International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 11 | Nov - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Bill Recognition and Structured Data Extraction from Handwritten Bills

Patil Pranav Nimba [1], Dr. Manoj S. Ishi [2], Deore Chetan Anil [3], Patil Gopal Dattu [4], Patil Chetan Arvind [5].

Department of Computer Engineering R. C. Patel Institute of Technology Shirpur, India
Department of Computer Engineering R. C. Patel Institute of Technology Shirpur, India
Department of Computer Engineering R. C. Patel Institute of Technology Shirpur, India
Department of Computer Engineering R. C. Patel Institute of Technology Shirpur, India
Department of Computer Engineering R. C. Patel Institute of Technology Shirpur, India
Email: patil.pranav4415@gmail.com, manoj.ishi@rcpit.ac.in, deorechetan24@gmail.com, patilgopal3589@gmail.com, chetanpatil.master@gmail.com

Abstract :- Handwritten bills represent the main transaction records for many small retailers, medical stores, and other small-scale enterprises, but getting the data into digital records is currently a laborious, error-prone data entry task. Conventional OCR engines are ill-suited to the challenges posed by recognizing handwritten content, and they return unstructured text rather than the field-level records desired by accountants and inventory managers. This paper performs a literature review of recent research on handwritten text recognition, receipt/invoice OCR, camera-document preprocessing, and layout-aware information extraction. A gap in the literature is identified: there is no study that combines elements of all of these areas into a system that takes handwritten content captured by smartphones, extracts structured, valid field-level records, and exports them into Excel or a lightweight SQL database for accounting and inventory management purposes. A new Bill Recognition and Structured Data Extraction System is proposed, which implements camera-aware preprocessing, text-region detection, handwritten text recognition using a CNN-BiLSTM-CTC (optional Transformer) design, layout-aware field identification, rule-based data validation, and Excel or SQL export. Methodology, system design, dataset requirements, and an evaluation plan are discussed, including character and word error rate, precision-recall-F1-score at the field level, exact-match accuracy, and processing time, with comparisons to generic OCR engines, detection-followed-by-recognition pipelines, and multimodal large language models. The proposed design can reduce the labor and improve accuracy of converting handwritten transaction records into digital format for small businesses.

***Keywords*—handwritten bill recognition, deep learning, CRNN, BiLSTM, information extraction, data validation**

I. INTRODUCTION

Handwritten bills and invoices are still a significant part of day-to-day commerce, for small retailers, medical stores and other smaller establishments; even as digital record-keeping becomes more and more important in tax, accounting and inventory management. In the majority of such places, converting a handwritten bill into a digital record is a manual process. A clerk or shop-owner has to retype the bill number, date, items, prices and totals into a sheet or spreadsheet. Not only is this a non-scalable process, as the transaction volume increases it involves more and more laborious data entry work, which is a prime source of error (misread digits, transposed figures, missing fields) especially when turnover is high, or the staff is inexperienced.

Conventional optical character recognition (OCR) engines, which are optimized for clean, printed text, perform significantly worse when applied to handwritten bills. Handwriting styles vary broadly between individuals, characters are frequently joined, slanted, or incomplete, and the physical layout of a handwritten bill is rarely standardized: item columns may shift, totals may be written in margins, and ruling lines may be absent or inconsistent. Beyond recognition accuracy, conventional OCR pipelines typically output an unstructured stream of text rather than the structured, field-level representation (bill number, date, vendor, item list, quantities, prices, tax, and total) that is actually required for downstream accounting or inventory use. This gap between raw text recognition and usable structured data is a core obstacle to automating bill digitization for small businesses, and it motivates treating handwriting recognition and field-level architectures [4].

On the receipt-recognition side, Yu, Ma, and Kong integrate CLIP-based multimodal alignment with a lightweight of deep-learning-based key information extraction from information extraction as a single, unified problem rather than two independently solved tasks.

