout defined by users.

Text detection from scanned documents and images have gotten more precise and advanced in recent years, thanks to their widespread use. There are many frameworks and APIs developed for the given tasks, (i.e., text detection) Tesseract OCR engine [20, 21], Amazon Textract, Google vision API, and many other APIs and frameworks. Text extraction using geographical information is now a sophisticated technique. However, extracting important parameters from documents with

high accuracy and low latency remains a difficulty for services and applications that demand high precision.

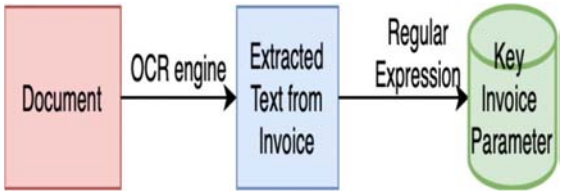

Fig. 1. The architecture of OCR's rule-based KIPE [2].

CloudScan is a method for analyzing invoices that intend to learn a single unified invoice model that can be applied to unspecified bills [22]. It employs a recurrent neural network (RNN) based on Long Short-Term Memory (LSTM), which excels at simulating long-term relationships. The network was trained on a dataset of 326,471 invoices in order to extract 8 distinct entities. The system may be thought of as a NER system that extracts named entities from text, such as people or places. Cloudscan received an F1-score of 0.891 overall. This model was compared to a logistic regression-based baseline model with an F1-score of 0.887.

The mapping of invoice text to a grid is a new technique that has recently been presented. The research paper encodes each invoice page as a two-dimensional grid of characters Chargrid: Towards Understanding 2-Dimensional Documents [23]. A fully convolutional encoder-decoder network is employed in this grid-based technique, with semantic segmentation using one encoder and bounding box regression with two decoders. Date, invoice number, amount, supplier name, and supplier address were all retrieved from the dataset. The data collection included 12000 scanned invoices from a wide range of suppliers and languages. Approaches based on consecutive text or document pictures performed worse than Chargrid.

A study was conducted in order to automate the process of extracting data from PDF invoices, and a graph-based technique was presented [24]. The content of PDFs was retrieved using graph convolutional neural networks once the text was represented as graphs (GCN). The invoice number, total amount, and invoice date were the most important data points collected. There were 1129 English invoices from 277 distinct merchants in the sample. With an F1-score of 0.875, the suggested model retrieved these important components from several invoices.

Stanford's group of researchers describe a bag-of-words approach in order to identify the linguistic and structural information for business document identification [25]. They used 97 raw invoice images and created 8000 features from 2095 tokens from these invoices. SVM machine learning technology gives the best result with minimum training and testing error with 157 selected features out of 8000 created features such as Side to side aligned tokens, vertically stacked tokens, and Adjacent tokens (Distance Threshold), Vertical location, and type. This approach is more suitable for defined layout information extraction as compared to multi-layout invoice documents.

Google has also looked at the subject of information extraction from templatic documents. One of their most recent contributions is the publication "Representation Learning for Information Extraction from Form-like Documents" [30], which introduces a unique technique for automatically structured data extraction from Templatic documents using representation learning. The method creates extraction candidates based on field type knowledge, and then a neural network discovers the best representation for all contenders based on surrounding words. When tested using unseen invoices, this method received an F1-score of 0.878.

Pattern matching

The ruthless attitude is to find the key fields from invoices that required pattern matching techniques with OCR invoice extracted text. Figure 1 refers to the rule-based architecture. The KIPE pattern matching based on rules is simple and robust, but it is not particularly accurate. The structure of a document cannot be generalized. This strategy will not work in sectors where greater accuracy criteria are required. Invoices may contain several occurrences of the text pattern that corresponds to the key; in such circumstances, this technique may cause uncertainty.

Classification approach for KIPE

SVM, Decision Tree, and Logistic Regression are examples of classification algorithms. Using these algorithms, we may categorize the key fields using their positional coordinates information into different classes, like vendor name, payee name, due date, invoice date, etc. Since we have multiple key fields, our classification task can be solved with a multi-class classifier. Here, they have assumed each geolocation is atomic in nature, i-e. each box represents a single class. Since we have numerous bounding boxes with the same key tag, it might be difficult to discover the needed key-value pair. We are not using any sequential characteristics in this technique; hence the outcome will be unsatisfactory.

