Handling Big Tabular Data Of ICT Supply Chains: A Multi-Task, Machine-Interpretable Approach

Priti P Kohale
Department of Computer Science & Engineering
pmbihade29@gmail.com

Dr. Shailija Sharma
Department of Computer Science & Engineering
shailija.sharma@aiscetuniversity.ac.in

Abstract- **The essential details of ICT devices are frequently distilled into large tabular data sets that are distributed throughout supply chains as a result of the features of Information and Communications Technology (ICT) goods. With the explosion of electronic assets, it is crucial to automatically analyse tabular structures. We develop a Table Structure Recognition (TSR) work and a Table Cell Type Classification (CTC) task to convert the tabular data in electronic documents into a machine-interpretable format and give layout and semantic information for information extraction and interpretation. For the TSR job, complicated table structures are represented using a graph. Table cells are divided into three groups—Header, Attribute, and Data—based on how they work for the CTC job. Then, utilising the text modal and picture modal characteristics, we provide a multi-task model to accomplish the two tasks concurrently. Our test findings demonstrate that, using the ICDAR2013 and UNLV datasets, our suggested strategy can beat cutting-edge approaches.**

Keywords- Big Data Analytics, Supply Chain Optimization, Image Processing, Table Structure Recognition, Table Cell Type Classification

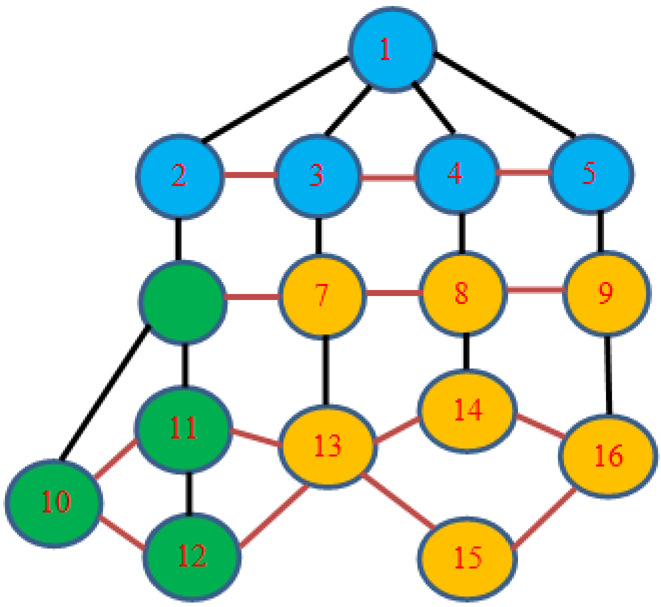
I. INTRODUCTION

It is crucial to extract and understand tables from electronic documents because electronic devices frequently contain different parameters, units, or other crucial information presented as tables in Information and Communications Technology (ICT) supply chains. Despite being user-friendly for human users, tables in electronic documents are frequently unstructured and difficult for machines to understand. The reading, extraction, and interpretation of tables from millions of electronic documents by humans is likewise impractical. As a result, in order to handle the enormous volume of electronic documents in the ICT supply chain and make unstructured tabular data machine-interpretable, we define the Table Structure Recognition (TSR) problem, which can recover a complex table structure with a graph, and the Table Cell Type Classification (CTC) problem based on the functional roles of the cell. In order to express the locational linkages between table cells, which may be represented by the edges in a graph, each cell in a table is represented with a vertex in a graph. Three types of cell connections, namely vertical connection, horizontal connection, and no connection, are specified. In the meanwhile, we establish three sorts of cells, namely Header, Attribute, and Data, which are comparable to several recent research [1], [2], and [3] tackling tabular cell classification difficulties using various taxonomies of cell types. A sample table from the CTC issue with defined types is shown in Figure 1, along with a graph of the table's numbered component. Typically, a table's headers and attributes may convey the meaning of the table's columns and rows, respectively, while data cells are used to display the precise information of each header and attribute. In other words, a table's facts and information may be found and retrieved quickly and readily based on the three types of cells that have been described and the machine-readable table structure. It is important to note that the majority of the current research [1, 2, 3] on CTC problems focuses on spreadsheet tables, which makes problem description considerably simpler because spreadsheets can store more meta-information and their default units are cells. Some research [4], [5] have attempted to extract entities and information from photos, which is similar to how we characterised the CTC issue. However, because they have not given attention to the tables in document images, their problem definitions cannot be solved simultaneously with the TSR problem.

This study's three primary contributions are as follows: 1) this paper creates a benchmark for the CTC problem and extends it to the tables in picture documents and Portable Document Format (PDF) documents. 2) We provide a multi-task strategy that concurrently resolves the specified TSR and CTC issues, producing cutting-edge outcomes. According to experimental findings, using the ICDAR2013 and UNLV datasets for the CTC task, our suggested technique may raise the F1 score from 70.76% to 88.17% and from 71.90 to 83.74, respectively. The suggested strategy may raise the F1 score for the TSR task from 87.24% to 89.51% on the UNLV dataset and from 92.30% to 93.04% on the ICDAR2013 dataset. 3) We examine the many facets of the suggested approach and use tests to demonstrate its efficacy.

