Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.063255

ARTICLE

A Convolutional Neural Network Based Optical Character Recognition for Purely Handwritten Characters and Digits

Syed Atir Raza[1,2], Muhammad Shoaib Farooq[1], Uzma Farooq[3], Hanen Karamti[4], Tahir Khurshaid[5,*] and Imran Ashraf[6,*]

[1]Department of Computer Science, School of System and Technology, University of Management and Technology, Lahore, 54000, Pakistan
[2]Department of Applied Computing Technologies, FoIT&CS, University of Central Punjab, Lahore, 54590, Pakistan
[3]Department of Software Engineering, University of Management and Technology, Lahore, 54000, Pakistan
[4]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia
[5]Department of Electrical Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea
[6]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea

*Corresponding Authors: Tahir Khurshaid. Email: tahir@ynu.ac.kr; Imran Ashraf. Email: imranashraf@ynu.ac.kr

Received: 09 January 2025; Accepted: 16 May 2025; Published: 03 July 2025

ABSTRACT: Urdu, a prominent subcontinental language, serves as a versatile means of communication. However, its handwritten expressions present challenges for optical character recognition (OCR). While various OCR techniques have been proposed, most of them focus on recognizing printed Urdu characters and digits. To the best of our knowledge, very little research has focused solely on Urdu pure handwriting recognition, and the results of such proposed methods are often inadequate. In this study, we introduce a novel approach to recognizing Urdu pure handwritten digits and characters using Convolutional Neural Networks (CNN). Our proposed method utilizes convolutional layers to extract important features from input images and classifies them using fully connected layers, enabling efficient and accurate detection of Urdu handwritten digits and characters. We implemented the proposed technique on a large publicly available dataset of Urdu handwritten digits and characters. The findings demonstrate that the CNN model achieves an accuracy of 98.30% and an F1 score of 88.6%, indicating its effectiveness in detecting and classifying Urdu handwritten digits and characters. These results have far-reaching implications for various applications, including document analysis, text recognition, and language understanding, which have previously been unexplored in the context of Urdu handwriting data. This work lays a solid foundation for future research and development in Urdu language detection and processing, opening up new opportunities for advancement in this field.

KEYWORDS: Image processing; natural language processing; handwritten Urdu characters; optical character recognition; deep learning; feature extraction; classification

1 Introduction

Urdu is an Indo-Aryan language that emerged during the Mughal Empire in the Indian subcontinent [1]. Its rich history can be traced back to the 11th century, a period when Persian served as the official language of the Mughal courts. Urdu was developed as a hybrid language, skillfully incorporating vocabulary, syntax, and grammatical structures from Persian, Arabic, and various regional languages, which contributed to its unique linguistic identity. Throughout the 18th and 19th centuries, Urdu solidified its status as the lingua franca

of the Indian subcontinent, facilitating communication among diverse linguistic communities and gaining prominence during the British colonial period [1]. Today, Urdu is recognized as the national language of Pakistan and is spoken by nearly 100 million people worldwide. Its vibrant literary heritage, characterized by poetry, prose, and a wealth of cultural expressions, significantly enriches its legacy and continues to influence contemporary literature and art forms [2].

Current optical character recognition (OCR) systems designed for Urdu primarily focus on digital or printed materials, leaving a notable gap in effective methodologies for accurately recognizing pure handwritten Urdu digits and characters [3,4]. This limitation presents obstacles across various applications including document analysis, postal services, and banking, where precise text recognition is crucial. Consequently, there is a pressing need for a robust OCR system that can effectively recognize pure handwritten Urdu numerals and characters. This area has emerged as a prominent research topic within the broader field of OCR learning and classification, attracting considerable interest from researchers and practitioners alike.

Recent advances in deep learning (DL) and computer vision [5] have demonstrated the remarkable potential of convolutional neural networks (CNNs) in achieving state-of-the-art performance across a variety of image recognition tasks [6–8], including OCR [9]. CNNs have proven to be particularly effective in extracting robust features from images, which makes them exceptionally well-suited for the recognition of handwritten characters and digits [10,11]. Consequently, numerous researchers have conducted comparative studies to evaluate the performance of different methodologies based on similar criteria [12].

The motivation for this study arises from the notable scarcity of effective methods for detecting and recognizing pure handwritten Urdu digits and characters produced with a pen or pencil. While significant progress has been made in the field of Urdu character recognition, the majority of contemporary techniques predominantly focus on printed or typed text, thereby overlooking the variations inherent in handwritten Urdu data. This study aims to address this critical gap by developing a novel approach that facilitates the reliable identification and recognition of pure handwritten Urdu numerals and characters. By leveraging the strengths of CNNs, this research contributes to OCR technologies specifically tailored for handwritten Urdu, ultimately enhancing applications in various domains. Through this work, we hope to pave the way for future innovations in the recognition of handwritten scripts, thereby enriching the landscape of Urdu language processing.

