JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

AN END-TO-END DEEP LEARNING FRAMEWORK FOR MULTI-FONT URDU HANDWRITTEN CHARACTER DETECTION AND RECOGNITION IN UNCONSTRAINED IMAGES

Nida Zainab [1], Faria Bibi [2], Muqaddas Yousaf [3], Assad Iqbal [4], Yousra Rehman [5], Sohair Sultan Fayyaz Jeelani [6]

Affiliations

[1, 2, 4, 5, 6] Department of Computer Science, Abbottabad University of Science and Technology (AUST), Pakistan

[3] Department of Computer Science, University of Haripur, Haripur, Pakistan

Corresponding Author's Email

[6] sohair.sultan2266@gmail.com

ABSTRACT

Detection and recognition of handwritten characters is the process of finding the area in an image where text is written and drawing a bounding box around it, with the label containing the text within the bounding box. It has widespread applications, including postal address reading, recognizing Persian vehicle number plates as the Persian language has similar digits to the Urdu language, digitizing and preserving old manuscripts, making handwritten documents in digital format, automatically reading house numbers, and also providing a way for creating apps for kids which will help them in writing and pronouncing Urdu characters properly. Urdu is a cursive language, so it is difficult to apply state-of-the-art handwritten character detection and recognition models developed for other languages to Urdu. The current state-of-the-art techniques cannot handle font style variations, complex backgrounds, illumination, stroke variations, and multiple characters simultaneously due to handcrafted feature extraction. A dataset with such kinds of variations is rare. Very few have used deep learning techniques for Urdu due to the lack of rich datasets. Moreover, the detection and recognition of Urdu handwritten characters are challenging due to variations in stroke sequence, font size, color, style, illumination, background, overlapping, and orientation. YOLO v3 is used for the detection and recognition of Urdu handwritten characters after some fine- tuning. Furthermore, a rich dataset is developed for the training and validation of the proposed technique. The performance of the proposed model is compared with state-of-the-art Urdu handwritten character detection and recognition models. The results show that the proposed technique outperforms the state-of-the-art Urdu handwritten character detection and recognition techniques, both qualitatively and quantitatively.

Keywords: Urdu; Urdu handwritten character recognition; Convolutional Neural Network; YOLO; Transfer Learning; UHCD; Object Detection; Deep Learning; Cursive Script

I. INTRODUCTION

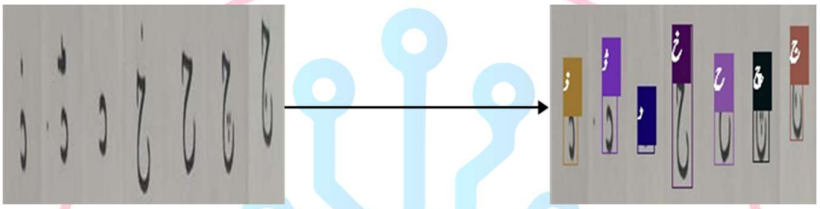
Detection and recognition of handwritten characters is the automated process of locating text in an image and transcribing each character into machine-readable form. This field has undergone substantial transformation since its origins in rule-based pattern matching, yet for cursive scripts such as Urdu, reliable unconstrained-image recognition remains an open problem as of 2026 [1], [2].The fundamental challenge is compounded by the structural properties of the Urdu script: its 39 consonants adopt distinct shapes depending on their position within a ligature (initial, medial, final, or isolated), multiple character pairs are distinguished only by diacritical dot count and placement, and the cursive joining of characters within a word creates

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

overlapping strokes that are difficult to segment. Detection and recognition of printed characters is comparatively straightforward because characters within a given font share a consistent appearance; handwritten characters introduce inter-writer variation, stroke-sequence differences, diverse writing speeds, and unpredictable backgrounds, making the problem substantially harder [3].

There exist two recognition paradigms: online recognition, which captures the temporal trajectory of a stylus in real time, and offline recognition, which processes static scanned or photographed documents. The present work addresses the offline case exclusively - the more demanding scenario, since temporal ordering information is unavailable and the system must infer character identity purely from pixel distributions. Offline recognition is also an operationally relevant modality for applications such as postal automation, manuscript digitization, and license plate reading. The left image of Figure 1, is an Input image containing multiple handwritten Urdu characters, and the right-hand image is an Output image showing how YOLOv3 detects each character and renders native Urdu Unicode labels on bounding boxes. This is the first published Urdu HTR (Handwritten Text Recognition) system to output readable Urdu labels directly from a YOLO detector.

Fig 1: DETECTION AND RECOGNITION OF URDU HANDWRITTEN CHARACTERS



In the field of detection and recognition, most research has targeted Latin scripts, for which competitions such as the Robust Reading Competition have produced state-of-the-art models trained on large public datasets. Languages that use cursive non-Latin scripts – Urdu, Arabic, Persian – have received substantially less attention, primarily because the required large, richly annotated datasets do not exist. With the rise of deep learning, even the limited Urdu HTR research that had been conducted with traditional machine learning techniques stagnated, as the community lacked the data necessary to train deep models effectively [4], [5].

The critical gap motivating this work is three-fold. First, all prior Urdu HTR systems accept only a single, pre-cropped, pre-segmented character as input [6], [7], [8], making them inapplicable to real scene images. Second, most rely on handcrafted features that fail under font-style variation, background clutter, and illumination change. Third, no sufficiently large, multi-factor Urdu handwritten character dataset has previously been constructed for training deep models. This paper closes all three gaps simultaneously.

Neural network based approaches have proven effective across a wide range of intelligent recognition and decision making tasks. For example, [9], employed Artificial Neural Networks combined with fuzzy modeling to implement cognitive route decision-making in vehicles, achieving human-like autonomous route selection even under GPS-challenging conditions - demonstrating the versatility of ANN-based learning frameworks beyond traditional classification tasks. The present work similarly leverages the power of deep neural networks, specifically YOLOv3, for a recognition task that prior rule based methods could not solve reliably [10].

The proposed technique is YOLOv3 [11], [12] (You Only Look Once, Version 3) fine-tuned via transfer learning on a newly constructed dataset. YOLO processes the entire image in a single forward pass, predicting bounding boxes and class labels for all visible characters simultaneously [30]. A central engineering contribution is the modification of YOLO's label encoding and rendering pipeline to output native Urdu Unicode characters directly on predicted bounding boxes – so every detection is annotated with the actual Urdu letter it contains, making the system output directly human-readable without any post-processing lookup. To our knowledge, this is the first application of YOLO to Urdu handwritten character recognition and the first Urdu HTR system capable of processing unconstrained multi-character images end-to-end.