Recent research has addressed closely related problems from several directions. At the level of document-wide synthesis, Rombach and Fettke's systematic literature review business documents identifies the field's progression through preprocessing, detection, recognition, and layout-aware extraction stages, and highlights the handling of noisy, non-standardized, and handwritten business documents as a persistent open problem rather than a solved one [1]. On invoice-specific extraction, Yashwant et al. propose a set of field-level evaluation metrics, specifically field-level precision, consistency-check failure rate, and exact-match accuracy, for assessing how well an automated pipeline extracts structured fields such as invoice number, date, vendor, and total amount against a verified ground truth [2]. For handwritten documents specifically, Wei and Sun combine an improved Differentiable Binarization Network (DBNet) for text detection with an improved Convolutional Recurrent Neural Network (CRNN) for recognition, applied to handwritten tax receipts, and report that their detection-then-recognition design substantially outperforms earlier baselines on their own annotated dataset [3]. Lodh et al. propose a transformer-based handwriting recognizer, HATCharClassifier, that fuses offline pen-stroke images with online stroke trajectory information and reports state-of-the-art accuracy on established online-handwriting benchmarks, illustrating the continued migration of handwriting recognition research toward transformer and attention-based

Bidirectional Gated Recurrent Unit for edge-deployable receipt recognition, demonstrating that compact sequence models can retain competitive accuracy while remaining deployable in

resource-constrained environments, and further highlighting the versatility of CLIP in receipt-related tasks [5].

deployable on low-cost hardware [5]. Layout-aware document understanding has also advanced: Yan et al. propose DocExtractNet, a LayoutLMv3-based framework enhanced with an image-enhancement module for low-quality scans and a hint-recovery strategy for missing key-value pairs, achieving strong field-extraction performance across several receipt and form benchmarks [6], while Mifsud et al. combine a neural multimodal transformer with a rule-based correction model for receipt information extraction and report that the hybrid approach outperforms the transformer alone [7]. Text detection specific to receipts has also evolved: Ginting, Sitorus Pane, and Nababan combine the EAST text detector with a CRNN recognizer in a domain-specific OCR framework for shopping receipts, reporting improved accuracy over both a generic OCR engine and a standalone CRNN baseline [8]. Due to the use of ordinary smartphones to capture bills, image-preprocessing quality also becomes an area of active research: Chaves et al. report a large-scale evaluation of binarization algorithms for camera-captured, photographed documents under varying lighting, flash, and capture-angle conditions, noting that binarization behavior for photographed documents cannot be taken as given based on scanner-based results [9]. At the recognition-architecture level, Bagade et al. report a hybrid CNN-LSTM-CTC handwriting recognizer evaluated on the IAM Handwriting Database and applied to an automated answer-scoring pipeline, offering direct, recent evidence for the CNN-BiLSTM/LSTM-CTC design pattern supporting handwriting recognition systems of this kind [10]. Lastly, given the rise of large multimodal models as an alternative to classical OCR pipelines, Berghaus et al. benchmark vision-capable large language models against a structured OCR-to-markdown parsing strategy for invoice processing, finding that neither approach consistently dominates the other and performance depends on document layout complexity [11].

This body of 2025–2026 work demonstrates steady progress on individual stages of the problem (detection, recognition, layout-aware extraction, and preprocessing), but it also highlights a research gap that is relevant to small-business bill digitization: few of the reviewed systems jointly address all three characteristics of a small-business handwritten bill: free-form, camera-captured images rather than scanned forms; genuine handwritten, rather than printed or typed, content; and the need to provide a complete, validated, field-based structured record rather than isolated text transcriptions or single-field extractions. While existing layout-aware frameworks such as DocExtractNet and the joint transformer/rule-based model of Mifsud et al. are developed around printed or semi-structured receipts, the detection-and-recognition pipelines of Wei and Sun and of Ginting et al. are not coupled with an explicit field-validation and structured-storage stage appropriate to small-business record-keeping workflows such as Excel or lightweight databases.

Motivated by this gap, this paper proposes a Bill Recogni-tion and Structured Data Extraction System that includes camera-aware image preprocessing, deep-learning-based handwritten text recognition based on the CNN, BiLSTM, and optional Transformer design patterns described in the re-

viewed literature, layout-aware field identification, data validation, and structured storage within a single pipeline for small-business handwritten bills. The system is presented here as a proposal: a problem definition and review of the relevant

then recognition) justifies our method choice of first

2025–2026 literature are provided, along with a methodology, system architecture, and evaluation plan, but no implementation or reporting of experimental results is included.

II. LITERATURE REVIEW

This section discusses the research works directly related to the given one, including recent studies on OCR of handwritten text, bills/receipts OCR, preprocessing of document images, text detection and recognition algorithms (CNN, CRNN, BiLSTM, Transformer), document layout-aware recognition, and invoice and receipt information extraction.