In this study, a CNN-based OCR approach is proposed for pure handwritten Urdu digits and characters. The proposed approach employs several image preprocessing and data augmentation techniques to improve the model's robustness. The approach employs convolutional layers to extract significant features which are then utilized by fully connected layers to accurately and efficiently detect Urdu handwritten digits and characters. The proposed technique achieves an accuracy of 98.3% on a comprehensive evaluation, outperforming existing state-of-the-art Urdu OCR systems. The proposed CNN architecture and methodology can be extended to recognize other scripts and languages.

Further, Section 2 explains the related work. Section 3 is about the materials and methods which has further subsections that explain the working of the proposed model. Section 4 provides experiments and results and Section 5 provides the conclusion.

2 Related Work

Much work has been done in the field of Urdu OCR but most of the work is to recognize digital Urdu as mentioned previously. Digital fonts are Nastaliq, Koufi, Thuluthi, Diwani, Rouqa, and Naskh, and several approaches exist for their identification [12]. No work has been found so far to recognize the pure handwritten Urdu digits and characters written by different people in different handwriting styles with the help of pen

or pencil. The OCRs have been built for Urdu to recognize the Urdu characters shown in Fig. 1 from the image data.

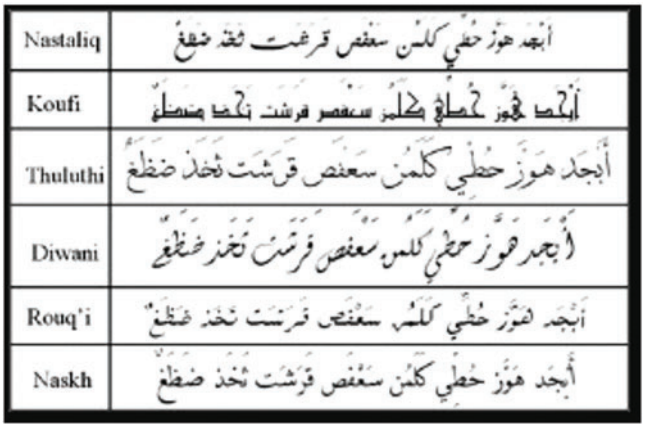

Figure 1: Digital Urdu major fonts

Kashif [13] implemented the ResNet18 model on the Urdu Nastaliq Handwritten Dataset (UNHD), achieving 94.4% accuracy in text recognition. The approach demonstrates the effectiveness of DL in handling the complexities of Urdu Nastaliq handwriting, highlighting the potential for further advancements in character and digit recognition.

KO and Porunan [14] introduced OCR-Nets, utilizing AlexNet and GoogleNet for handwritten Urdu character recognition through transfer learning. The approach was evaluated using a combined dataset, which was further enhanced by a manually generated Urdu character dataset containing diverse fonts and sizes. This additional dataset was incorporated to ensure a fair comparison with traditional character recognition methods and to assess the effectiveness of DL models in recognizing handwritten Urdu text. The experimental results demonstrated that OCR-AlexNet and OCR-GoogleNet achieved improvements, with averaged accuracy of 96.3% and 94.7%, respectively. While the study effectively explored DL-based OCR for Urdu characters, it primarily focused on printed and semi-handwritten text rather than purely handwritten Urdu digits and characters. The structural complexity of the Urdu script, which includes intricate curves, loops, and varying stroke thicknesses, presents intricate challenges in handwritten OCR. Moreover, the study did not extensively address the recognition of handwritten numerals, which are crucial for many practical applications such as financial and official document processing. This gap highlights the need for further research on OCR models specifically designed for recognizing purely handwritten Urdu digits and characters, ensuring improved accuracy and adaptability to diverse handwriting styles.

Rizvi et al. [15] developed an Urdu OCR system that achieved an impressive accuracy of 98% in recognizing the Nastalique Urdu script using the discriminative elastic matching learning (DEML) algorithm. Nastalique is a widely used digital font in Urdu typing software, including Google Urdu Keyboard and Harf Kar, making it an essential component of digital Urdu text processing. The study effectively demonstrated the capability of DEML in recognizing digitally rendered Urdu scripts with high accuracy. However, the work primarily focused on printed and digitally generated Nastalique text rather than purely handwritten Urdu characters and digits. Handwritten Urdu script presents additional complexities including variations in individual writing styles, inconsistent stroke thickness, and irregular spacing, which make OCR significantly more challenging. Furthermore, while the study contributed to digital font recognition, it did not address the

critical need for recognizing handwritten Urdu numerals, which are essential in various applications, such as banking, historical document digitization, and administrative record-keeping.