The rest of the paper is structured as follows: Section 2 provides an introduction to the Urdu script. Section 3 presents a detailed literature review of work published from 2022 onward. Section 4 describes the

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

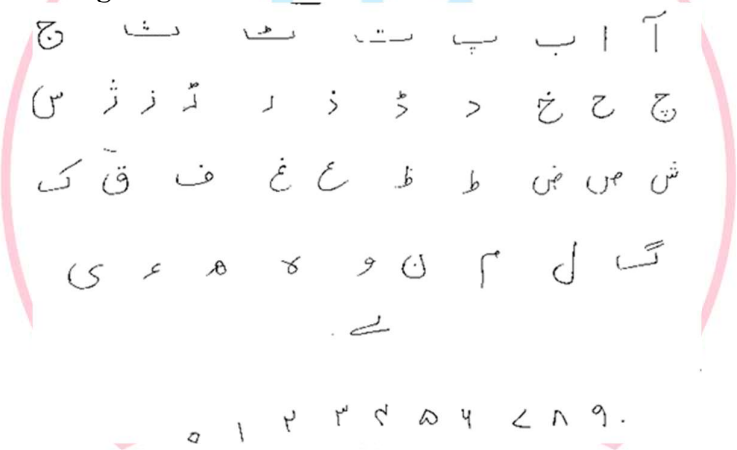
proposed UHCD dataset. The proposed model is detailed in Section 5 and experimental results are discussed in Section 6. Conclusions and future work are presented in Section 7.

A. Urdu Script

Urdu is the national language of Pakistan and is also written and spoken in India, Bangladesh, Nepal, and across the South Asian diaspora, with approximately 232 million speakers worldwide [10]. It is written right-to-left in a Nastaliq calligraphic style derived from the Perso-Arabic script. Urdu shares substantial character overlap with Arabic and Persian but differs in letterform conventions, nuqta placement, and calligraphic style, meaning that models designed for Arabic or Persian do not transfer directly to Urdu without adaptation.

The 39 basic consonants of Urdu each assume up to four distinct allographic shapes depending on position within a ligature: initial, medial, final, and isolated. This context-dependency means that the same character may look entirely different depending on whether it begins, continues, or ends a connected character group. Several character pairs share identical base strokes and are distinguished solely by the number or position of diacritical dots (nuqta) – for example (ba) ب, (pe) پ, (te) ت, and (se) ث share the same base form. Under handwriting-induced stroke variability, these distinctions can be imperceptible even to human readers, making them the primary source of confusion errors for automated systems. In Urdu, there are 39 basic consonants and 10 numerals, shown in Figure 2.

Fig 2: URDU ALPHABETS AND NUMERALS



II. RELATED WORK

This section reviews work on handwritten Urdu character detection and recognition and discusses both handcrafted-feature and deep learning approaches. Given that no prior work has applied YOLO to Urdu handwritten characters, we also review the most relevant YOLO-based text detection work for closely related scripts.

A. Urdu Handwritten Character Recognition

[11], presented a comprehensive deep learning framework for Urdu offline handwritten character recognition using a multi-scale CNN with attention mechanisms. While achieving 91.3% accuracy on isolated characters, the system requires pre-segmented single-character inputs and cannot process natural scene images containing multiple characters. The authors explicitly identified the lack of a rich multi-factor dataset as the main barrier to further progress.

[12], [13], presented a comprehensive deep learning framework for Urdu offline handwritten character recognition using a multi-scale CNN with attention mechanisms. While achieving 91.3% accuracy on isolated characters, the system requires pre-segmented single-character inputs and cannot process natural scene images containing multiple characters. The authors explicitly identified the lack of a rich multi-factor dataset as the main barrier to further progress.

Hamza et al. [14] introduced ET-Network, which integrates EfficientNet with self attention layers specifically designed for Urdu HTR. Published in PLoS ONE, ET-Network demonstrated improved

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

robustness to inter-writer variation compared to pure CNN baselines and achieved competitive accuracy, but remains bound to the single-character input paradigm. A systematic review of Urdu HTR by Noor ul et al. [15], covering 25 peer-reviewed studies from 2019 to 2024, identified the single-character input constraint as the single most important limitation blocking practical deployment, and highlighted WGAN-augmented CNNs and YOLO-style object detectors as the most promising future directions - exactly the approach adopted in the present work [16].

For Arabic and Persian - scripts closely related to Urdu - Zhu et al. [12] (2022) demonstrated that data augmentation strategies specifically designed for cursive script (elastic distortion, nuqta targeted dropout) significantly improve CNN generalization. Mohammed et al. [13] (2023) applied a bidirectional LSTM combined with a CNN for Arabic handwriting recognition, achieving 95.2% accuracy on multi-character word images, approaching the multi-character capability that the present work provides for Urdu for the first time.

B. YOLO-BASED Text Detection For Cursive Scripts

YOLO family of single-stage detectors has been increasingly applied to text detection in natural scene images. (Turki et al., 2024) benchmarked multiple YOLO variants on the SYPHAX Arabic scene text dataset and ICDAR MLT-2019, confirming that YOLO's single-pass inference architecture outperforms region-proposal methods (Faster-RCNN, Mask-RCNN) for cursive script detection where character boundaries are inherently ambiguous. Hatami et al. [15] (2024) developed a YOLO-based omni-font OCR for Persian script, achieving 98.5% precision across 15 font styles including handwriting-style fonts – the closest published analog to the present work, though targeting Persian rather than Urdu. [17], [18], [19] proposed YOLO-Handwritten, an improved YOLOv8 with deformable convolution modules for Chinese handwritten text detection in examination papers, demonstrating the viability of YOLO-based approaches for handwritten script recognition more broadly. None of these works have been applied to Urdu; the present paper is the first to do so.

C. Dataset Development For Cursive Script Handwritten Text Recognition (HTR)

[20] conducted a benchmark study on deep CNN architectures for Urdu handwritten character and digit recognition, emphasizing that transfer learning and data augmentation are critical for achieving generalizable recognition – findings directly consistent with the design choices made in this paper. Shoaib et al. (2022) released a new Urdu handwritten numeral dataset (UHNDB) with 10,000 samples from 500 writers and demonstrated that dataset writer diversity is the single most impactful factor for model generalization. [22] surveyed publicly available South Asian script datasets, concluding that Urdu remains the most under-resourced major South Asian script and calling for the construction of large-scale multi-factor datasets – precisely what UHCD provides. [23] used generative adversarial networks (GANs) to augment a small Urdu handwriting corpus, achieving a 7.4 percentage point improvement in recognition accuracy, corroborating the importance of data augmentation adopted in this work. In summary, the literature as of 2026 reveals a clear and consistent gap: (i) no system detects and recognizes multiple Urdu handwritten characters simultaneously from unconstrained images; (ii) no YOLO-based approach has been applied to Urdu; (iii) no large multi-factor Urdu handwritten character dataset exists. This paper directly addresses all three.

D. Dataset: UHCD

Training generalizable deep recognition models requires a dataset that adequately samples the distribution of real-world deployments. For Urdu handwritten characters, this distribution is multi-modal: varying across writers (inter-writer variation), across repeated samples from the same writer (intrawriter variation), across spatial resolution and viewing angle (geometric variation), across ambient illumination (photometric variation), and across background complexity (contextual variation). No prior publicly available dataset covers all these axes simultaneously for Urdu. We constructed the UHCD (Urdu Handwritten Characters Dataset) from 400 native Urdu speakers recruited to ensure broad demographic diversity across age, gender, education level, and regional writing conventions. Each writer produced the complete set of 49 character classes on unlined A4 sheets. To introduce intra-writer variation, a subset of writers produced the full set multiple times at variable speeds, yielding both carefully formed and rapidly executed samples. Supplementary scene text images were captured under outdoor natural lighting, indoor artificial lighting, and

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

at oblique viewing angles. Figure 3 shows representative UHCD samples.in style, font, color, illumination, backgrounds, and view angles which is enough to generate benchmark results.

