

TABLE 1: Literature of information extraction from various types of documents.

S/N	Document type	Technique	Approach	Authors
1	Invoices	(i) Bidirectional LSTM deep neural network and trained data extracted end-to-end from invoice	Machine-based	[2]
		(ii) Named entity recognition using BERT (bidirectional encoder representations from transformers)	Machine-based	[54]
		(iii) Optical character recognition and graph convolution network from invoice images	Machine-based	[53]
2	Financial reports	(i) Detection of key performance indicators (KPI) from a report using the density of alpha-numeric characters in a rule-based fashion	Rule-based	[16]
3	Medical clinical notes	Parse meaningful critical values from clinical notes and perform a semantic lookup	Rule-based	[21, 55]
4	Legal documents: (i) Court record docs (CRDs) (ii) Compliance documents	(i) Bidirectional LSTM for training and extracting information	Machine-based	[17]
		(ii) Context-free grammar for complex rule interpretation	Rule-based	[56]
5	Software requirements documents	Syntactic and semantic analysis approach to align with standard writing best practices	Rule-based	[15]
6	CVs	Rule-based text extraction from CV	Rule-based	[49, 57]
7	Academia: literature research	Optical character recognition and graph convolution network from invoice images	Machine-based	[19, 20]