To conclude, previous research focuses on key field extraction tasks on documents that have comparable layouts, such as receipts. There is no dataset in the literature that contains high-quality, many, or diversified layout documents. The accuracy of AI and NLP models is largely determined by the availability of adequate and diverse training data. This study also uses the SVM with RBF kernel approach to handling our multi-layout invoices obtained from various vendors without utilizing templates.

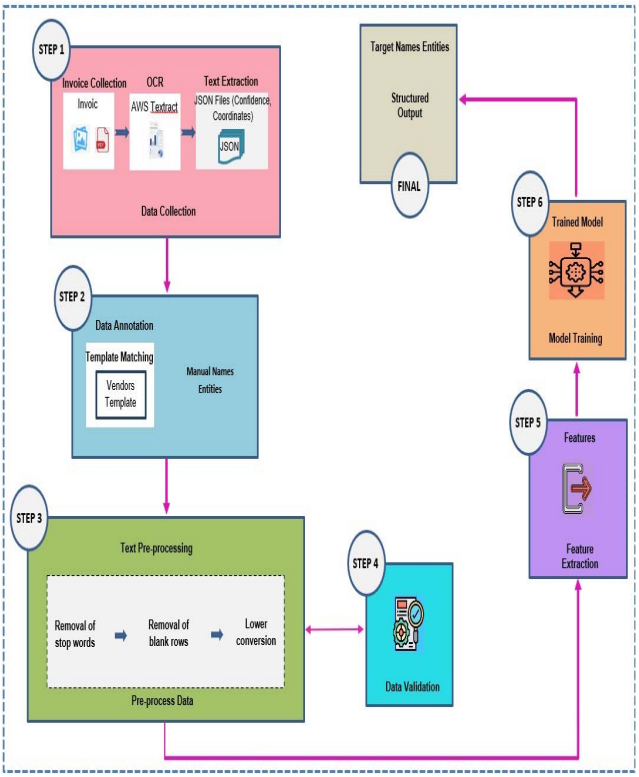

Fig. 2. Proposed Framework for Extracting Key Fields from Unstructured Invoice Documents with Multiple Layouts

2. METHODOLOGY

In this research, an improved technique has been presented for key fields extraction from invoices text. I have ensemble various features from extracted invoice text followed by bounding box level multiclass classification by using simple machine learning techniques. We are using the surrounding coordinates for each text block and calculating their angle with other surrounding text boxes and their distance. Each key field text block is classified into different categories, our invoice use-cases, each of them may be categorized into a class, Vendor Name, Payee Name, Due Date, PO Number, and None. Where none does represent any category. We are going, to begin with, one basic assumption that every key field text box (TB) nature is atomic, i.e., each TB corresponds to only 1 class, this may or may not be followed once we have a document with extremely dense information, wherever there may be a risk that extracting text from totally various categories into one class.

Data acquisition

To generate the variance in invoice patterns invoices from various supplier companies are gathered. All the collected invoices are in a soft copy format and do not need to scan. Each supplier's invoices have a distinct design and format. The distinct design and format layouts of invoices later open door to design a layout for the new distinct invoice. Figure 2 depicts this, as invoices PDFs are gathered, as the initial step all these PDFs are converted into images and fed into the AWS OCR. Textract engine for data extraction and text detection. After that, all the extracted data and their related information are stored in JSON file format. Consequently, it was discovered that the AWS Textract performed very well without missing any text from the documents. The user-defined model templates are used for annotating the key fields with their matched vendor template and all the other fields are categorized as a 'none' class. The label data is further used for extracting new features by using their surrounding text box information. More statistical data about the proposed dataset is summarized in Tab 1. The Invoices obtained for this research belong to the advertising firm that uses many vendors for a variety of material invoices.

The AWS Textract OCR service was used to extract the textual data from these invoices. When the API receives a picture of a document, it returns a JSON-object containing all textual data discovered by its OCR capabilities. Individual properties of sub-images identified in the supplied picture are also included in the answer. The textual data is grouped together by its locations, detection of content and categories, and individual characteristics of sub-images found in the provided image are also included in the response.

Table 1. Statistical overview of Multi layout document dataset.