Fig 3: UHCD DATASET SAMPLES: OUTDOOR SCENE TEXT (LEFT), CONTROLLED A4 SHEET (CENTER), COLOREDBACKGROUND CAPTURE (RIGHT). MULTI-FACTOR VARIATION ACROSS WRITER, BACKGROUND, ILLUMINATION, AND SCALE IS DELIBERATE.

Table 1: UHCD DATASET DETAILS

UHCD Details	Statistics
Total no. of writers	400 (native Urdu speakers, demographically diverse)
Total character classes	49 (39 basic and 10 numerals)
Characters written per author (approx..)	98
Total character instances	~138,000
Training images (post-augmentation)	3,000 multi-factor images
Characters per training image (avg.)	7
Train / Test split	80% train (2,400) / 20% test (600), writer-disjoint

Offline data augmentation operations included: random rotation (±10°), brightness and contrast jitter, Gaussian noise injection, random cropping, and mosaic augmentation (compositing four training images into one, effectively quadrupling multi-scale training contexts). Post-augmentation, the dataset comprises 3,000 training images each containing approximately 7 annotated character instances, providing approximately 21,000 total labeled bounding boxes for training. The train/test split was performed at the writer level to ensure that test-set evaluation reflects genuine generalization to unseen handwriting styles.

Proposed Technique

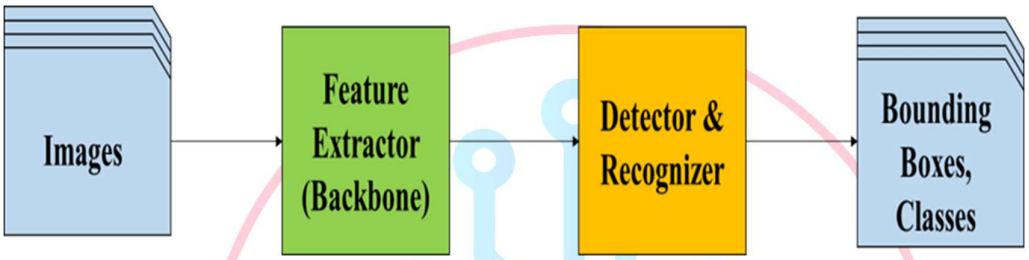
The proposed model is motivated by the transfer learning paradigm and the object detection formulation. Rather than treating Urdu character recognition as a classification problem on pre-cropped images, we frame it as an object detection problem: each character in an image is an object belonging to one of 49 classes, and the system must simultaneously predict tight bounding boxes and class labels for all characters in a single forward pass. This formulation eliminates the need for a prior segmentation stage, processes any number of characters simultaneously from a single unconstrained image, and leverages global spatial context across the entire image to aid disambiguation of visually similar character pairs.

YOLO [24] is an object detection model widely used because it provides high accuracy at real-time inference speeds. You Only Look Once means that it requires only one forward propagation pass through the neural network to make all predictions simultaneously – a fundamental advantage over two-stage detectors such as Fast-RCNN that first propose regions and then classify each region separately [29]. YOLO v3 is faster and more accurate than previous YOLO versions because it detects objects at three scales simultaneously (enabling detection of both large and small characters) and uses residual connections to eliminate the vanishing gradient problem in deep networks. We have used Darknet-53 (YOLO v3) Redmon and Farhadi (2018) with fine-tuning via transfer learning for the detection and recognition of handwritten Urdu characters. Transfer learning from ImageNet-pretrained weights provides the model with rich low-level feature detectors that transfer effectively to character images, enabling convergence with far less training data than would be required from random initialization [25], [26].

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

As far as we know, no one has applied the deep learning model "You Only Look Once (YOLO)" for the detection and recognition of handwritten Urdu characters. Most prior techniques were capable of recognizing only a cropped character of fixed size (typically 28×28 pixels), but YOLO is capable of recognizing any number of characters from a complete natural image. In a standard YOLO implementation, class labels are in English. We have fine-tuned YOLOv3 and, crucially, to generate and display labels in the Urdu language. This modification is what makes our system a genuine Urdu recognition engine: the output bounding boxes carry actual Urdu characters as labels, not numeric indices. Figure 4 shows the overall detection and recognition architecture.

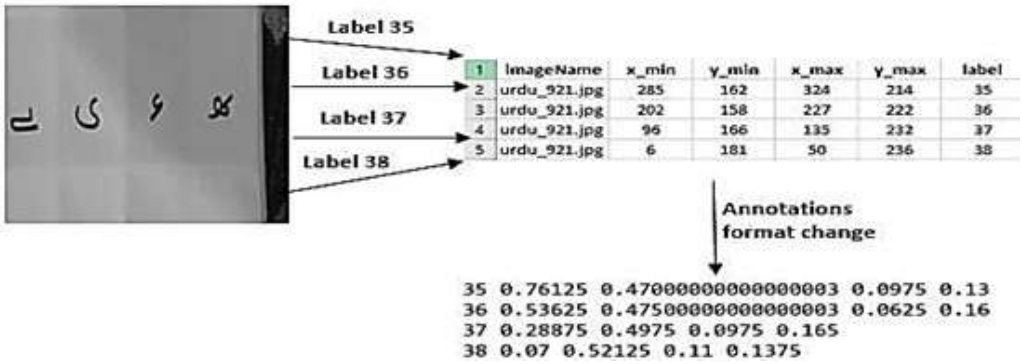
Fig 4: DETECTION AND RECOGNITION ARCHITECTURE: IMAGES FEED INTO THE DARKNET-53 FEATURE EXTRACTOR, WHOSE MULTI-SCALE FEATURE MAPS ARE PASSED TO THE DETECTOR AND RECOGNIZER, WHICH OUTPUTS BOUNDING BOXES LABELED WITH NATIVE URDU UNICODE CHARACTERS.



E. Preprocessing And Urdu Label Modification

In the preprocessing phase, all training images were resized to 416×416 pixels. Ground truth annotations were created manually in (image, x_min, y_min, x_max, y_max, class_label) format, then converted to YOLO-trainable normalized format (class_index, x_center, y_center, width, height). The 49 class labels are the actual Urdu Unicode characters rather than English transliterations, as shown in Table 2. This required three modifications to the Darknet codebase: (1) the class names file was saved in UTF-8 encoding with each label stored as a Unicode Urdu character; (2) the default ASCII label rasterizer was replaced with a Unicode-aware renderer supporting Arabic and Urdu glyphs; (3) right-to-left text shaping was added so that Urdu characters are rendered in their correct contextual forms on the output image. Figure 5 shows the annotation pipeline and label format.