Rizvi et al. [16] developed a supervised learning-based OCR system for recognizing Urdu text for the Nastalique script, achieving an accuracy of 97.3%. The work demonstrated the effectiveness of supervised learning techniques in processing digitally rendered Urdu text, which is commonly used in print media and digital platforms. The study contributed to the growing field of Urdu OCR by enhancing recognition accuracy for the standardized Nastalique font, which is widely utilized in Urdu typing applications. However, the research was limited to printed or digitally generated Urdu text and did not address the complexities associated with recognizing purely handwritten Urdu numerals and characters. Handwritten Urdu script varies significantly among individuals due to differences in stroke formation, writing speed, and stylistic inconsistencies, making character recognition much more challenging. The absence of a robust OCR solution for handwritten Urdu script represents a crucial research gap, as handwritten text is still prevalent in official documents, educational settings, and historical manuscripts. Addressing this challenge requires advanced DL models capable of handling the variability inherent in handwritten Urdu, ensuring accurate recognition across diverse handwriting styles.

Nasir et al. [17] proposed a DL framework to improve the accuracy and efficiency of Urdu text recognition. The study focused on developing a robust model capable of processing Urdu script with greater precision, addressing challenges posed by the complex structure of the language. The research highlighted the limitations of existing OCR systems, particularly in handling Urdu text, which often features varying ligatures, overlapping characters, and diverse writing styles. By leveraging DL techniques, the framework demonstrated significant improvements in text recognition, contributing to advancements in the field of Urdu OCR. However, the work focused mainly on printed or digitally generated Urdu text and did not extensively explore the recognition of purely handwritten Urdu characters and numerals. A major challenge in handwritten Urdu OCR is the lack of large, well-annotated datasets that encompass the natural variations in handwriting styles. Without such datasets, developing highly accurate models remains difficult. Their research underscored the need for further advancements in this area, as effective recognition of handwritten Urdu text is essential for applications such as document digitization, historical manuscript preservation, and automated handwriting analysis. This gap highlights the necessity for more sophisticated DL approaches specifically tailored for handwritten Urdu script.

Naseer et al. [18] developed an OCR system for recognizing calligraphic Nastalique Urdu script using the normal-to-tangent line (NTL) approach, achieving an accuracy of 90% and 75% for different datasets. The research focused on improving the recognition of digitally written Urdu text, particularly in calligraphic styles, which pose unique challenges due to the script's inherent complexity, cursive nature, and overlapping ligatures. By employing the NTL approach, the system demonstrated notable improvements in processing Nastalique script, contributing to advancements in Urdu OCR technology. However, the study was limited to calligraphic and digitally generated Urdu text and did not address the recognition of purely handwritten Urdu characters and numerals. Handwritten Urdu script presents a significantly complex challenge due to variations in individual writing styles, inconsistencies in stroke formation, and noise introduced during the writing process. The lack of a comprehensive dataset for handwritten Urdu text further complicates this issue, making recognition tasks more difficult.

Misgar et al. [19] used recurrent neural network (RNN) and long short-term memory (LSTM) models to create a recognition system for Urdu characters and digits. The suggested system showed high accuracy of 96.9% and 97% for character recognition using the RNN and LSTM models, respectively, and 85% for digit recognition using the RNN model. These findings show the efficacy of DL-based algorithms

for recognizing Urdu characters and numbers, making a significant contribution to the field of Urdu handwriting recognition.

Mosbah et al. [20] integrated CNNs and bidirectional LSTM (BiLSTM) with the connectionist temporal classificatioin (CTC) algorithm to address the complexities of Arabic script. Evaluated on printed and handwritten datasets (P-KHATT, APTI, IFN/ENIT), ADOCRNet achieves state-of-the-art recognition accuracy, significantly reducing error rates compared to existing OCR solutions. Despite advancements in OCR for Arabic script, existing models struggle with the complexities of Arabic writing, including cursiveness, diacritics, and varied fonts. Most current approaches lack efficiency in handling both printed and handwritten Arabic text with high accuracy. ADOCRNet bridges this gap by integrating CNNs and BiLSTM, significantly reducing error rates, yet further optimization for real-world applications is still needed.

Nabi et al. [21] proposed a DL model specifically designed for identifying writers of offline Urdu handwritten documents, utilizing DL's automatic feature extraction capabilities. The proposed system, inspired by the VGG-16 architecture, was trained on a novel dataset comprising samples from 318 distinct Urdu writers, achieving better accuracy. However, challenges remain in addressing the unique characteristics of Urdu handwriting, which have not been thoroughly explored in existing writer identification systems.