Data Annotation

Since the extracted invoice data was not annotated but supervised machine learning approach required annotated data to train the machine learning model to identify the desired field of interest from all the extracted raw information. We have used a user-defined different template model to annotate the key fields using the existing mechanism. The model template consists of key fields' positional coordinates with some flexibility in their coordinate values. Some of the invoices are not completely mapped to their defined template but manually all these cases handle during the data annotation phase. The complete flowchart of training dataset creation is shown in figure 3.

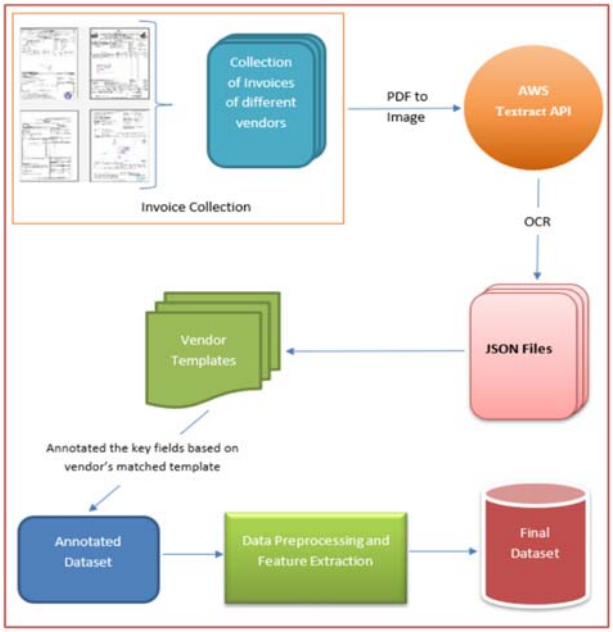

Fig. 3. Flowchart of Annotated data creation.

Data Pre-Processing

Stop words including pronouns (we, his), connectives (and, or), and articles (a, an, the) don't help you grasp what a phrase or document is saying. As a result, stop words must be removed in order to minimize memory space and processing costs. As a first and crucial step in the pre-processing procedure, articles, pronouns, and connectives are deleted. After the Stop-word is removed, all tokens are transformed into lowercase letters. It's a must if you want to keep the output in a consistent format. This is a crucial data pre-processing step for removing duplicate records and blank rows that may have been introduced as a result of OCR problems.

S #	Parameters	Properties
1.	Period of Invoice collection	From 2020 to 2022 for different vendors
2.	Each invoice page has a set number of words.	400 words per page on average.
3.	Size of one invoice PDF of any layout.	6 KB minimum 300 KB average 3 MB maximum
4.	Software used for Multi layout Document annotation construction	Jupyter Notebook, Google Colab platform, AWS Textract OCR for text extraction from invoice PDF, and PyCharm. User defined invoice layout for NER annotations.

Fig. 4. Pipeline for pre-processing data

Feature Extraction

Features engineering selection is one of the important areas in applied machine learning. It is the process of producing new features from existing data in order to train a machine learning model. Since a machine learning algorithm could only learn with input data that we provide it, this step is frequently more important than the actual model itself [26]. It is typically a time-consuming manual procedure that relies on domain expertise, intuition, and data manipulation. In this thesis, we have used some domain knowledge and existing methods to define new features using existing information. Each text field is surrounded by different other text fields. In Figure 5, the yellow box field is our desired field to extract so we have used all the other surrounding text information to sure about using some mathematics and trigonometry approaches as well page contextual information. All the other desired fields used the same kind of information for understanding more about their surrounding fields. This approach is simpler and more robust as compared to other Name Entity Recognition complex machine learning techniques and using the existing system information.