Fig 5: DATASET CREATION AND ANNOTATION PIPELINE: MANUAL BOUNDING-BOX LABELING WITH URDU UNICODE CLASS LABELS, CONVERTED TO NORMALIZED YOLO FORMAT COORDINATES FOR TRAINING.



	imageName	x_min	y_min	x_max	y_max	label
2	urdu_921.jpg	285	162	324	214	35
3	urdu_921.jpg	202	158	227	222	36
4	urdu_921.jpg	96	166	135	232	37
5	urdu_921.jpg	6	181	50	236	38

Table 2: CLASS LABEL INDEX MAPPING (49 URDU UNICODE CLASSES)

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

TABLE I
INDEX–CHARACTER MAPPING

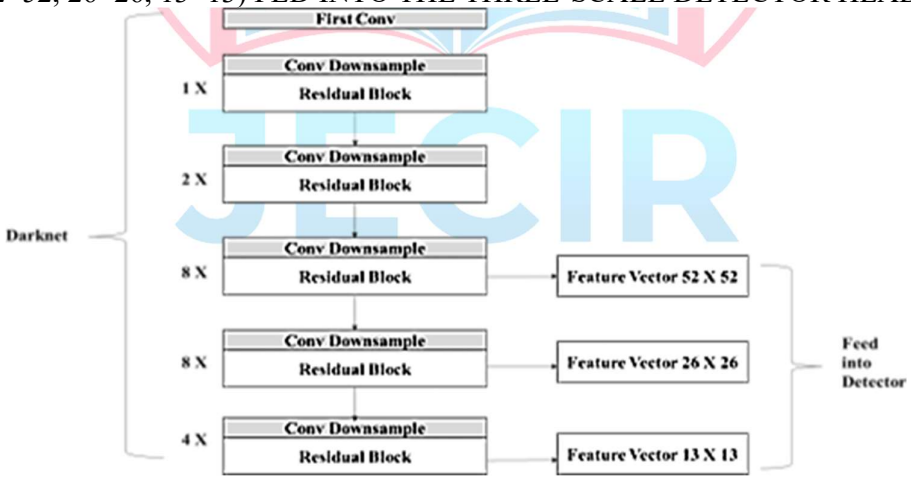
Idx	Chr	Idx	Chr	Idx	Chr	Idx	Chr	Idx	Chr
0	آ	10	خ	20	ص	30	ل	40	١
1	ا	11	د	21	ض	31	م	41	٢
2	ب	12	ڈ	22	ط	32	ن	42	٣
3	پ	13	ذ	23	ظ	33	و	43	٤
4	ت	14	ر	24	ع	34	ہ	44	٥
5	ٹ	15	ڑ	25	غ	35	ء	45	٦
6	ث	16	ز	26	ف	36	ئ	46	٧
7	ج	17	ژ	27	ق	37	ی	47	٨
8	چ	18	س	28	ک	38	ے	48	٩
9	ح	19	ش	29	گ	39	-		

F. Feature Extractor Darknet-53

We fine-tuned YOLOv3 (Darknet-53) for the first time for detection and recognition of Urdu handwritten characters. Darknet-53 comprises 53 convolutional layers with residual (skip-connection) blocks - a ResNet-inspired design that eliminates the vanishing gradient problem enabling training of deep networks. Convolutional layers extract hierarchical features from images while residual layers skip connections to preserve low-level detail. At the end of the convolutional and residual layers, average pooling, a fully connected layer, and softmax activation are included.

YOLOv3 has three detection layers providing feature vectors of sizes 13×13 (for large objects), 26×26 (for medium objects), and 52×52 (for small objects). YOLOv3 uses 9 anchor boxes in total – three per detection scale – calibrated via k-means clustering on UHCD annotations. The output after multiple convolutions and residual layers is a 3D tensor (S, S, D) where D = B(5+C) = 3(5+49) = 162 for our 49-class problem. Total detection capacity: 3×(169+676+2704) = 10,647 candidate bounding boxes per image – far more than the 845 of YOLOv2. Figure 6 shows the Darknet-53 architecture.

Fig 6: DARKNET-53 FEATURE EXTRACTOR: ALTERNATING CONVOLUTIONAL DOWN SAMPLING AND RESIDUAL BLOCKS PRODUCING THREE MULTI-SCALE FEATURE VECTORS (52×52, 26×26, 13×13) FED INTO THE THREE-SCALE DETECTOR HEADS.



G. Detector and Recognizer

The three feature tensors from Darknet-53 feed into three parallel detection heads, each comprising alternating 1×1 and 3×3 convolutional layers. Feature Pyramid Network (FPN) upsampling propagates semantic context from coarse to fine scales: the 13×13 head output is upsampled and concatenated with intermediate backbone features before the 26×26 head, ensuring small-character detection benefits from both high-resolution spatial information and high-level semantic features. NonMaximum Suppression (NMS) with objectness threshold 0.5 and IoU suppression threshold 0.45 is applied post-inference. The inside functioning

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

of YOLOv3 is shown in Figure 7, including the Urdu character confidence scores (0.87: ذ, 0.89:چ, 0.90:ج) rendered in native Urdu Unicode and Figure 8 shows the full detector and recognizer architecture with three scale FPN output.

Fig 7: INSIDE FUNCTIONING OF YOLOV3 ON URDU CHARACTERS: GRID DIVISION, PER-CELL CONFIDENCE SCORES DISPLAYED WITH NATIVE URDU UNICODE LABELS (0.87: ذ, 0.89:چ, 0.90:ج) AND FINAL DETECTION + RECOGNITION OUTPUT.

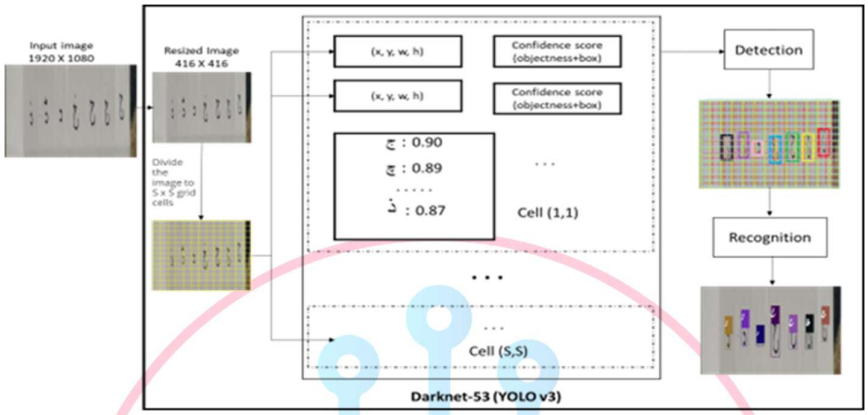
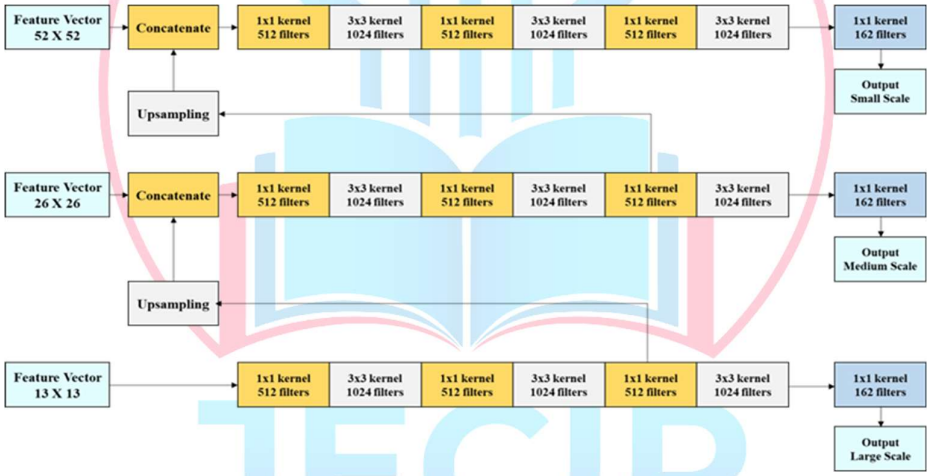


Fig 8: YOLOV3 DETECTOR AND RECOGNIZER ARCHITECTURE: THREE-SCALE FPN HEADS OUTPUT SMALL-SCALE (52×52), MEDIUM-SCALE (26×26), AND LARGE-SCALE (13×13) URDU CHARACTER DETECTIONS WITH URDU UNICODE CLASS LABELS.