Rombach and Fettke analyze 130 deep-learning-based approaches to key information extraction from business documents published in 2017-2024, grouping them according to a common task: document preprocessing, layout encoding, and field-specific information extraction [1]. The study does not evaluate the accuracy of a particular model but summarizes the conclusions of the papers under review. The main value of this research is that it highlights the existing trends in key information extraction from business documents, providing a taxonomy of tasks with a brief description of each. In addition, the authors identify gaps in the current research, specifically the lack of methods for processing noisy scans, documents with non-standard layout, and handwritten text. This work served as the starting point for our research because it illustrates, from a bird's eye view, the current level of development of key information extraction from business documents. Thus, our paper also addresses the challenge of extracting information from non-standard documents, including handwritten texts, which is not covered in the reviewed articles.

Yashwant et al. propose to tackle the problem of invoice information extraction using Docling- and LlamaCloud-based approaches and suggest using field-level precision, consistency-check failure rate, and exact-match accuracy to assess the quality of the results [2]. The paper does not consider handwritten invoices but evaluates methods in terms of accuracy on digital invoices. This study's value is in the suggested format of the comparison of different invoice information extraction methods, which we used as a basis for comparison in our experiments with handwritten bills and receipts. Apart from that, although the paper does not consider handwritten documents, the authors highlight the main categories used in invoice documents: invoice number, supplier name, customer name, date, due date, total amount due, goods/servcies description, itemized list of goods/services purchased, and VAT.

Wei and Sun propose a method for recognizing handwritten tax receipts, which combines an enhanced Differentiable Binarization Network for text detection and an optimized CRNN for text recognition [3]. They report that the performance of their approach on their own dataset achieves a detection precision of 89.1% and recall of 86.4% and recognition precision of 95.6%, recall of 94.2%, and F1-score of 0.94, outperforming several state-of-the-art methods in detection and multimodal tasks. The article directly addresses the problem of recognizing handwritten receipts, which is at the heart of our research. However, unlike our study, the authors only evaluate their method on tax receipts, which are probably more structured than average small-business receipts. Nevertheless, this paper's value is that it provides a baseline for our own experiment on small-business bill recognition. The authors' two-step approach (first detection,

separating text regions from the rest of the receipt image and then applying text recognition.

Lodh et al. propose HATCharClassifier, a transformer-based handwriting recognizer which performs early fusion of offline pen-stroke images and online stroke-trajectory data in a shared latent space [4]. Evaluated on the IAMOn-DB and VNOn-DB online-handwriting benchmarks, the model reports state-of-the-art character-level accuracy, up to 96.4% and 95.8% respectively. A limitation is that the approach assumes access to online stroke-trajectory data, which is unavailable for offline, camera-captured bill images, its relevance however is in demonstrating the value of attention-based, cross-modal fusion; this motivates considering an attention-augmented sequence model as an enhancement over a plain BiLSTM in the proposed recognition stage.

Yu, Ma and Kong propose a receipt-recognition framework which integrates CLIP-based multimodal alignment and a lightweight Bidirectional Gated Recurrent Unit and position-aware attention, for edge deployment [5]. On the CORD receipt dataset, they report a text-detection F1-score of 93.1% and a character error rate of 5.1%, with an 18 MB compressed model running at 25 frames per second on a Jetson AGX Orin device. The evaluation is limited to printed receipts rather than handwritten bills, but the demonstrated feasibility of lightweight, resource-constrained sequence models directly informs the deployment of the proposed system in a small-business setting with limited computing infrastructure.

Yan et al. introduce DocExtractNet, a LayoutLMv3-based framework for receipt information extraction, which features an image-enhancement module for low-quality scans and a hint-recovery strategy for incomplete key-value pairs [6]. On the Finance-Receipts, FUNSD and CORD benchmarks, the framework reportedly achieves F1-scores of 97.07%, 91.80% and 97.38%, respectively. The underlying LayoutLMv3 backbone is pre-trained mostly on printed and typed documents, and the paper does not report results on genuinely handwritten bills; its relevance is in demonstrating the value of explicit low-quality-image handling, which directly informs the proposed system's layout-aware field-identification stage.

Mifsud et al. study the joint use of a neural multimodal transformer (LayoutLMv3 with a classifier head, supported by the DocTR OCR engine) and a rule-based correction model for receipt information extraction, on a dataset which combines real and synthetic receipts [7]. They report an overall F1-score of 0.98 for classifying textual fields, including line items, after hyperparameter tuning and integration of the rule-based component, indicating that the hybrid combination outperforms the transformer used in isolation. The study is again focused on printed and semi-structured receipts rather than handwritten bills, but the finding that rule-based validation meaningfully improves transformer-based extraction directly informs the inclusion of an explicit data-validation stage in the proposed pipeline.