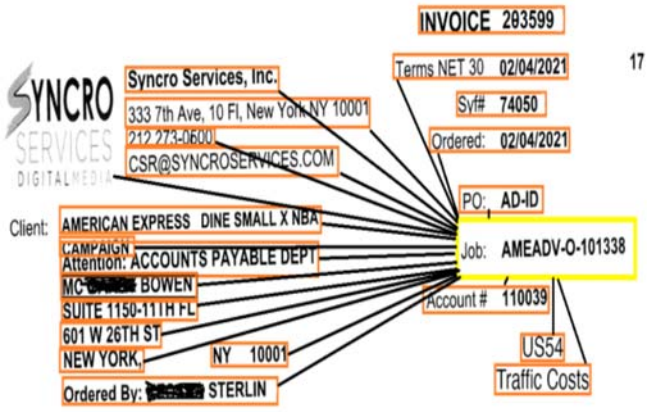

Fig. 5. Feature extraction approach for key fields extraction

Tool Selection

The objective of this research is to select a more robust framework for training a machine learning model. So, the data collected and prepared in the above stage and now is ready to be used. The tool chooses to perform this task is Python 3.0 Jupyter Notebook. Reason of choosing Python over R it gives robust powerful packages for machine learning model training and its open-source auto-machine learning frameworks using which machine learning model creation tasks can be accomplished in a most efficient and quick way. There is basic to advance statistical machine learning packages available for both supervised and unsupervised learning in python which allows almost every type of data to be dealt with. In this research report, we present an improved technique for key fields extraction from invoices text. Various features have been assembled from extracted invoice text followed by bounding box level multi-class classification by using simple machine learning techniques.

Model Selection

For supervised machine learning classification tasks, various machine learning models can be used like logistics regression, k nearest neighbors, support vector machines i.e., SVM, multi-layer perceptron (artificial neural network), and some more. In this research paper, nine different machine learning models were trained and used to make predictions: SVM with different parameters settings using a grid search approach. These open-source frames are widely used in areas and provide very good results, but the main issues are related to these frameworks. It cannot create new valuable features on its own due to domain knowledge area gaps.

As an initial step, we start by converting our task into a binary classification task with labels -1, 1 for all classes using the one-vs-rest method. SVC with RBF kernel trick provides very good results on the training dataset. But SVC algorithms are more based on the label name "2114" known as the "Invoice Number" field. We have an imbalance dataset for machine learning model training. The results can be improved by using some imbalanced dataset technique like smote, etc. Figure 6 represents the results of SVM training model.

3. RESULTS AND DISCUSSION

One for annotating essential fields and obtaining all necessary and beneficial information from the bounding box using multi-layout vendor templates. Second, a machine learning model is trained to categorize each bounding box is divided into numerous categories, including total, address, firm identity, dates, order number, and so on. The existing systems mainly focus on extracting information from scanned documents with a similar template or layout. In practice, however, companies get invoices from a variety of vendors, each of them having its own structure and layout. Open-source datasets has lacked a collection of invoice documents which contains varied numbers of different invoice layouts.

Classification Report :	precision	recall	f1-score	support
1114	1.00	0.96	0.98	46
1118	1.00	1.00	1.00	105
1218	1.00	0.50	0.67	2
1318	1.00	1.00	1.00	2
2114	0.91	1.00	0.95	126
2118	1.00	0.99	1.00	112
2216	1.00	1.00	1.00	23
2220	1.00	0.99	0.99	100
3114	1.00	1.00	1.00	41
3118	1.00	1.00	1.00	117
3214	1.00	1.00	1.00	19
3220	1.00	1.00	1.00	136
3316	1.00	1.00	1.00	21
4114	1.00	1.00	1.00	65
4116	1.00	1.00	1.00	32
4118	1.00	1.00	1.00	32
4218	1.00	1.00	1.00	21
4320	1.00	1.00	1.00	32
5118	1.00	0.98	0.99	114
5130	1.00	0.94	0.97	17
5220	1.00	1.00	1.00	108
6114	1.00	0.97	0.98	33
6116	1.00	1.00	1.00	14
6118	1.00	1.00	1.00	98
6130	1.00	1.00	1.00	19
6214	1.00	1.00	1.00	4
6216	1.00	1.00	1.00	41
6220	1.00	1.00	1.00	110
6316	1.00	1.00	1.00	26
7116	1.00	1.00	1.00	4
7120	1.00	1.00	1.00	41
8114	1.00	0.97	0.98	31
8116	1.00	1.00	1.00	9
8118	1.00	1.00	1.00	128
8130	1.00	1.00	1.00	3
8214	1.00	0.92	0.96	12
8216	1.00	1.00	1.00	14
8220	1.00	1.00	1.00	106
8260	1.00	0.80	0.89	5
8316	1.00	1.00	1.00	17
accuracy			0.99	1986
macro avg	1.00	0.98	0.98	1986
weighted avg	0.99	0.99	0.99	1986

Accuracy (RBF Kernel): 99.40
F1 (RBF Kernel): 99.39

Fig. 6. Training report of SVM model.