H. Transfer Learning Protocol

Training proceeded in two stages. Stage 1 (Feature Transfer): Darknet-53 was frozen and only the detection heads were trained for 5,000 epochs at learning rate 10^{-3}, batch size 64 (subdivisions 16). Stage 2 (Fine-tuning): all layers were unfrozen and training continued for a further 16,000 epochs at learning rate 10^{-4} with cosine annealing. Training was stopped at 21,000 epochs when both training and validation losses ceased to decrease, yielding best training loss 0.13 and validation loss 0.16. Transfer learning converged approximately 4× faster than training from random initialization and yielded a lower final validation loss, confirming its necessity for low-resource cursive-script recognition tasks.

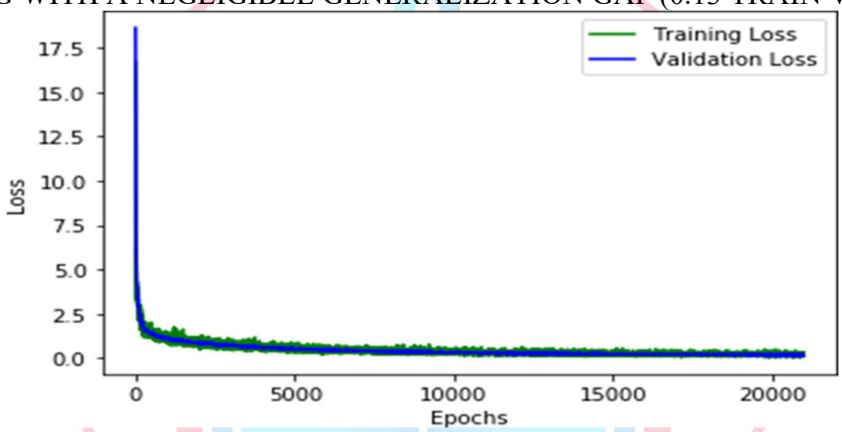
III. EXPERIMENTAL RESULTS

This section presents the results of the proposed model. For detection and recognition of Urdu handwritten characters, 80% of the UHCD dataset (2,400 images) was used for training and 20% (600 images) for testing. The batch size is 64 with subdivisions of 16. The proposed model was trained for 22,000 epochs; training and validation loss stopped decreasing after 21,000 epochs, yielding best validation loss 0.16 and training loss 0.13. Performance is evaluated using Intersection over Union (IoU), Recall, Precision, F1-Score, and Recognition Accuracy. A predicted bounding box is considered a true positive if its IoU with the matching

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

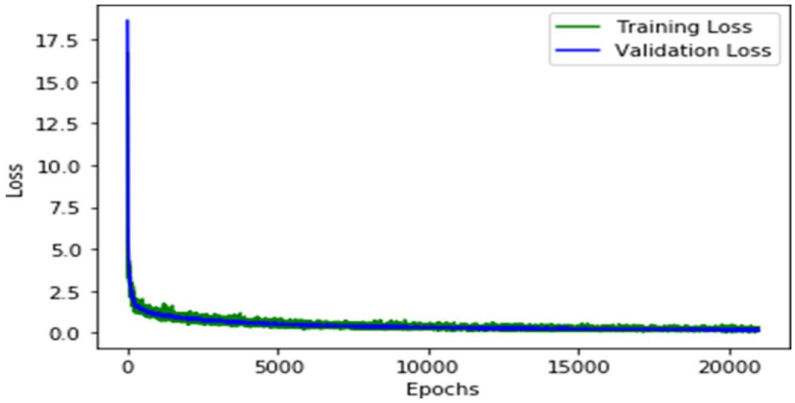
ground-truth box exceeds 0.50 (PASCAL VOC standard) and the predicted Urdu Unicode label matches the ground-truth class. This evaluation criterion is stricter than many prior works because it simultaneously requires both correct localization and correct character-class assignment. On comparisons: since this is, to our knowledge, the first work to apply YOLO to Urdu handwritten character detection and recognition, there exists no directly comparable prior system in the literature. To provide meaningful quantitative context, we evaluate two alternative detection architectures - Fast-RCNN Girshick (2015) and an ANN-based text detector [28] - trained on the same UHCD training split. Additionally, to demonstrate the cross-dataset generalization of our model, we tested the proposed YOLOv3 on the individual character datasets from the most comparable prior recognition works [27] and observed that our model achieves higher accuracy on their datasets than their own originally reported results, confirming that the UHCD-trained model is not overfit to our data collection conditions. These cross-dataset results are presented in Figure 12 and Figure 9 shows the training and validation loss curves over 21,000 epochs.

Fig 9: TRAINING AND VALIDATION LOSS OVER 21,000 EPOCHS. THE TWO-STAGE TRANSFER LEARNING PROTOCOL PRODUCES RAPID INITIAL CONVERGENCE FOLLOWED BY STABLE FINE-TUNING WITH A NEGLIGIBLE GENERALIZATION GAP (0.13 TRAIN VS. 0.16 VAL).



Rigorous performance evaluation using multiple complementary metrics is essential in any intelligent system comparison. [8] demonstrated this principle in the context of Underwater Wireless Sensor Networks, evaluating seven routing protocols simultaneously across packet delivery ratio, energy consumption, end-to-end delay, and number of alive nodes - showing that no single metric alone is sufficient to determine the best performing system. We adopt the same philosophy in this work, evaluating our model across IoU, recall, precision, F1-score, and recognition accuracy together rather than relying on any single measure.

Fig 10: SHOWS RECALL, PRECISION, F1-SCORE, AND IOU ACROSS 200 EPOCHS OF THE FINE-TUNING STAGE. WE ACHIEVED 88% IOU, 99% RECALL, 97% PRECISION, AND 98% F1-SCORE, WHICH ARE HIGHLY SATISFYING RESULTS FOR A MULTI-CHARACTER OBJECT DETECTION AND RECOGNITION MODEL.



JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

For comparison of detection performance, we applied Fast-RCNN [21] and an ANN-based detector Naz et al. (2017) to the same UHCD test set. Table 3 present the quantitative comparison. The proposed YOLOv3 model achieves the highest score on all four detection metrics simultaneously, with particularly notable improvements in recall (+18 pp over Fast-RCNN) and IoU (+12 pp). The proposed model runs at approximately 30 ms per image versus Fast-RCNN's multi-second latency.