Ginting, Sitorus Pane and Nababan propose a domain-specific OCR framework for shopping receipts which combines the EAST text detector with a CRNN recognizer [8]. Trained and evaluated on 320 annotated training and 84 test receipt images, the framework reportedly achieves 92.6% character-level and 86.4% word-level accuracy, excelling both a generic Tesseract baseline and a standalone CRNN. The evaluation dataset is comparatively small and limited to printed receipts; its relevance is in confirming that a two-stage detect-then-recognize CRNN pipeline is a practical, reproducible baseline design even with a modestly-sized annotated dataset, which is a realistic constraint for a small-business bill dataset that remains to be collected.

Chaves et al. present a large-scale, non-handwriting-specific study which compares the quality, time and space cost of ninety-eight classical and five newly proposed binarization algorithms applied specifically to photographed, camera-captured documents under various lighting, flash and capture-angle conditions [9]. As a benchmarking and assessment study rather than a recognition system, it does not report end-to-end recognition accuracy; its direct relevance is methodological, in that it establishes that binarization behavior for camera-photographed documents differs materially from flatbed-scanned documents, directly motivating the noise-removal, contrast-enhancement and skew-correction sub-stages proposed alongside binarization in the present system's preprocessing stage.

Bagade et al. report a hybrid CNN-LSTM model with a Connectionist Temporal Classification (CTC) decoder for handwritten text recognition, which is trained and evaluated on the IAM Handwriting Database, and further apply the recognized text to an automated short-answer scoring pipeline using several similarity metrics [10]. They report that BERT-embedding-based similarity achieved the highest scoring accuracy (94.44%), among the similarity metrics compared, and observe a sign of overfitting in the handwriting recognizer after approximately forty training epochs on the IAM validation set. A limitation of this study is its evaluation on the general-purpose IAM dataset rather than on bill- or receipt-specific handwriting, and its downstream task (answer scoring) differs from structured field extraction; its relevance is architectural, as it offers recent, concrete evidence for the CNN+LSTM/BiLSTM+CTC design pattern which constitutes the handwritten text recognition stage of the proposed system.

Finally, Berghaus et al. benchmark eight multimodal large language models, ranging from the GPT-5, Gemini 2.5 and Gemma 3 families, against a structured OCR-to-markdown parsing strategy (using Docling) for invoice field extraction across several openly available invoice datasets [11]. They find that neither direct multimodal image processing nor the structured text-parsing strategy uniformly dominates the other, with relative performance depending on document layout complexity and the specific field being extracted. The study is restricted to typed and scanned invoices rather than handwritten bills, and does not address handwriting recognition; its relevance lies in illustrating that even state-of-the-art multimodal large language models do not yet offer a dominant, layout-agnostic solution for structured document field extraction, reinforcing the need for a dedicated, purpose-built pipeline such as the one proposed in this paper, rather than relying on general-purpose multimodal models alone.

TABLE I
COMPARATIVE SUMMARY OF REVIEWED 2025–2026

Ref. / Author, Year	Research Problem	Method / Model	Dataset/Document Type	Key Findings	Limitation
[1] Rombach & Fettke, 2025	Mapping the state of DL-based key information extraction	Systematic literature review (130 studies)	Business documents (survey)	Identifies handwritten/noisy documents as under-addressed	Synthesis only; no new system or accuracy figures
[2] Yashwant *et al.*,	Standardizing invoice	Docling / LlamaCloud	Digital/scanned	Proposes field-level precision,	Not evaluated on