- The suggested multi layout key field retrieval technique has a variety of consequences for extracting different entities as structured output from a huge proportion of unstructured purchase orders documents. Furthermore, end-to-end automation of the invoice information extraction procedure assists every organization's finance department in completing invoices quickly and verifying accounts due and receivable.

- Customer acquisition and validation procedures are impacted by automated key field extraction from accounting reports such as bills, which has an influence on company performance. It has the potential to drastically decrease the costs associated with human record keeping and invoice verification for thousands of vendors invoices each day.

For detecting and extracting named items from multi-layout unstructured invoice documents, several AI techniques are applied. The following AI techniques are used to analyze the supplied multi-layout unstructured invoice documents dataset. BiLSTM is a Recurrent Neural Network (RNN) that uses two LSTM, one forward and one backward, to analyze the text sequence. It is utilized when prediction tasks require knowledge of both the previous and future input sequences. This example shows how word embeddings like as Word2Vec, GloVe, and FastText may be used in conjunction with BiLSTM to extract critical fields from a multi-layout unstructured invoice document dataset. Word embedding approaches such as thatWord2Vec and GloVe are commonly employed with language models to obtain contextual embeddings by integrating hidden layers. However, training GloVe andWord2Vec on unstructured documents such as invoices is not a smart idea because the forms of Invoice Numbers, Total Invoice Amounts, and Invoice Dates are all different [27]. The extracted features obtained for such uncommon words will be of poor quality and useless for key field extraction. Fast-Text outperformsWord2Vec and GloVe in the majority of key field extraction tasks [28]. Another sequence processing model with knowledge of Bidirectional LSTM and data labeling logic via CRF is BiLSTM-CONDITIONAL RANDOM FIELD. The weights associated with data samples are learned by BiLSTM using the CRF model. It outperforms BiLSTM in tasks involving natural language processing and analysis [29].

Figure 7 shows the summary picture of overall performance of most similar Recognition. Here we can clearly see that the proposed solution performs very well in the above setting to extract the key fields from multi-layout multi-page financial documents.

	BiLSTM – Word2Vec	BiLSTM – GloVe	BiLSTM – FastText	BiLSTM-CRF	Proposed
UBIAI	✓	✓	✓	✓	×
Textual Data	✓	✓	✓	✓	×
Positional Coordinates	×	×	×	×	✓
Multi - Layout	✓	✓	✓	✓	✓
Multi - Pages	×	×	×	×	✓

Fig. 7. SVM-RBF result evaluation based on given settings as compared to BiLSTM and its variants.

We have used Vendor defined templates to annotate the key fields from invoices. However, all the other systems used previous build tools called UBIAI to annotate the key fields. The result of our proposed model is quite good as compared to BiLSTM and its variant in given setting. The performance of the model is measured with precision, recall, and F1-score, as shown in Figure 8. BiLSTM with word2Vec, GloVe, and FastText perform very bad because of automatically bad features selection with any prior knowledge while BiLSTM-CRF with word embedding layer perform very well, it also used their coordinate information for identify each field and understand the text of each field. On the other hand, this proposed model does not used any text data information, this framework mainly focus on coordinates of each entity and their surrounding field coordinates. We have different Vendor invoices and it's had different naming conventional to identify the same quantity such as Invoice Number or Invoice #, etc. This issue can efficiently resolve with given solution.