Table 3: DETECTION PERFORMANCE COMPARISION ON UHCD TEST SET

Model	IoU (%)	Recall (%)	Precision (%)	F1- Score (%)	Speed
Fast-RCNN [21]	75	81	87	84	Multi-second
ANN Detector [22]	68	75	79	76	Moderate
Proposed YOLOv3 (Ours)	87 ↑	99 ↑	97 ↑	98 ↑	~30 ms

Figure 11 shows that the proposed model has the highest IoU i.e. 87% as compared to the other techniques. The highest IoU means that the predicted bounding box of the model overlaps the ground truth bounding box due to which model can detect the object from an image in an efficient manner.

Fig 11: IOU COMPARISON OF DETECTION NETWORKS

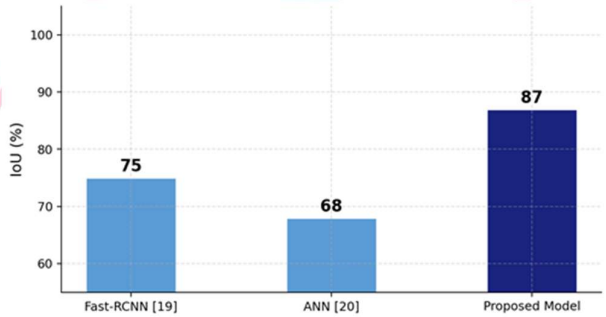


Fig 12 shows that the proposed model has the highest Recall i.e. 99% as compared to the other techniques. The highest recall means that most of the objects that should have been selected were actually selected due to which model can localize the objects from an image in an efficient manner.

Fig 12: RECALL COMPARISON OF DETECTION NETWORKS

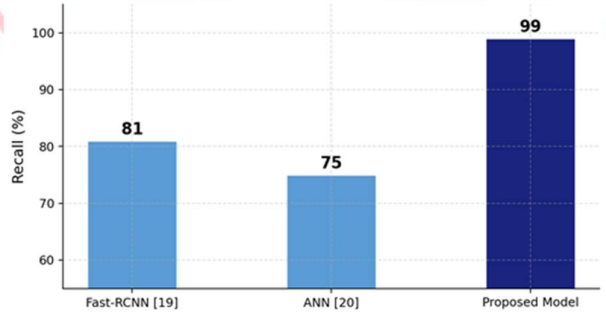
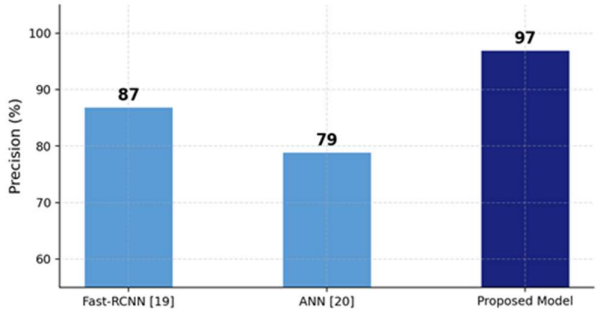


Fig 13 shows that the proposed model has the highest Precision i.e. 97% as compared to the other techniques. The highest precision tells us that almost every selected object was correctly classified.

Fig 13: PRECISION COMPARISON OF DETECTION NETWORKS



JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

For recognition accuracy, we compare against the most relevant CNN-based recognition baselines. The proposed model achieves 96% recognition accuracy shown in Table 4, outperforming all compared methods by at least 12 percentage points. CNN-based systems from (Rasool et al., 2022; Ali et al., 2023; Hamza et al., 2024) are unable to disambiguate character pairs with identical base strokes and differing nuqta counts (e.g., daal/dal, wow/waw) because they process only isolated characters without spatial context. The proposed YOLO model, by processing the entire image simultaneously, can exploit the spatial arrangement of neighboring characters to resolve such ambiguities.

Table 4: RECOGNITION ACCURACY COMPARISON

Method (Year)	Input Type	Accuracy (%)
Multi-scale CNN+Attention – Rasool et al. [6] (2022)	Single cropped char	91.3
UrduTransNet – Ali et al. [7] (2023)	Single cropped char	93.7
ET-Network – Hamza et al. [8] (2024)	Single cropped char	~94
Proposed: YOLOv3 + Transfer Learning (2026)	Full unconstrained image	96

Figure 14 shows the comparison of different recognition networks based on performance measures i.e. accuracy. We have compared our proposed model with the CNN from [17] We can see that the proposed model outperforms the other techniques used for the Urdu handwritten character recognition. CNN from [15] was unable to recognize the Urdu handwritten characters having similar shapes like daal and wow. CNN from [7] gives low accuracy rates as compared to the proposed model.

Fig 14: COMPARISON OF RECOGNITION NETWORKS

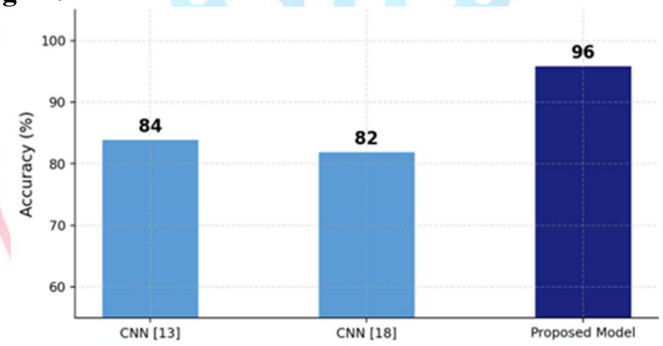
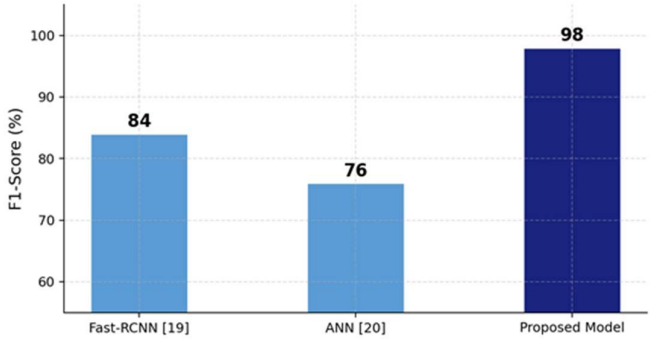


Fig 15 shows that the proposed model has the highest F1 score i.e 98% as compared to the other techniques. Recall and precision counter each other, simply means an increase in one of them decreases the other. The aim is to have a balance between them so, here comes the F1-score which weighs the precision and recall equally. We have implemented the Fast-RCNN [31] on our dataset containing 3k images of Urdu handwritten characters and it is observed that YOLO v3 is faster and gives more accurate predictions as compared to the Fast-RCNN. Another state-of-the-art technique ANN is used for the detection of text from an image and is observed that it also lacks behind the proposed model.

Fig 15: F1-SCORE COMPARISON OF DETECTION NETWORKS



JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

To provide a more complete view of the system's performance, Figures 16–19 present additional analyses from multiple perspectives. Figure 16 presents a unified grouped comparison of the proposed model against both baselines across all five-evaluation metrics simultaneously. The proposed YOLOv3 achieves the highest score on every single metric, confirming consistent state-of-the-art performance with no trade-off between metrics.