MAXIMUM RATINGS AND ELECTRICAL CHARACTERISTICS (T_A= 25 ∘ C UNLESS OTHERWISE NOTED)						
PARAMETER			SYMBOL	TSF30U60C		UNIT
MAXIMUM REPETITIVE PEAK REVERSE VOLTAGE	PER DEVICE		V_{FAV}	60		V
	PER VDIODE					
PEAL FORWARD SURGE CURRENT, 8.3 MS SINGLE HALF SINE- WAVE SUPERIMPOSED ON RATED LOAD PER DIODE			I_{FSM}	30		A
				15		
				250		A
VOLTAGE RATE OF CHANGE (RATED VS)			DV/DT	10000		V/μS
				TYP	MAX	
INSTANTANEOUS FORWARD VOLTAGE PER DIODE (NOTE 1)	$I_F = 15A$	$T_J = 25 \circ C$	V_F	0.48	0.57	V
	$I_F = 15A$	$T_J = 125 \circ C$	V_F	0.43	0.52	μA
INSTANTANEIUS REVERSE CURRENT PER DIODE AT RATED REVERSE VOLTAGE		$T_J = 25 \circ C$	I_R	-	500	mA
		$T_J = 125 \circ C$		-	60	
TYPICAL THEMAL RESISTANCE PER DIODE			$R_{\theta JC}$	4		°C/W
OPERATING JUNCTION TEMPERATURE RANGE			T_J	-55 TO+150		°C
STORAGE TEMPERATURE RANGE			T_{STG}	-55 TO +150		°C

(a)



(b)

Fig. 1. Figure (a) shows three types of table cells, namely header, attribute, and data of an ICT device. Figure (b) is part of the table cells' graph representation, in which each vertex represents a cell, red lines stand for horizontal connection, and black lines stand for vertical connection.

The structure of this essay is as follows: The relevant studies are included in Section II. The entire process of converting tabular data into a machine-interpretable format is shown in Section III, along with the suggested multi-task approach. The experimental findings are presented and discussed in Section IV. In section V, we present our conclusion and potential course of action.

II. Related Work

It is essential but frequently difficult to convert unstructured tabular data into structured and machine-interpretable format because it requires several separate steps to complete this task in order to extract and interpret crucial information from the growing number of electronic documents in the global ICT supply chain. We exclusively concentrate on deep learning-based methods since the most popular research in recent years first converted PDF files into document pictures before using deep learning-based techniques.

A. Tabular Structure Recognition

Due to the complicated table structures, including merging cells and non-explicit boundary lines, TSR's goal is to detect table cells in identical rows and columns. It is common practise to use a graph to describe a complicated table structure. In this design, the table cells are represented by graph nodes, and the linkages between the cells are defined as three types: vertical connections, horizontal connections, and no connections. This issue formulation is followed by several research [6], [7], which provide bottom-up strategies. Contrarily, other research utilising top-down methods, such as DeepDeSRT [8], CascadeTabNet [9], and TableDet [10], would describe structural identification as an object detection or segmentation problem, frequently in conjunction with a table detection challenge. To enhance performance, these techniques frequently use and expand Cascade R-CNN [11], Mask-RNN [12], transfer learning, and data augmentation techniques.

B. Table Cell Type Classification

For the spreadsheet CTC problem, several research use text embeddings and aesthetic elements. In their study [1], Elvis et al. employ Weka [13] to choose features and test several tree-based classifiers. Pre-trained cell embedding and stylistic features are suggested by Majid et al. in their study [2]. In order to create cell contexts, they develop a neighbor-based method, suggest an embedding model based on the InferSent model [14], and create a classifier using LSTM. Additionally, language models have been applied to the categorization of tabular cells. A vast corpus and numerous computing resources are needed for the training of a large language model. The pre-trained language model is initially utilised to create the feature embeddings and hone a cell type classification model before being applied to the CTC job. A huge language model for a broadly organised table called TUTA [3] was developed using a large corpus of spreadsheets. In addition, other language models, such BERT [15], can also be used to the CTC issue. These research all concentrate on spreadsheets because they provide intrinsic structural data that may be used to train and improve context-based language models. Tabular data in PDF documents, on the other hand, do not have this kind of internal structural information, which makes the work more difficult.

III. Proposed Method

As shown in Fig. 2, manufacturers produce "large tabular data," which contains important information in tables but is difficult to communicate with other supply chain participants. By converting unstructured tabular data into a structured and machine-interpretable format, our suggested approach paves the way for effective information exchange throughout the whole supply chain.