2025	extraction evaluation	pipelines	invoices	consistency-check failure, and exact-match metrics	handwritten content
[3] Wei & Sun, 2026	Handwritten tax receipt recognition	Improved DBNet + improved CRNN	Handwritten tax receipts (own dataset)	Detection P = 89.1%, R = 86.4%; recognition P = 95.6%, R = 94.2%, F1 = 0.94	Dataset not independently validated on other bill formats
[4] Lodh *et al.*, 2025	Robust handwriting recognition via online + offline fusion	Transformer early fusion (HATCharClassifier)	IAMOn-DB, VNOn-DB (online handwriting)	Up to 96.4% / 95.8% character accuracy	Requires online stroke data, unavailable for offline images
[5] Yu *et al.*, 2025	Lightweight receipt recognition for edge devices	CLIP + BiGRU, position-aware attention	CORD receipt dataset	F1 = 93.1%, CER = 5.1%; 18 MB model at 25 FPS	Printed receipts only
[6] Yan *et al.*, 2025	Enhanced receipt information extraction	LayoutLMv3 + ImageEnhance + Precision Hints (DocExtractNet)	Finance-Receipts, FUNSD, CORD	F1 = 97.07%, 91.80%, 97.38%	Backbone pre-trained on printed documents
[7] Mifsud *et al.*, 2025	Improving receipt field classification	Joint LayoutLMv3 transformer + rule-based model	Real + synthetic receipts	Overall F1 = 0.98 after rule-based integration	Printed/semi-structured receipts
[8] Ginting *et al.*, 2026	Domain-specific receipt OCR	EAST detector + CRNN	320 train / 84 test shopping receipts	92.6% character, 86.4% word accuracy	Small dataset; printed receipts
[9] Chaves *et al.*, 2025	Binarization of camera-photographed documents	Benchmark of 98 classical + 5 new algorithms	Photographed laser/deskjet documents	Quantifies quality/time/space trade-offs under lighting/angle variation	No recognition accuracy reported
[10] Bagade *et al.*, 2026	Handwriting recognition for automated answer scoring	CNN-LSTM-CTC	IAM Handwriting Database	BERT-similarity scoring accuracy 94.44%; overfitting after ~40 epochs	General handwriting, not bill-specific
[11] Berghaus *et al.*, 2025	Comparing LLM strategies for invoice field extraction	Multimodal LLMs (GPT-5, Gemini 2.5, Gemma 3) vs. OCR-to-markdown parsing	Open invoice datasets	Neither strategy uniformly dominates; depends on layout complexity	Typed/scanned invoices; no handwriting

III. METHODOLOGY

The section describes the proposed Bill Recognition and Structured Data Extraction System. Consistent with the re-search gap identified in Section II, the system is presented as a proposed framework: its components, architecture, and train-ing strategy are specified in detail, but no implementation has yet been carried out, and no experimental results are reported.

A. System Overview and Pipeline

The proposed methodology follows a linear processing pipeline that transforms an input bill image into structured, stored data, as shown in Fig. 1:

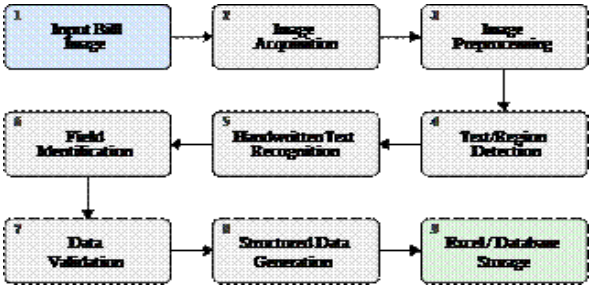

Fig. 1. Flow diagram of the proposed bill recognition and structured data extraction pipeline.

B. Bill Image Acquisition

The proposed system accepts bill images captured through a web or mobile interface using either an ordinary smartphones camera or a flatbed scanner. Since it is expected that the majority of small-

business bills will be captured by smartphone camera rather than scanner, the system is designed around the variable lighting, capture angle, and image-quality conditions documented for photographed documents by Chaves et al. [9], rather than assuming scanner-quality input.

C. Image Preprocessing

The image is then passed through a preprocessing sub-pipeline consisting of: (i) noise removal, to suppress speckle and background artefacts introduced by photography; (ii) binarization, to convert the image to a black-and-white representation, separating ink strokes from background using the choice of binarization algorithm informed by the comparative quality-time-space assessment of photographed-document binarization reported by Chaves et al.

[9]; (iii) image enhancement (contrast adjustment), to sharpen strokes and improve legibility under low-light capture, using the image-enhancement strategy used for low-quality receipt scans by Yan et al. [6]; (iv) skew correction, to rotate a tilted image so that text lines are horizontally aligned; and (v) resizing/normalization, to bring images to a consistent resolution for the downstream recognition model.

D. Text and Region Detection

The preprocessed image is analyzed to detect candidate text regions, corresponding to individual lines or fields, that serve as input units for recognition. The proposed design exploits the detection-then-recognition split evidenced for handwritten tax receipts via an improved Differentiable Binarization Network by Wei and Sun [3] and by the EAST-based detection strategy used for shopping receipts by Ginting et al. [8]. An alternative, trainable text-detection stage is proposed to tolerate the variable, non-standardized layouts typical of small-business bills.