Model	Features	Entities Extracted	Precision	Recall	F1-score
BiLSTM	Word2Vec	Invoice Number	0.00	0.00	0.00
		Invoice Date	0.00	0.00	0.00
		Buyer Name	0.00	0.02	0.00
		Buyer GST Number	0.00	0.00	0.00
		Supplier Name	0.29	0.40	0.32
		Supplier GST Number	0.00	0.00	0.00
		Grand Total Amount	0.00	0.00	0.00
BiLSTM	GloVe	Invoice Number	0.10	0.08	0.09
		Invoice Date	0.01	0.27	0.02
		Buyer Name	0.17	0.17	0.16
		Buyer GST Number	0.02	0.01	0.01
		Supplier Name	0.02	0.01	0.01
		Supplier GST Number	0.01	0.06	0.02
		Grand Total Amount	0.00	0.00	0.00
BiLSTM	FastText	Invoice Number	0.04	0.03	0.03
		Invoice Date	0.00	0.01	0.00
		Buyer Name	0.11	0.53	0.13
		Buyer GST Number	0.00	0.00	0.00
		Supplier Name	0.33	0.61	0.40
		Supplier GST Number	0.00	0.00	0.00
		Grand Total Amount	0.00	0.00	0.00
BiLSTM-CRF	Word Embedding Layer	Invoice Number	0.98	0.88	0.93
		Invoice Date	0.76	0.87	0.81
		Buyer Name	0.95	0.97	0.96
		Buyer GST Number	0.60	0.96	0.74
		Supplier Name	0.96	0.72	0.83
		Supplier GST Number	0.97	0.73	0.83
		Grand Total Amount	0.90	0.75	0.82
SVM-RBF	**spatial features**	Invoice Number (FP-Label)	0.91	1.00	0.95
		Invoice Number (FP-Value)	1.00	1.00	1.00
		Invoice Number (AP-Label)	1.00	0.99	1.00
		Invoice Number (AP-Value)	1.00	0.99	0.99
		Invoice Date (FP-Label)	1.00	1.00	1.00
		Invoice Date (FP-Value)	1.00	1.00	1.00
		Invoice Date (AP-Label)	1.00	1.00	1.00
		Invoice Date (AP-Value)	1.00	1.00	1.00

Fig. 8. SVM Evaluation Results using Spatial Features Comparison with BiLSTM and its variant with different Features Extraction Methods

4. CONCLUSION

This research work concludes the proposed solution ICRS (Intelligent Character Recognition System for Account Payable Automation) for invoice key fields data extraction. Its advantages are in different industries, especially in the advertisement invoices industry. The proposed solution is based on Support Vector Machine (SVM). Many previously developed solutions work on text data whereas the proposed solution works with image datasets, some solutions do not work well with multipage invoices whereas our proposed solution does, positional coordinates need not be fixed as compared to other approaches as shown in Figure 7. The overall accuracy is 97 percent. After conducting several experiments, we can conclude that simple machine learning techniques (SVM) gets better results as compared to deep learning complex techniques. This research brings up new possibilities in the realm of key field extraction from multi-layout invoices. The existing research work can be extended by adding the following contributions to the proposed solution:

- Increase the size of the multi-layout invoices dataset. The main goal is to increase the number of different vendor invoices to have a more diverse dataset.
- Data annotation process needs to improve in terms of making it simpler and quicker. As a result, we are looking for a strategy to automatically annotate the different layout invoices.
- Pre-trained neural networks are used. To check the performance of multi-layout invoices, pre-trained neural networks such as BERT and its variations can be used.

REFERENCES

[1] Integromat. 8 Easy Ways to Automate your Invoices (and Save Hours of Your Time) [Internet]. Integromat Blog. [cited 2022 Jun 28]. Available from: https://www.integromat.com/en/blog/invoice-automation

[2] Patel S, Bhatt D. Abstractive information extraction from scanned invoices (AIESI) using end-to-end sequential approach. arXiv preprint arXiv:2009.05728. 2020 Sep 12.

[3] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991. 2015 Aug 9.

[4] Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360. 2016 Mar 4.

[5] Baviskar D, Ahirrao S, Kotecha K. A bibliometric survey on cognitive document processing. Library Philosophy and Practice. 2020 Oct 1:1-31.

[6] Adnan K, Akbar R. Limitations of information extraction methods and techniques for heterogeneous unstructured big data. International Journal of Engineering Business Management. 2019 Dec 9;11:1847979019890771.

[7] Adnan K, Akbar R. An analytical study of information extraction from unstructured and multidimensional big data. Journal of Big Data. 2019 Dec;6(1):1-38.

[8] Baviskar D, Ahirrao S, Kotecha K. Multi-Layout Invoice Document Dataset (MIDD): A Dataset for Named Entity Recognition. Data. 2021 Jul 20;6(7):78.

[9] Palm RB, Laws F, Winther O. Attend, copy, parse end-to-end information extraction from documents. In2019 International Conference on Document Analysis and Recognition (ICDAR) 2019 Sep 20 (pp. 329-336). IEEE.

[10] Reul C, Christ D, Hartelt A, Balbach N, Wehner M, Springmann U, Wick C, Grundig C, Büttner A, Puppe F. OCR4all—An open-source tool providing a (semi-) automatic OCR workflow for historical printings. Applied Sciences. 2019 Nov 13;9(22):4853.