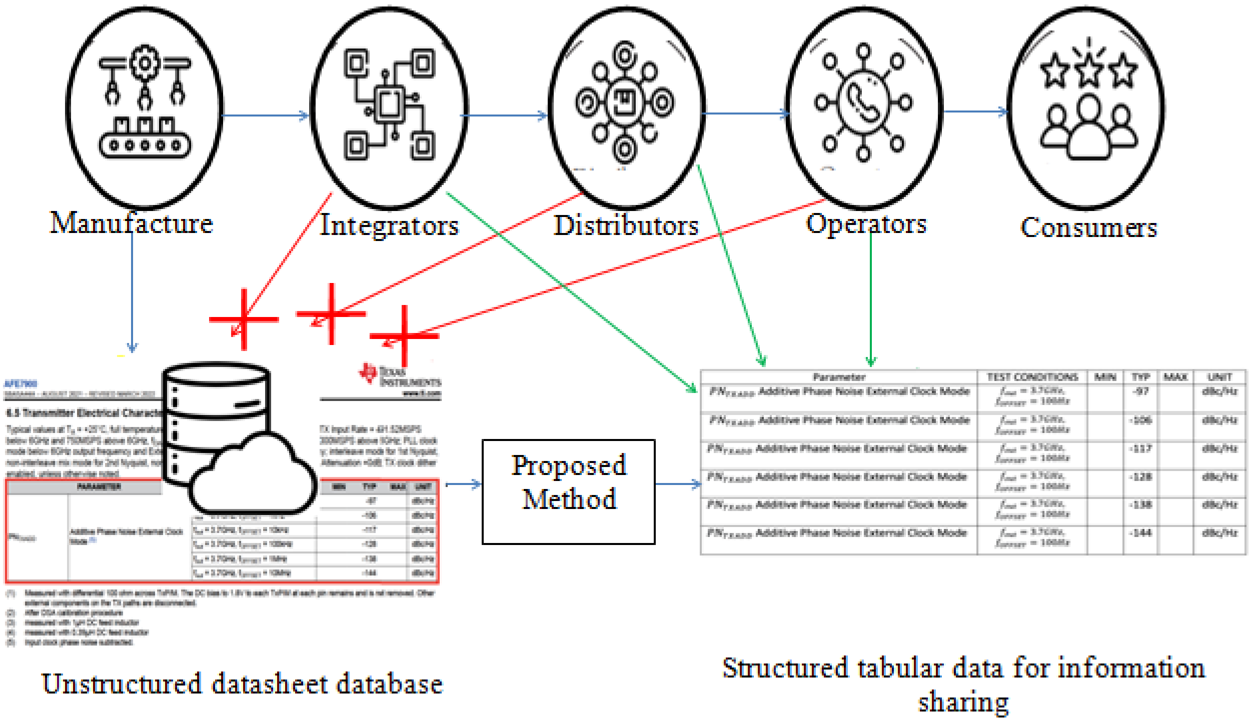


Fig. 2. A simplified ICT supply chain sample.

A. Problem Definition

The goal of TSR is to recover complicated table structures into a format that is organised and machine-interpretable. We also employ a graph to describe a complicated table structure, as illustrated in Figure 1. Graph vertexes are used to represent table cells and graph edges are used to signify the relationships between two cells in a table, following the problem specification in some current bottom-up techniques [7]. Assuming that the bounding box for each table cell is known, it may be written as $\{x_i^1, x_i^2, y_i^1, y_i^2\}$ where i is the cell number. Then, a collection of cells may be used to represent the kth table (t_k) in a table set $T = \{t_k; k \in K\}$ comprising K tables $C_k = \{c_i^k; i \in N_k\}$, where c_i^k denotes the ith cell in the table t_k and

N_k is the total number of cells in the table. Given a training set of K tables, the table $t_k's$ cell association set $R_k = \{r^k_{\{i,j\}}; i \neq j, i \in N_k, k \in K\}$, and $r^k_{\{i,j\}} = \{\{c^k_i, c^k_j\}; i \neq j, i \in N_k, k \in K\}$ the associated label set $Y^{tsr}_k = \{y^k_{\{i,j\}}; y \in \{0,1,2\}, i \in N_k, j \in N_k, k \in K\}$ of the association set, where 0, 1, and 2 denote horizontal, vertical, and no connectivity, respectively, between the two cells in the association. In order to recognise table structures, it is necessary to train a predictive model with training data that can calculate the probability of a cell association between two cells, or $P_\theta(y^k_{\{i,j\}} = \hat{y}^k_{\{i,j\}} | r^k_{\{i,j\}}), \hat{y}^k_{\{i,j\}} \in \{0,1,2\}$.

TABLE I. AN OVERVIEW OF THE NOTATIONS USED IN THIS PAPER.

T	***Set of a table***
C_k	This table's kth table's cell set
R_k	Cell associations in the kth table
Y^{tsr}_k	The kth table's label set for the TSR
Y^{ctc}_k	The kth table's label set for the CTC
$\{x^1_i, x^2_i, y^1_i, y^2_i\}$	The coordinates of a table's ith cell
c_i	ith cell in a table
t_k	The kth table in a table set T containing K tables
$r^k_{\{i,j\}}$	The relationship between the ith and jth cells in the kth table.
$\mathcal{L}$	The loss functions
$\theta, \emptyset, \omega, \sigma$	The variables that each function can be trained on
$\mathcal{F}$	A fully connected layer
F	A completely linked layer's output
CAT	The function that can represent the fusion layer
C	The classifier
$\mathcal{E}$	The embedding network
x	The input of a function
Y	The output of the proposed method
$\mathbb{R}^{ch*h*w}$	A real value set with ch channel, h height, w width

A complex table structure's functional functions are to be determined by the CTC issue. For the table cell classification algorithm, each table t_k contains a set of cells $C_k = \{c^k_i; i \in N_k\}$ and their corresponding types $Y^{ctc}_k = y^k_i; i \in N_k$, given a training set made up of K tables. We specify three cell types: header, attribute, and data, which means that $y^k_i \in \{0,1,2\}$respectively. Therefore, the goal of table cell classification is to develop a prediction model using training data that can estimate the likelihood that a cell belongs to one of a set of specified cell types, namely$P_\emptyset(y^k_i = \hat{y}^k_i | c^k_i), \hat{y}^k_i \in \{0,1,2\}$ .Notably, the suggested method is a bottom-up strategy and necessitates the knowledge of each cell's enclosing boxes. MMOCR [16] is used in our solution to identify table cells and extract text from table pictures. Table I provides a summary of the notations used in this article.