A. Handwritten Text Recognition

Each detected region is proposed to be passed to a handwritten text recognition module combining a Convolutional Neural Network (CNN) visual front end for feature extraction with a Bidirectional Long Short-Term Memory (BiLSTM) sequence model trained under a Connectionist Temporal Classification (CTC) objective. This CNN+BiLSTM+CTC design follows the pattern directly evidenced for handwriting recognition by Bagade et al. [10] and for handwritten receipts by Wei and Sun [3]. A Transformer-based or cross-modal attention-augmented alternative to the BiLSTM stage, following the fusion strategy of Lodh et al. [4] and the multimodal alignment strategy of Yu et al. [5], is proposed as an optional enhancement to be

compared against the BiLSTM baseline during implementation, and not as an assumed superior strategy. As an additional point of comparison, the recent benchmarking of multimodal large language models against structured OCR-based parsing by Berghaus et al. [11] is noted as a possible future alternative recognition strategy, though it is not adopted as the primary design in this proposal because it has not yet been evaluated on handwritten content.

B. Layout-Aware Field Identification

Recognized text is mapped to specific bill fields (bill number, date, vendor name, customer name, item description, quantity, unit price, tax, and total amount) using a combination of positional cues (where text appears on the bill) and pattern- or learned classification. This stage follows the layout-aware extraction approach of DocExtractNet [6] and the joint transformer/rule-based correction strategy of Mifsud et al. [7], adapted from printed-receipt settings to the less standardized layouts expected of handwritten small-business bills.

C. Data Validation

Extracted field values are proposed to be checked for internal consistency or flagged for user review when inconsistencies are detected. This validation stage adapts, for a handwritten-bill context, the field-level precision and consistency-check-failure evaluation framework proposed by Yashwant et al. [2], and reflects the finding of Mifsud et al. that rule-based correction measurably improves transformer-based field extraction [7].

D. Structured Data Generation and Storage

Validated field values are assembled into a structured record (structured tabular or JSON form, one record per bill) and the proposed system is de-signed to export this record to an Excel spreadsheet or to insert it into a lightweight SQL database (MySQL or SQLite), allowing the extracted data to integrate with the

record-keeping workflows already used by small businesses.

E. Proposed System Architecture

At module level, the proposed system decomposes into co-operating modules: a user interface for bill upload or capture and result review; an image preprocessing module; a text/region detection module; a handwritten text recognition module (comprising CNN, BiLSTM, optional Trans-former components); a field identification module; a data vali dation module; and a data storage/export module. A user (shopkeeper or business operator) interacts with the system through a web or mobile interface; the image passes through the preprocessing and detection modules, then the prerequisite for implementation and is treated as future work, not as part of the methodology already executed.

A. Algorithm BEGIN

INPUT bill_image

IF quality(bill_image) is unacceptable THEN REQUEST re-capture
END IF
pre_img <- PREPROCESS(bill_image)
# denoise, binarize (Otsu), correct skew, enhance, regions <- DETECT_TEXT_REGIONS(pre_img)

recognition module, then the field identification and validation modules, before being written to Excel or a database and displayed back to the user.



Fig. 1 System Architecture

F. Training and Testing Approach, and Algorithms/Models

The proposed training strategy follows standard supervised deep-learning practice for document recognition: the detection module (DBNet- or EAST-style, following [3], [8]) and the recognition module (CNN+BiLSTM+CTC, optionally Trans-former-augmented, following [10], [4], [5]) are proposed to be trained separately or jointly on annotated bill images, with the field-identification module trained or fine-tuned with a lay-out-aware pre-trained backbone in the manner of DocExtract-Net [6]. Python, with TensorFlow for model development and OpenCV for classical image preprocessing, is proposed as the implementation environment, consistent with the tooling used across the reviewed recognition and detection literature.

G. Dataset Requirements, Preprocessing, and Annotation

Implementing and training the proposed system requires a dataset of small-business handwritten bills which does not, to the author's knowledge, currently exist in the reviewed literature; the closest available resources are the handwritten tax-receipt dataset of Wei and Sun [3] and general handwriting datasets such as IAM, used by Bagade et al. [10], neither of which match the layout and content of small-business bills. The proposed dataset would require: (i) bill images captured under realistic, varied smartphone-camera conditions (lighting, angle, device), following the capture-condition variation studied by Chaves et al. [9]; (ii) line- and field-level ground-truth annotation of the nine target fields listed above; and (iii) a held-out test partition reserved exclusively for the evaluation plan described in Section IV. Compiling and annotating this dataset is itself a necessary

FOR EACH region IN regions DO text[region] <- RECOGNIZE(region)
# CNN + BiLSTM / Transformer model END FOR fields <- IDENTIFY_FIELDS(text)
# bill no., date, vendor, items, price, tax, total IF VALIDATE(fields) is consistent THEN
record <- STRUCTURE(fields) STORE(record) TO Excel DISPLAY(record) TO user
ELSE FLAG(fields) FOR user review END IF END

IV. RESULTS AND EVALUATION PLAN

Experimental numerical results are not reported because the proposed system has not yet been experimentally implemented and evaluated. This section instead specifies the evaluation plan by which the proposed system is intended to be assessed once implemented, and is careful throughout to distinguish proposed evaluation procedures from completed measure-ments.