[11] Abbas A, Afzal M, Hussain J, Lee S. Meaningful information extraction from unstructured clinical documents. Proc. Asia Pac. Adv. Netw. 2019 Oct;48:42-7.

[12] Steinkamp JM, Bala W, Sharma A, Kantrowitz JJ. Task definition, annotated dataset, and supervised natural language processing models for symptom extraction from unstructured clinical notes. Journal of biomedical informatics. 2020 Feb 1;102:103354.

[13] Joshi S, Shah P, Pandey AK. Location identification, extraction and disambiguation using machine learning in legal contracts. In2018 4th International Conference on Computing Communication and Automation (ICCCA) 2018 Dec 14 (pp. 1-5). IEEE.

[14] Shah P, Joshi S, Pandey AK. Legal clause extraction from contract using machine learning with heuristics improvement. In2018 4th International Conference on Computing Communication and Automation (ICCCA) 2018 Dec 14 (pp. 1-3). IEEE.

[15] Tkaczyk D, Szostek P, Bolikowski L. GROTOAP2—the methodology of creating a large ground truth dataset of scientific articles. D-Lib Magazine. 2014 Nov;20(11/12).

[16] Yang J, Liu Y, Qian M, Guan C, Yuan X. Information extraction from electronic medical records using multitask recurrent neural network with contextual word embedding. Applied Sciences. 2019 Sep 4;9(18):3658.

[17] Eberendu AC. Unstructured Data: an overview of the data of Big Data. International Journal of Computer Trends and Technology. 2016 Aug;38(1):46-50.

[18] Davis B, Morse B, Cohen S, Price B, Tensmeyer C. Deep visual template-free form parsing. In2019 International Conference on Document Analysis and Recognition (ICDAR) 2019 Sep 20 (pp. 134-141). IEEE.

[19] Zhao X, Niu E, Wu Z, Wang X. Cutie: Learning to understand documents with convolutional universal text information extractor. arXiv preprint arXiv:1903.12363. 2019 Mar 29.

[20] Smith R. An overview of the Tesseract OCR engine. InNinth international conference on document analysis and recognition (ICDAR 2007) 2007 Sep 23 (Vol. 2, pp. 629-633). IEEE.

[21] Smith R. Tesseract ocr engine. Lecture. Google Code. Google Inc. 2007 Jul.

[22] Palm RB, Winther O, Laws F. Cloudscan-a configuration-free invoice analysis system using recurrent neural networks. In2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) 2017 Nov 9 (Vol. 1, pp. 406-413). IEEE.

[23] Katti AR, Reisswig C, Guder C, Brarda S, Bickel S, Höhne J, Faddoul JB. Chargrid: Towards understanding 2d documents. arXiv preprint arXiv:1809.08799. 2018 Sep 24.

[24] Krieger F, Drews P, Funk B, Wobbe T. Information extraction from invoices: A graph neural network approach for datasets with high layout variety. InInternational Conference on Wirtschaftsinformatik 2021 Mar 9 (pp. 5-20). Springer, Cham.

[25] Liu W, Zhang Y, Wan B. Unstructured document recognition on business invoice. Mach. Learn., Stanford iTunes Univ., Stanford, CA, USA, Tech. Rep. 2016.

[26] Schaeffer MS. Essentials of accounts payable. John Wiley & Sons; 2002 Oct 15.

[27] Si Y, Wang J, Xu H, Roberts K. Enhancing clinical concept extraction with contextual embeddings. Journal of the American Medical Informatics Association. 2019 Nov 1;26(11):1297-304.

[28] Wang B, Wang A, Chen F, Wang Y, Kuo CC. Evaluating word embedding models: methods and experimental results. APSIPA transactions on signal and information processing. 2019;8.

[29] Li Y, Liu T, Li D, Li Q, Shi J, Wang Y. Character-based bilstm-crf incorporating pos and dictionaries for chinese opinion target extraction. InAsian Conference on Machine Learning 2018 Nov 4 (pp. 518-533). PMLR.

[30] Majumder BP, Potti N, Tata S, Wendt JB, Zhao Q, Najork M. Representation learning for information extraction from form-like documents. Inproceedings of the 58th annual meeting of the Association for Computational Linguistics 2020 Jul (pp. 6495-6504).