B. Multi-task Table Structure Recognition and Table Cell Type Classification

Table detection, table structure identification, and classification are the three phases required to convert unstructured tabular data into a structured and machine-interpretable format. Fortunately, some research on table detection has produced encouraging results, as mentioned in section II. In reality, we first use TableDet [10] to extract all the tables from the document images in order to produce the table set $T = \{t_k; k \in K\}$. We next convert all of the PDF documents into document pictures. We apply the well-known KD-tree based K-nearest technique, which is also employed in previous works [7], [17], to construct the cell association set required by the table structure identification task. More specifically, we may use the bounding boxes of two cells to compute the distance between them using a predefined distance metric, often Euclidean Distance. So, for the ith cell in the kth table t_k, we can simply perform this method by locating its closest top M neighbour cells to generate association pairings. Notably, in our implementation, M is a hyperparameter with a value of 20.

Despite the fact that PDF documents and document pictures can only offer a very limited amount of meta-information, they may nonetheless naturally contain both text and visual information, which allows features to be retrieved from both text and image modalities. The suggested model has two branches, one for the TSR task and the other for the CTC task, as illustrated in Figure 3. This is typical of multi-task techniques. Only the characteristics from the picture modal, which may be retrieved by an embedding network, are used by the TSR branch. The CTC branch, in contrast, draws its characteristics from the coordinates of the cells as well as the picture and text modals. More specifically, we use the cropped cell images and context images as the input in the table structure recognition branch in the TSR branch, as shown in the input sample portion of Figure 3. This is done in accordance with the design of study [17], where we crop both the cell images in each cell association pair and their context images. After that, to extract visual feature maps, the resulting pictures are put into an embedding network, represented by the function $\varepsilon_\emptyset$. Then, a fully connected layer that may be represented by a function $F^{tsr}_{\theta_1}$ is fed the extracted feature maps. The classifier $C^{tsr}_{\sigma_1}$ is then further fed the outputs of $F^{tsr}_{\theta_1}$ using a softmax function to determine the likelihood of each class. Equation 1, where $ch, h, and\ \omega$ are the number of channels, image height, and image width, respectively, and all $\theta, \sigma, \emptyset$ are trainable parameters, may be used to express the entire process.

$$\boldsymbol{y_{tsr}} = C^{tsr}_{\sigma_1}(f_{tsr}), f_{tsr} = F^{tsr}_{\theta_1}\left(\varepsilon_\emptyset(x_{tsr})\right), x_{tsr} \in \mathbb{R}^{ch*h*w}, f_{tsr} \in \mathbb{R}^{l_1}, \boldsymbol{y_{tsr}} \boldsymbol{\epsilon} \mathbb{R}^{\boldsymbol{3}} \quad (1)$$

The cropped cell pictures and their related words are utilised as the inputs of the CTC branch in conjunction with the coordinates of the cells. The text features may be determined using Equation 2 where f_2 is the output feature and l_{text} is the number of tokens in the input phrase. Assuming that the InferSent [14] module can be represented by function I_w, the fully connected layer for the text modal can be symbolised by function $F^{ctc}_{\theta_2}$.

$$f_{text} = F^{ctc}_{\theta_2}\left(I_w(x_{text})\right), x_{text} \in \mathbb{R}^{l_{text}}, f_{text} \in \mathbb{R}^{l_2} \quad (2)$$