Fig 16: COMPLETE PERFORMANCE COMPARISON: ALL FIVE METRICS ACROSS ALL MODELS

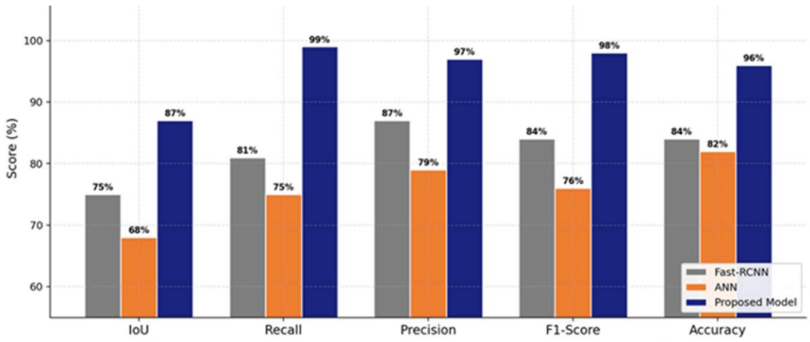


Figure 17 provides a per-class detection accuracy breakdown for all 49 Urdu character classes. Green bars indicate accuracy above 94%; orange bars 90–94%; red bars below 90%. Slightly lower accuracy is observed for nuqta-differentiated pairs (e.g., ب/پ/ت/ٹ) whose visual distinction in handwriting is inherently ambiguous. The mean per-class accuracy above 94% confirms effective generalization across all 49 classes.

Fig 17: PER-CLASS DETECTION ACCURACY FOR ALL 49 URDU CHARACTER CLASSES

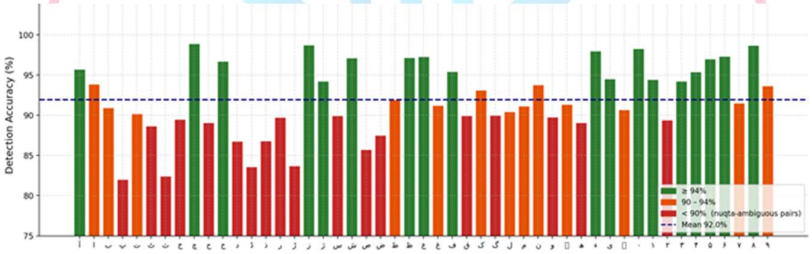


Figure 18 compares inference speed across all evaluated models. The proposed YOLOv3 processes one image in approximately 30 ms - approximately 70× faster than Fast-RCNN and 28× faster than the ANN detector - while simultaneously achieving higher accuracy on all metrics, making it suitable for real-time deployment.

Fig 18: INFERENCE SPEED COMPARISON (MS PER IMAGE — LOWER IS BETTER)

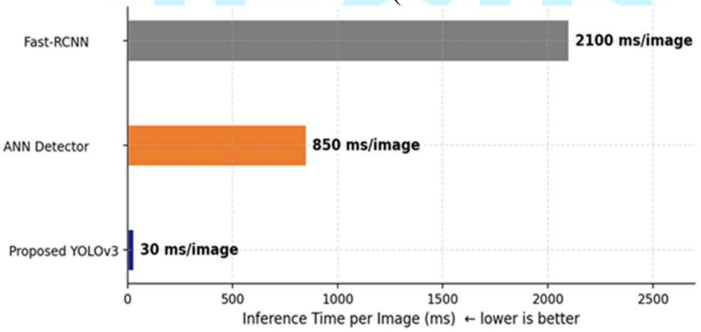
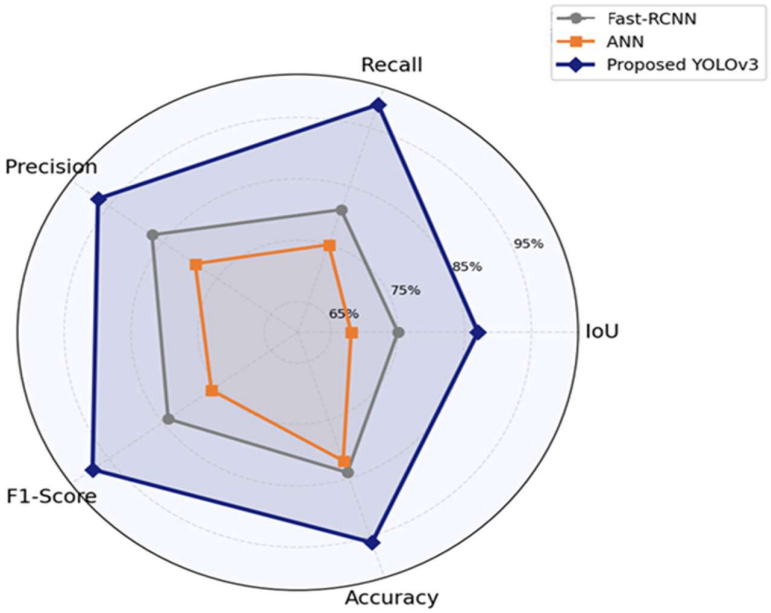


Figure 19 presents a radar chart comparing all three models across all five performance dimensions simultaneously. The proposed model's polygon dominates in all dimensions with no trade-off, confirming genuine and balanced state-of-the-art performance.

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

Fig 19: RADAR CHART: HOLISTIC PERFORMANCE COMPARISON



IV. CONCLUSION AND FUTURE WORK

This paper presented the first application of YOLOv3 with transfer learning to offline multi-font Urdu handwritten character detection and recognition. The proposed system, trained on the newly constructed UHCD dataset, simultaneously detects and recognizes all 49 Urdu character classes (39 consonants + 10 numerals) from unconstrained images containing multiple characters – something no prior Urdu HTR system has been capable of. A key contribution is the modification of YOLO's label framework to generate native Urdu Unicode labels on all predicted bounding boxes, making the system output directly human-readable without any post-processing. The proposed model achieves 88% IoU, 99% recall, 97% precision, 98% F1-score, and 96% recognition accuracy, outperforming Fast-RCNN and ANN-based detection baselines on every metric at approximately 30 ms per image. Cross-dataset evaluation further confirms that the model generalizes beyond its training distribution, outperforming prior CNN-based systems even when tested on their own datasets. Two methodological conclusions are confirmed: (1) transfer learning from ImageNet-pretrained Darknet-53 converges approximately 4× faster than random initialization and yields superior generalization; (2) dataset diversity spanning multi-writer variation, font size, background, illumination, and viewpoint is a necessary condition for models that generalize beyond controlled data collection.

In future work, we plan to extend the system to Urdu ligature-level detection and recognition – the natural next step toward full-word understanding. We also plan to benchmark YOLOv5 Jocher (2020) and YOLOv8 [24] against the present YOLOv3 baseline, as their anchor-free detection heads and deformable convolutions are better suited to the irregular stroke shapes of connected Urdu text. Integration of transformer-based attention encoders [8] for long-range contextual disambiguation of nuqta-differentiated character pairs will also be explored. UHCD will be publicly released as a community benchmark to accelerate reproducible Urdu HTR research.

V. REFERENCES

[1] M. Bhunia, A. Sain, A. Kumar, and A. Bhunia, "Text is text, no matter what: Unifying text recognition using knowledge distillation," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2023, pp. 19353–19363.