A. Dataset Split and Training/Validation/Testing Strategy

Once the bill-image dataset described in Section III-K has been collected and annotated, it is proposed to be partitioned into training, validation, and test subsets (for example, in a conventional 70/15/15 or 80/10/10 split, to be finalized once the full dataset size is known), with the test subset held out entirely from model development and used only for final evaluation. Cross-validation is proposed as an additional safeguard against overfitting, particularly given the comparatively small annotated datasets typical of this domain, such as the 320 training and 84 test images used by Ginting et al. [8], and the overfitting behavior after approximately 40 training epochs reported for a comparable CNN-LSTM handwriting recognizer by Bagade et al. [10].

B. Evaluation Metrics

The proposed evaluation plan adopts standard, well-established document-recognition and information-extraction metrics, following the evaluation practice used for handwritten receipts by Wei and Sun [3] and for invoice extraction by Yashwant et al. [2], rather than reporting ad hoc or unverified figures:

- Character Error Rate (CER) and Word Error Rate (WER), computed against manually verified ground-truth transcriptions using the standard edit-distance formulation

$$CER = (S + D + I) / N, \quad (1)$$

where S, D, and I are the numbers of substitution, deletion, and insertion operations required to transform the recognized text into the ground truth, and N is the total number of ground-truth characters (WER is defined analogously at the word level).

- Field-level accuracy, measured as the proportion of extracted fields (bill number, date, vendor, item description, quantity, unit price, tax, total amount) that exactly match verified ground-truth values.
- Precision, recall, and F1-score for the field-identification and data-validation stages, treating each correctly identified and correctly valued field as a true positive, following the field-level and consistency-check framing proposed by Yashwant et al. [2].
- Exact-match accuracy, measuring the proportion of bills for which every target field is extracted correctly in full.
- Bill-level extraction accuracy, summarizing end-to-end performance across an entire bill record rather than a single field.
- Processing time per bill, measured from image input to structured-record output, to assess practical feasibility for small-business use, in the spirit of the frame-rate and model-size reporting used for light- weight receipt recognition by Yu et al. [5].

C. Comparison with Baseline Methods

The proposed system is intended to be compared against at least three categories of baseline once implemented: (i) a conventional, non-handwriting-specialized OCR engine applied directly to the same bill images, to quantify the benefit of the proposed handwriting-specific pipeline over general OCR; (ii) a detection-then-recognition baseline resembling the EAST+CRNN design of Ginting et al. [8] or the DBNet+CRNN design of Wei and Sun [3], without the proposed BiLSTM/Transformer recognition enhancement or the layout-aware field-identification stage, to isolate the contribution of each added component; and (iii), where feasible, a general-purpose multinodal large language model approach resembling the comparison carried out by Berghaus et al. [11], to assess whether a purpose-built pipeline offers measurable advantages over a general-purpose model on handwritten, small-business bills specifically.

D. Planned Evaluation of Individual Stages

Handwritten text recognition is proposed to be evaluated using CER and WER on held-out bill-image lines; text/region detection is proposed to be evaluated using standard detection precision and recall against annotated text-region bounding boxes, following the detection-metric convention used by Wei and Sun [3]; field extraction is proposed to be evaluated using field-level precision, recall, F1-score, and exact-match accuracy; and overall structured-data extraction is proposed to be evaluated using bill-level extraction accuracy and end-to-end processing time, considered jointly so that recognition accuracy is not assessed in isolation from practical throughput.