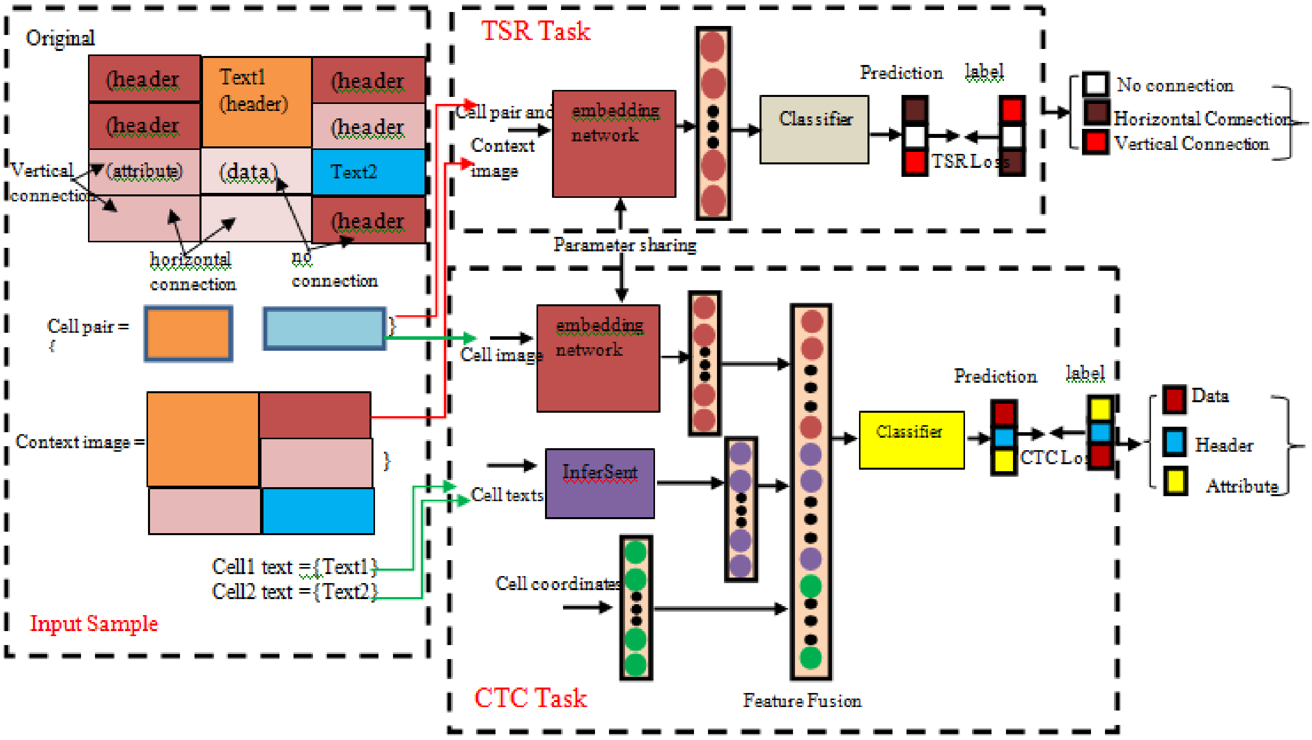


Fig. 3. Overall architecture of the proposed multi-task model. Notably, header, attribute and data are three types of cell in the CTC task, and horizontal connection, vertical connection and no connection are three possible outputs in the TSR task.

The Equations 3 and 4 may also be used to derive the coordinate features and image features in the CTC branch. Finally, using Equation 5, all features are combined using the fusion function CAT and sent into the classifier $C_{\sigma_2}^{ctc}$. It is important to note that $\varepsilon_\emptyset$ shares characteristics in both branches and that l_1, l_2, l_3 and l_4 depend on the number of neurons in each completely connected layer.

$$f_{img} = F_{\theta_2}^{ctc}\left(\varepsilon_\emptyset\left(x_{img}\right)\right), x_{img} \in \mathbb{R}^{ch*h*w}, f_{img} \in \mathbb{R}^{l_3} \quad (3)$$

$$f_{coord} = F_{\theta_3}^{ctc}(x_{coord}), x_{coord} \in \mathbb{R}^4, f_{coord} \in \mathbb{R}^{l_4} \quad (4)$$

$$\boldsymbol{y}_{ctc} = C_{\sigma_2}^{ctc}\left(CAT\left(f_{img}, f_{text}, f_{coord}\right)\right), \boldsymbol{y}_{ctc} \in \mathbb{R}^3 \quad (5)$$

The loss function $\mathcal{L}_{mt}$ may be described as Equation 6, which incorporates a TSR loss represented by $\mathcal{L}_{tsr}$, a CTC loss denoted by $\mathcal{L}_{ctc}$, and a hyper parameter to balance $\mathcal{L}_{tsr}$ and $\mathcal{L}_{ctc}$ since the suggested technique uses multi-task architecture, which means that all the parameters should be trained jointly. In our implementation, we employ cross entropy loss, which is defined as Equation 7, where $\boldsymbol{y}_i$ is the ith prediction and $\boldsymbol{y}_i$ represents the matching ground truth.

$$\mathcal{L}_{mt} = (1-\lambda) * \mathcal{L}_{tsr} + \lambda * \mathcal{L}_{ctc} \quad (6)$$

$$\mathcal{L}_{ce}(y, \hat{y}) = -\sum_{i=1}^{N} \hat{y}_i \log y_i \quad (7)$$

IV. Experiments and Analysis

Table identification and TSR issues are frequently addressed using the 156 and 540 table ICDAR2013 and UNLV datasets, respectively. As described in section III-A, we additionally annotate these two datasets with "Header," "Attribute," or "Data," and we exclude all of the empty cells from the CTC task labelling since empty cells are simple to recognise without the use of any machine learning models. The ICDAR2013 dataset is divided into a validation set with 33 tables and a testing set with 123 tables at random for the trials. Additionally, the UNLV dataset is randomly divided into three sets, each comprising 323, 107, and 110 tables: training, validation, and testing. In keeping with the experimental design of the work [17], we delete 18 tables from the UNLV dataset due to the uncertainty of their labels and utilise the training set of the UNLV dataset for testing the performance on the ICDAR2013 dataset. For the ICDAR2013 and UNLV datasets, we utilise MMOCR [16] to extract text contents from the table cells.

A. Implementation Details and Experimental Results

The suggested approach includes an embedding network E that is utilised in both the TSR task and the CTC job, as was covered in section III-B. We employ a straightforward ConvNet-4 [18] in our implementation, which has four convolutional layers. The input pictures of the embedding network are first downsized to the dimension $3 * 84 * 84$ with zero padding to maintain the original height-width ratio, in accordance with the procedure stated in the paper [17]. The embedding network $\varepsilon_\emptyset$ is then fed the shrunk pictures, and a fully connected layer F is added after that to retrieve the features. In our implementation, l_1 and l_3 are both equal to 128 since we set the number of neurons in $F_{\theta_1}^{tsr}$ and $F_{\theta_2}^{ctc}$to 128. Similarly, for the ICDAR2013 dataset and the UNLV dataset, l_2 and l_4 are set to 128 and λ 32, respectively, and in the loss, function is set to 0.3. A fully connected layer with three neurons and a softmax function are used to convert the logits into a distribution for each classifier, as illustrated in Figure 3. We implement the fusion function CAT, which lacks any trainable parameters, using a straightforward concatenation function. Notably, we employ a cost-sensitive strategy in the development of the loss functions due to the imbalanced datasets utilised in the trials.

