Initial Research and Literature Review

Title	Methodology	Strengths	Limitations
An attention-based method for offline handwritten Urdu text recognition (ICFHR, 2020)(Anjum and Khan 2020) [1]	•Encoder Decoder Based •CNN, DenseNet, BLSTM	•Position Change •Attention Mechanism •Relevant Context	•Prediction by text character •Separate model for error correction
A convolutional recursive deep architecture for unconstrained Urdu handwriting recognition (Neural Computing & application, 2021) [2]	•Urdu Handwriting •Convolution + BLSTM •Lastly n-gram model	•State of art approaches •Information loss avoid •Image quality	•Separate n-gram model •Lack of longer context •Prone to Error
A Computationally Efficient Pipeline Approach to Full Page Offline Handwritten Text Recognition (ICDARW, 2019) [3]	•Word by word Localization •CNN-BLSTM + Language model	•Line by Line less expensive •Multi-down sampled •Combining BLSTM	•Localize the text word by word •Space issue •Localize only English text
An online cursive handwritten medical words recognition system for busy doctors in developing countries for ensuring efficient healthcare service delivery (Scientific reports, 2022) [4]	•Line sequence from image •BLSTM network	•SOTA results •The whole sequence of characters •Taking context in account	•Works only for online handwritten text extraction •Requires doctors to use a smartpen to write.
Multilingual handwritten numeral recognition using a robust deep network joint with transfer learning (Information Sciences (Elsevier),2021) (Fateh, et al. 2021) [5]	**•Multilingual Handwritten numbers •Deep CNN •Limit to Chinese, Arabic, English, Kannada, Persian, and Urdu •Language recognition and digit recognition •Best model parameters for the recognition of digits**	**•SOTA Result •Employing Transfer Learning**	**•Limited to numbers only •Multiple languages can lead to error in text localization •High error rate possible**

Initial Research and Literature Review

Title	Methodology	Strengths	Limitations
TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models (Minghao Li, et al. 2021)	•Transformer Architecture •Uses pre-trained CN and NLP models •Splits image into sequence of patches that are used as inputs	•SOTA Results •Uses pre-trained CN and NLP models, which take advantage of large-scale unlabeled data for image understanding and language modeling, with no need for an external language model •Does not require CNN for backbone, so image-specific biases are avoided	•Requires huge amount of data •Not suitable for low-resource languages (e.g Urdu), due to the nature of transformer architecture as it requires large amounts of data
LayoutLM: Pre-training of Text and Layout for Document Image Understanding (ACM, 2019)(Yiheng Xu, et al. 2019)	•Uses both text and document layouts for training •Joint training in textual and layout information •BERT is used as the backbone, and adds two new input embeddings: Positional and image embedding •Positional embeddings to capture relationship among tokens within a document	•SOTA Results •Takes into account both textual and layout information, which is beneficial for a great number of real-world document image understanding tasks such as information extraction from scanned documents	•Only works for English (will not work for Urdu, or multilingual use cases) •Needs a separate model for text extraction and localization, alongside the LayoutLM model itself, which is quite computationally heavy