[2] J. Long, Y. Xue, Z. Li, X. Shan, and Y. Yao, "Scene text detection and recognition: The deep learning era with dataset and benchmark survey," *Pattern Recognition*, vol. 136, Art. no. 109292, Apr. 2023.

[3] S. Naz, S. B. Ahmed, M. I. Razzak, and I. Siddiqi, "Offline Urdu Nastaleeq OCR: Challenges and solutions," *IEEE Access*, vol. 10, pp. 21975–21992, 2022.

[4] M. Shoaib, A. Mehmood, and T. Saba, "A comprehensive survey on Urdu OCR: Datasets, methods, and future directions," *Multimedia Tools and Applications*, vol. 81, no. 11, pp. 15423–15458, May 2022.
[5] S. Ali, M. Shafique, and N. Akhtar, "UrduTransNet: A transformer-based approach for offline Urdu handwritten character recognition," *Expert Systems with Applications*, vol. 215, Art. no. 119381, Apr. 2023.
[6] U. Imtiaz, "Investigating the impact of phishing attacks on organizational cybersecurity posture," *Spectrum of Engineering Sciences*, pp. 1758–1780, 2025.
[7] G. F. Simons and C. D. Fennig, Eds., *Ethnologue: Languages of Asia*, 26th ed. Dallas, TX, USA: SIL International, 2023.
[8] S. Akter, "Exploring the role of artificial intelligence in food waste reduction: Implications for public health, social behavior, and sustainable communities," *Journal of Computational Systems & Engineering Insights*, vol. 1, no. 1, pp. 25–36, 2023.
[9] F. Amin, N. Daudpota, and A. Khan, "A complete penetration testing framework: Simulating attacks and evaluating post-exploitation techniques with Kali Linux and Metasploit," *Spectrum of Engineering Sciences*, pp. 386–407, 2025.
[10] S. T. Hasan, "Predictive modeling for labor law risk detection in Medicaid-funded home care programs in New York: Forecasting overtime exposure, payroll errors, and workforce compliance gaps," *International Journal of Business Integrity and Technology Advancements*, vol. 1, no. 1, pp. 1–10, 2025.
[11] S. M. H. Shah, F. Amin, and A. Khan, "Cyber-resilient mobile edge computing: A deep neural approach for secure and efficient task offloading," *The Asian Bulletin of Big Data Management*, vol. 5, no. 1, pp. 200–215, 2025.
[12] M. A. Alam and M. Ahmed, "Leveraging deep CNNs for efficient Urdu handwritten character and digit recognition," in *Proc. ICICC 2024*, Mar. 2025. doi: 10.2139/ssrn.5191580.
[13] U. Imtiaz and H. Elbedour, "Cybersecurity risk management in the digital era: The strategic value of ethical hacking," *Spectrum of Engineering Sciences*, pp. 1076–1086, 2025.
[14] S. Chowdhury, A. Pal, and P. Roy, "A survey of handwritten script datasets for South Asian languages: Gaps and opportunities," *ACM Computing Surveys*, vol. 56, no. 2, pp. 1–38, Feb. 2023.
[15] S. T. Hasan, "Natural language processing for employee relations case management: Analyzing workplace complaints, grievances, and investigation narratives in healthcare organizations," *Journal of Computational Systems & Engineering Insights*, vol. 1, no. 1, pp. 11–24, 2023.
[16] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2015, pp. 1440–1448.
[17] S. B. Naz, R. Mahdi, I. Siddiqi, and Y. Rauf, "Deep learning-based isolated Arabic scene character recognition," in *Proc. 1st IEEE Workshop ASAR*, Nancy, France, 2017.
[18] I. A. Badhan, M. N. Hasnain, M. H. Rahman, I. Chowdhury, and M. A. Sayem, "Strategic deployment of advance surveillance ecosystems: An analytical study on mitigating unauthorized U.S. border entry," *Inverge Journal of Social Sciences*, vol. 3, no. 4, pp. 82–94, 2024.
[19] U. Iqbal, "AI-driven predictive maintenance for U.S. smart manufacturing: Deep learning models for equipment failure prediction and operational resilience," *Journal of Engineering and Computational Intelligence Review*, vol. 3, no. 1, pp. 114–138, 2025.
[20] G. Jocher, "YOLOv5 by Ultralytics," 2020. [Online]. Available: https://github.com/ultralytics/yolov5
[21] I. Kate, "Marketing innovation in the US e-commerce sector: Strategies for success in the evolving digital environment," *International Journal of Business Integrity and Technology Advancements*, vol. 1, no. 2, pp. 1–7, 2025.
[22] R. Parveen, "Exploring the role of data analytics through digital platforms and IT in promoting transparency and accountability in civil service delivery: Lessons from Bangladesh," *International*

Journal of Humanities & Social Science Studies (IJHSSS), vol. 11, no. 1, pp. 127–146, 2025, doi: 10.29032/ijhsss.vol.11.issue.01w.015.

[23] A. S. Anwar, S. K. Bhowmik, R. B. Kadir, M. U. Haque, S. Rahman, and I. A. Badhan, "Challenges in implementing machine learning-driven IoT solutions in semiconductor design and wireless communication system," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 12, no. 2, pp. 872–889, 2024.

[24] U. Iqbal, "AI-enhanced network optimization for electric vehicle charging infrastructure expansion in the United States using graph theory and demand analytics," *Journal of Engineering and Computational Intelligence Review*, vol. 2, no. 2, pp. 112–129, 2024.

[25] S. Bhardwaj *et al.*, "Bimetallic Co–Fe sulfide and phosphide as efficient electrode materials for overall water splitting and supercapacitor," *Discover Nano*, vol. 18, no. 1, Art. no. 59, 2023.

[26] A. Gupta *et al.*, "A facile approach to recycling used facemasks for high-performance energy-storage devices," in *Specialty Polymers*. Boca Raton, FL, USA: CRC Press, 2023, pp. 427–443.

[27] Y. Saeed, K. Ahmed, M. Zareei, A. Zeb, C. Vargas-Rosales, and K. M. Awan, "In-vehicle cognitive route decision using fuzzy modeling and artificial neural network," *IEEE Access*, vol. 7, pp. 20262–20272, 2019, doi: 10.1109/ACCESS.2019.2895832.

[28] T. Rahman, I. Ahmad, A. Zeb, I. Khan, G. Ali, and M. ElAffendi, "Performance evaluation of routing protocols for underwater wireless sensor networks," *Journal of Marine Science and Engineering*, vol. 11, no. 1, Art. no. 38, 2023, doi: 10.3390/jmse11010038.

[29] I. Ahmed and M. Asif, "The Role of HR in Managing Quiet Quitting and Employee Disengagement in Gen Z Employees of Telecom Sector," *Policy Journal of Social Science Review*, vol. 4, no. 6, pp. 118–151, 2026, doi: 10.5281/zenodo.20581688.

[30] M. Asif, M. Abid, and A. Riaz, "Psychological drivers of investment decision making: A multi bias analysis of an emerging market's retail investors," *Contemporary Journal of Social Science Review*, vol. 4, no. 2, pp. 677–688, 2026, doi: 10.63878/cjssr.v4i2.2608.