We list the following models as benchmark models for the TSR task: TabbyPDF [19], GraphTSR [7], DeepDeSRT [8], and CATT-Net [17]. We implement four models by fine-tuning the Glove [20], FastText [21], and BERT [15] in accordance with the well-known pre-trained fine-tune paradigm in order to compare the state-of-the-art NLP techniques. This is because our proposed multi-task model depends on features from both image modal and text modal. Without utilising pretrained weights, we build two image classification models, ConvNeXt [22] and ResNet50 [23], for the image modal. Glove and FastText are two common examples of non-context-based models, whereas BERT is a well-known context-based language model, indicating that the performance of these two models largely depends on lengthy reliance in a context. ConvNeXt [22] is the most advanced model for the image classification problem, whereas ResNet50 is a traditional convolution network that is frequently employed in image classification tasks. Without taking into account the context information of the tables, we solely concentrate on the tables and use TableDet [10] to extract all of the tables from the document pictures. As a result, fine-tuning BERT performs very poorly for the CTC problem and is not included in Table 3.

The suggested method's overall Precision, Recall, and F1-score for the TSR task and CTC task, respectively, are displayed in Tables 2 and 3. The study [7] yielded the TabbyPDF, GraphTSR, and DeepDeSRT performance ratings on the ICDAR2013 dataset. The score is not given in linked investigations, as shown by the "-" in Table 2. The experimental findings demonstrate that the suggested strategy can perform better in the TSR and CTC tasks than the benchmark models.

TABLE II. EXPERIMENTAL RESULTS FOR THE TSR TASK.

	ICDAR2013			UNLV		
Method	Prec	Recall	F1	Prec	Recall	F1
TabbyPDF	78.90	84.50	81.60	-	-	-
GraphTSR	81.90	85.50	83.70	-	-	-
DeepDeSRT	57.30	56.40	56.80	-	-	-
CATT-Net	94.10	90.70	92.30	86.28	88.31	87.24
Ours	92.85	93.29	93.04	92.66	86.78	89.52

TABLE III. EXPERIMENTAL RESULTS FOR THE CTC TASK

	ICDAR2013			UNLV		
Method	Prec	Recall	F1	Prec	Recall	F1
Glove	56.47	60.88	57.74	55.00	62.03	57.33
FastText	67.85	64.16	65.46	63.72	68.85	65.85
ConvNeXt	70.15	71.42	70.76	65.88	66.72	66.21
ResNet50	72.81	69.03	70.16	72.62	71.83	71.90
Ours	87.94	88.41	88.17	82.19	85.51	83.74

B. Discussion and Analysis

1) Ablation Study: In this part, we run more tests to show how well each element of our suggested strategy works. First, we use the relevant branch design of the suggested technique to create two single-task models for the TSR task and CTC, respectively. We additionally design three models that only employ features from each modal because the CTC task makes use of features from the coordinate information, the text modal, and the picture modal. The experimental findings are displayed in Tables 4 and 5, where $S(I), S(T), S(C), and\ S(I + T + C)$ denote the single-task model with coordinate feature, text feature, and combined three types of features, respectively, while our denotes the suggested multi-task technique.

The experimental findings in Table 4 reveal that when paired with characteristics from the picture modal, features from the text modal and coordinate information are ineffective and seldom ever beneficial. As opposed to other benchmarks, picture modal visual aspects are more effective and can result in the best performance. Therefore, in our suggested multi-task solution, we solely employ visual features in the TSR task. In the meanwhile, Table 5's experimental findings demonstrate that the CTC issue may benefit from visual features, text features, and coordinate information, and that combining these two types of features can significantly boost performance. Therefore, in the CTC branch of the suggested multi-task technique, we make use of visual characteristics, text features, and coordinate information.

TABLE IV. ABLATION STUDY RESULTS FOR THE TSR TASK.

	ICDAR2013			UNLV		
Method	Prec	Recall	F1	Prec	Recall	F1
S (T)	36.36	34.56	34.32	40.03	35.91	35.94
S (T)	89.42	91.71	90.52	84.39	87.52	85.75
S (C)	50.81	77.69	45.30	60.38	83.75	61.17
S (T+I+C)	88.71	93.53	90.98	90.57	84.16	87.11
Ours	92.85	93.29	93.04	92.66	86.78	89.52

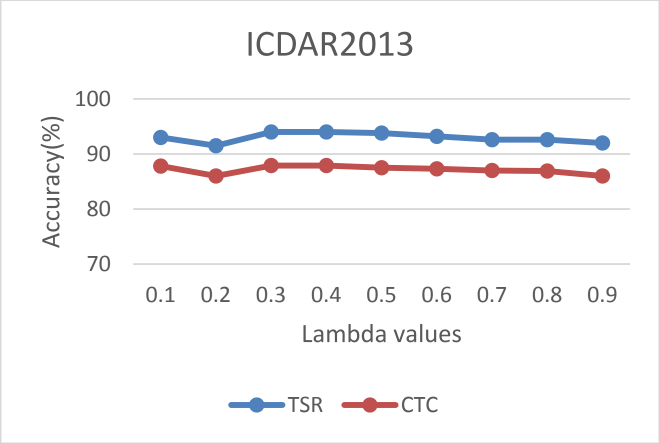
TABLE V. ABLATION STUDY RESULTS FOR THE CTC TASK.

	ICDAR2013			UNLV		
Method	Prec	Recall	F1	Prec	Recall	F1
S (T)	68.11	62.46	64.48	72.047	67.85	69.41
S (T)	70.12	67.52	68.46	63.98	76.49	68.62
S (C)	86.07	90.84	87.65	76.98	85.55	80.74
S (T+I+C)	85.44	88.68	86.53	77.75	81.59	78.70
Ours	87.94	88.41	88.17	82.19	85.51	83.74

2) The impact of hyper parameter lambda: Equation 6 illustrates the TSR branch and CTC branch in our suggested approach, both of which result in a loss. The loss of the CTC task and the loss of the TSR job are balanced using the hyperparameter. Figure 4 depicts the outcomes of further tests we ran to examine how this hyperparameter affected the suggested model. The figure demonstrates how the value of λ can affect how well the suggested strategy performs, but it can also maintain an overall stable performance.