No CER, WER, precision, recall, F1-score, exact-match accuracy, bill-level accuracy, or processing-time values are reported in this paper, since the system described in Section III has not yet been built, trained, or tested. The evaluation plan above is intended to guide, and to be executed during, a future

IV. CONCLUSION

This paper tackles the problem of digitizing handwritten bills from small retailers, medical stores, and other small-scale businesses where data entry is laborious due to the lack of capabilities of conventional OCR systems to recognize handwritten and non-standardized documents. Drawing insights from recent works on recognition of handwritten text, receipt and invoice OCR, document image preprocessing, and layout-aware information extraction, this paper proposes a Bill Recognition and Structured Data Extraction System that implements camera-aware image preprocessing, CNN+BiLSTM-based handwritten text recognition module (with an optional enhancement using a Transformer-based architecture), layout-aware field identification, explicit data validation, and structured data storage in Excel or a lightweight database. By combining deep learning and document image processing, the proposed solution aims to address the challenges of handwriting variations, smartphone-based image capture, bill layout irregularities, and field-level data extraction that are not fully addressed by any of the aforementioned techniques. The current paper's main contribution is defined by the proposed system's pipeline and evaluation methodology that is specifically focused on recognizing handwritten bills from small businesses as opposed to being centered on developing an entirely new recognition architecture. It is expected that such a system would reduce the amount of manual data entry needed to digitize a transaction bill while also enhancing data auditability by implementing recognition-specific data validation rules and facilitating structured data storage. However, as the current work only defines the system

pipeline and does not evaluate its accuracy or data entry reduction efficiency, no specific claims can be made regarding the advantages of the proposed approach over existing OCR, layout-aware field extraction, or multimodal large language model solutions. Moreover, the challenges inherent to handwriting variations, image quality degradations due to capturing mechanisms, and layout irregularities specific to small-scale business bills should be considered significant limitations of this study as no practical experiments were performed to assess the performance of the proposed approach. As such, future work will be focused on gathering a suitable dataset of handwritten bills, implementing the proposed system pipeline, and performing the experiments and evaluation described in this paper.

REFERENCES

[1] A. M. Rombach and P. Fettke, "Deep learning based key information extraction from business documents: Systematic literature review," *ACM Comput. Surveys*, vol. 58, no. 2, Art. 45, pp. 1–37, Sep. 2025.

[2] S. Yashwant, A. Dubey, P. Paikray, and G. Thulsiram, "Invoice information extraction: Methods and performance evaluation," *arXiv preprint* arXiv:2510.15727 [cs.AI], Oct. 2025.

[3] L. Wei and J. Sun, "Handwritten tax receipt recognition method based on improved DBNet and CRNN," *PLOS ONE*, vol. 21, no. 7, pp.

[7] X. Mifsud, L. Grech, A. Baldacchino, L. Keller, G. Valentino, and A. Muscat, "Receipt information extraction with joint multi-modal transformer and rule-based model," *Mach. Learn. Knowl. Extr.*, vol. 7, no. 4, Art. 167, Dec. 2025.

[8] D. E. P. Ginting, S. A. Sitorus Pane, and M. N. K. Nababan, "Deep learning-based OCR framework for receipts: Performance evaluation of EAST and CRNN integration," *JITK (Jurnal Ilmu Pengetahuan*

1–18, 2026.

[4] A. Lodh, R. Chakraborty, S. Palaiahnakote, and U. Pal, "A transformer based handwriting recognition system jointly using online and offline features," *arXiv preprint* arXiv:2506.20255 [cs.CV], Jun. 2025.

[5] J.-M. Yu, H.-J. Ma, and J.-L. Kong, "Receipt recognition technology driven by multimodal alignment and lightweight sequence modeling," *Electronics*, vol. 14, no. 9, Art. 1717, 2025.

[6] Z. Yan, Z. Ye, J. Ge, J. Qin, J. Liu, Y. Cheng, and C. Gurrin, "DocExtractNet: A novel framework for enhanced information extraction from business documents," *Inf. Process. Manage.*, vol. 62, no. 3, Art. 104046, May 2025.

dan Teknologi Komputer), vol. 11, no. 3, 2026.

[9] G. P. Chaves, T. Vieira, G. de F. e Silva, R. D. Lins, and S. J. Simske, "Binarizing photographed document images: 2025 quality, time and space assessment," in *Proc. ACM Symp. Document Eng. (DocEng '25)*, Nottingham, U.K., Sep. 2025.

[10] J. Bagade, S. Rajebhosale, M. Dhage, C. Hedaoo, K. Sultanpure, and

S. Godbole, "A hybrid CNN-LSTM model for handwritten text recognition with comparative analysis of similarity algorithms for automated answer scoring," *Ingénierie des Systèmes d'Information*, vol. 31, no. 3, pp. 965–975, 2026.

[11] D. Berghaus, A. Berger, L. Hillebrand, K. Cvejoski, and R. Sifa, "Multi-modal vision vs. text-based parsing: Benchmarking LLM strategies for invoice processing," in *Proc. 2025 IEEE Int. Conf. Big Data (BigData)*, 2025, pp. 2519–2526.