3) The impact of feature fusion: As seen in Figure 3, the features for the CTC task are derived from many modalities, such as the text modal, the picture modal, and the coordinates of the cells. Before being input into the classifier, these features are combined by the

feature fusion layer. As a result, we run additional tests to see how these variables are affected. Any integer greater than zero may represent the number of neurons in a fully linked layer. We assume $l_1 = l_2 = l_3 = l_4$ in this section in order to simplify the issue and take into account the results presented in Table 5.



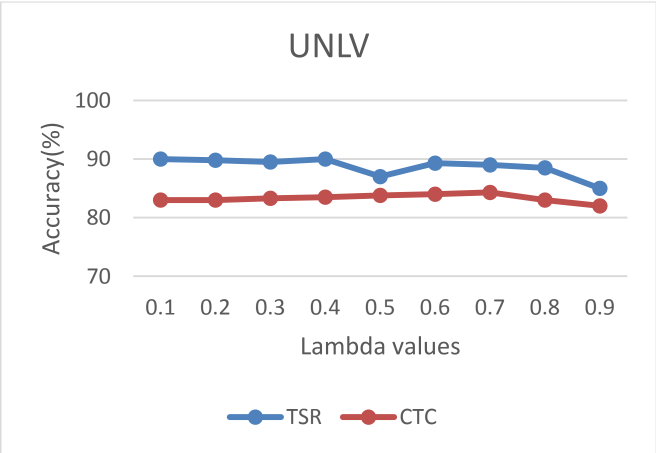


Fig. 4. FigureExperimental results with different λ values.

The experimental results are shown in Tables 6 and 7 when l_1 is l_6, 32, 64, and 128 correspondingly. The findings demonstrate that adding more neurons can improve performance to some extent, particularly for TSR.

TABLE VI. EXPERIMENTAL RESULTS FOR THE TSR TASK.

	ICDAR2013			UNLV		
Method	Prec	Recall	F1	Prec	Recall	F1
l16	81.24	92.57	85.86	81.65	82.25	81.65
I32	84.75	93.12	88.49	79.77	82.86	80.55
I64	85.69	92.51	88.75	79.85	85.74	82.53
I128	80.13	93.60	85.54	87.41	85.66	86.45

TABLE VII. EXPERIMENTAL RESULTS FOR THE CTC TASK.

	ICDAR2013			UNLV		
Method	Prec	Recall	F1	Prec	Recall	F1
l16	79.76	82.48	78.83	82.84	81.94	82.26
I32	83.54	83.50	82.29	81.02	85.04	82.92
I64	89.47	81.22	84.72	80.76	87.09	83.57
I128	90.83	84.33	87.22	79.36	87.18	82.86

V. CONCLUSION AND FUTURE WORK

ICT supply chain participants produce "huge tabular data," which is challenging to handle but frequently has significant information. In order to convert tabular data into a machine-interpretable format, this work presented a multi-task model that concurrently addresses the TSR and CTC issues. According to experimental findings, multitask designs are superior to single-task designs for both TSR and CTC tasks. The F1 score for the CTC task on the ICDAR2013 and UNLV datasets, respectively, increased from 70.76% to 88.17% and from 71.90 to 83.74 using the suggested technique, which can also significantly beat the state-of-the-art benchmark models. Increase the F1 score for the TSR task from 87.24% to 89.51% for the UNLV dataset and from 92.30% to 93.04% for the ICDAR2013 dataset. While the two issues can potentially be dealt with sequentially in future work, we handle the TSR and CTC together with a single multi-task model.

ACKNOWLEDGEMENTS

The authors would like to thank everyone, just everyone!

REFERENCES

[1] Burak Omer Saracoglu, Initialization of profile and social network analyses robot and platform with a concise systematic review, Machine Learning with Applications, Volume 7, 2022, 100249, ISSN 2666-8270, https://doi.org/10.1016/j.mlwa.2022.100249.

[2] Jannatul Ferdous Ruma, Sharmin Akter, Jesrin Jahan Laboni, Rashedur M. Rahman, A deep learning classification model for Persian Hafez poetry based on the poet's era, Decision Analytics Journal, Volume 4, 2022, 100111, ISSN 2772-6622, https://doi.org/10.1016/j.dajour.2022.100111.

[3] Lucheng Hong, Zehua Chen, Yifei Wang, Mohammad Shahidehpour, Minghe Wu,A novel SVM-based decision framework considering feature distribution for Power Transformer Fault Diagnosis, Energy Reports, Volume 8, 2022, Pages 9392-9401,ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2022.07.062.

[4] Sara Barja-Martinez, Mònica Aragüés-Peñalba, Íngrid Munné-Collado, Pau Lloret-Gallego, Eduard Bullich-Massagué, Roberto Villafafila-Robles, Artificial intelligence techniques for enabling Big Data services in distribution networks: A review, Renewable and Sustainable Energy Reviews, Volume 150, 2021, 111459, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2021.111459.

[5] G. Tang, L. Xie, L. Jin, J. Wang, J. Chen, Z. Xu, Q. Wang, Y. Wu, and H. Li, "Matchvie: Exploiting match relevancy between entities for visual information extraction," arXiv preprint arXiv:2106.12940, 2021.

[6] Jonathan Cohen, Jorge Gil, An entity-relationship model of the flow of waste and resources in city-regions: Improving knowledge management for the circular economy, Resources, Conservation & Recycling Advances, Volume 12, 2021, 200058, ISSN 2667-3789,https://doi.org/10.1016/j.rcradv.2021.200058.

[7] Zhaoyang Qu, Zhenming Zhang, Shuai Liu, Jie Cao, Xiaoyong Bo, Knowledge-driven recognition methodology for electricity safety hazard scenarios,Energy Reports, Volume 8, 2022, Pages 10006-10016, ISSN 2352-4847,https://doi.org/10.1016/j.egyr.2022.07.158.

[8] Yang Huang, Zizhen Li, Qiyang Hong, Lizhi Zhou, Yue Ma, Yisha Hu, Jiabao Xin, Tingting Li, Zhibo Kong, Qingbing Zheng, Yixin Chen, Qinjian Zhao, Ying Gu, Jun Zhang, Yingbin Wang, Hai Yu, Shaowei Li, Ningshao Xia, A stepwise docking molecular dynamics approach for simulating antibody recognition with substantial conformational changes, Computational and Structural Biotechnology Journal, Volume 20, 2022, Pages 710720, ISSN 2001-0370, https://doi.org/10.1016/j.csbj.2022.01.012.

[9] Stefania Fresca, Andrea Manzoni, POD-DL-ROM: Enhancing deep learning-based reduced order models for nonlinear parametrized PDEs by proper orthogonal decomposition,Computer Methods in Applied Mechanics and Engineering, Volume 388, 2022, 114181, ISSN 0045-7825, https://doi.org/10.1016/j.cma.2021.114181.

[10] Kazuhide Mimura, Shugo Minabe, Kentaro Nakamura, Kazutaka Yasukawa, Junichiro Ohta, Yasuhiro Kato, Automated detection of microfossil fish teeth from slide images using combined deep learning models, Applied Computing and Geosciences, 2022, 100092, ISSN 2590-1974, https://doi.org/10.1016/j.acags.2022.100092.

[11] Ruihui Xue, Wei Xiang, Yansong Deng, Improved Faster R-CNN Based On CSP-DPN, Procedia Computer Science, Volume 199, 2022, Pages 1490-1497, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.01.190.

[12] Hongjie He, Hongzhang Xu, Ying Zhang, Kyle Gao, Huxiong Li, Lingfei Ma, Jonathan Li,Mask R-CNN based automated identification and extraction of oil well sites, International Journal of Applied Earth Observation and Geoinformation, Volume 112, 2022, 102875, ISSN 1569-8432, https://doi.org/10.1016/j.jag.2022.102875.

[13] Judith Santos-Pereira, Le Gruenwald, Jorge Bernardino, Top data mining tools for the healthcare industry, Journal of King Saud University - Computer and Information Sciences,Volume 34, Issue 8, Part A, 2022, Pages 4968-4982, ISSN 1319-1578,https://doi.org/10.1016/j.jksuci.2021.06.002.

[14] Shuangyin Li, Weiwei Chen, Yu Zhang, Gansen Zhao, Rong Pan, Zhenhua Huang, Yong Tang,A context-enhanced sentence representation learning method for close domains with topic modeling, Information Sciences, Volume 607, 2022, Pages 186-210, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2022.05.113.

[15] Seonghyeon Moon, Seokho Chi, Seok-Been Im, Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT), Automation in Construction, Volume 142, 2022, 104465, ISSN 0926-5805, https://doi.org/10.1016/j.autcon.2022.104465.

[16] Jasper Schwenzow, Jochen Hartmann, Amos Schikowsky, Mark Heitmann, Understanding videos at scale: How to extract insights for business research, Journal of Business Research,Volume 123, 2021, Pages 367-379, ISSN 0148-2963, https://doi.org/10.1016/j.jbusres.2020.09.059.

[17] B. Xiao, M. Simsek, B. Kantarci, and A. A. Alkheir, “Table structure recognition with conditional attention,” Preprint: https://arxiv.org/abs/2203.03819, 2022.

[18] Arturs Polis, Alexander Ilin, A Relational Model for One-Shot Classification of Images and Pen Strokes, Neurocomputing, Volume 501, 2022, Pages 1-13, ISSN 0925-2312,https://doi.org/10.1016/j.neucom.2022.06.004.

[19] Duy Duc An Bui, Guilherme Del Fiol, Siddhartha Jonnalagadda, PDF text classification to leverage information extraction from publication reports Journal of Biomedical Informatics, Volume 61, 2016, Pages 141-148, ISSN 1532-0464, https://doi.org/10.1016/j.jbi.2016.03.026.

[20] Archana Kumari, D.K. Lobiyal, Efficient estimation of Hindi WSD with distributed word representation in vector space, Journal of King Saud University - Computer and Information Sciences, 2021, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2021.03.008.

[21] An Dao, Natthawut Kertkeidkachorn, Ryutaro Ichise, Comparison of Deep-Neural-Network-Based Models for Estimating Distributed Representations of Compound Words,Procedia Computer Science, Volume 192, 2021, Pages 1294-1303, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.08.133.

[22] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A convnet for the 2020s,” arXiv preprint arXiv:2201.03545, 2022.

[23] Siliang Chen, Xu Zhu, Kang Chen, Zexu Liu, Pengcheng Li, Xinbin Liang, Xinqiao Jin, Zhimin Du, Applying deep learning-based regional feature recognition from macro-scale image to assist energy saving and emission reduction in industrial energy systems, Journal of Advanced Research, 2022,,ISSN 2090-1232, https://doi.org/10.1016/j.jare.2022.07.003..