Cheema et al. [22] developed ViLanOCR, a bilingual OCR system attuned to Urdu and English, leveraging multilingual vision-language transformers to enhance recognition accuracy. The authors evaluated the approach using the character error rate (CER) metric on the Urdu UHWR dataset, achieving better performance with a CER of 1.1%. The system outperforms existing methods, demonstrating the effectiveness of transformer-based models in OCR tasks for low-resource languages. However, despite advancements in OCR, low-resource languages like Urdu remain underexplored, with existing models struggling to handle script complexities, diacritics, ligatures, and cursive variations. Most OCR solutions either lack fine-tuning for Urdu or fail to generalize well due to dataset limitations. This work addresses these challenges by adapting multilingual vision-language transformers specifically for Urdu OCR, bridging the gap between multilingual OCR and low-resource script optimization.

Hamza et al. [23] presented efficient transformer (ET)-Network that uses a vanilla transformer for language modeling and incorporates self-attention layers into EfficientNet for better feature extraction to improve the recognition of Urdu handwritten text. This technique addresses important issues in Urdu script recognition by improving text creation by prefix beam search, in contrast to conventional OCR methods that have trouble with cursive variations and writing inconsistencies. Due to script complexity and lack of annotated datasets, Urdu is still understudied concerning OCR, despite advancements in major languages. Long-range dependencies in handwritten text are not adequately captured by current models, which results in recognition mistakes. By tailoring transformer-based designs for Urdu, this study fills the gap and provides a more effective and flexible solution for low-resource handwritten OCR.

Mustafa et al. [24] suggested a word-level OCR model for digital Urdu text that addresses script complexity, font variances, and character reordering by utilizing transformer-based topologies and attention methods. By training on numerous token permutations, the permuted autoregressive sequence (PARSeq) architecture enhances context-aware inference and boosts recognition accuracy. Character reordering and contextual dependencies are not well handled by current models, and Urdu OCR is still difficult because of overlapping characters and script unpredictability. By incorporating PARSeq-based iterative refinement, this study fills these shortcomings and provides a more reliable and flexible method for Urdu text recognition.

Malik et al. [25] proposed a transfer learning-based framework for hate speech detection and target community identification in Nastaliq Urdu using Urdu-RoBERTa and Urdu-DistilBERT. The authors developed the HSTC corpus from Pakistani Facebook posts and employed fine-tuning with Grid search

to enhance classification accuracy. While hate speech detection in high-resource languages is well-studied, Nastaliq Urdu remains underexplored due to limited datasets and script complexity. Existing models struggle with fine-grained multi-class classification for target communities (religious, political, gender-based). This research bridges the gap by leveraging pre-trained Urdu transformers, offering an automated, robust approach for social media content moderation in under-resourced languages.

Urdu handwriting recognition has long posed challenges due to its complex script, varied writing styles, and the lack of extensive annotated datasets. While previous research has made strides in OCR for Urdu, most efforts have been centered on printed text. Some existing studies have explored Urdu handwriting recognition, however, the methodologies often fail to generalize well due to insufficient training data, limited feature extraction capabilities, or suboptimal classification techniques. In contrast, this study introduces a novel CNN-based approach specifically tailored for recognizing pure Urdu handwritten digits and characters. Unlike prior research, which either focuses on printed text or struggles with the variability of handwriting, this work leverages advanced image processing techniques and specialized data augmentation strategies to enhance recognition accuracy. A key innovation of the current approach lies in its strategic combination of convolutional feature extraction and fully connected classification layers, ensuring both robust feature learning and effective classification. By training the model on a large publicly available dataset, we achieve an unprecedented accuracy of 98.30% and an F1 score of 88.6%, surpassing previous benchmarks in Urdu handwriting recognition. This significant performance improvement underscores the efficiency of CNNs in handling the intricacies of Urdu script, including its cursive nature, ligatures, and diacritical marks. Furthermore, this work lays the foundation for future advancements in document digitization, automated text recognition, and Natural Language Processing (NLP) applications for Urdu. By addressing the shortcomings of earlier methodologies and setting a new benchmark, this research represents a groundbreaking contribution to Urdu language processing and opens new avenues for exploration in handwritten OCR systems, particularly for under-resourced languages (Table 1).

Table 1: Summary of discussed research works

Author	Contribution	Technique	Focus on hand written detection
This study	Urdu pure handwritten digits and characters written with a pen or lead pencil	CNN	Yes
Kashif [13]	Printed Urdu font detection only	ResNet18	No
KO and Porunan [14]	Printed Urdu detected using variants of AlexNet & GoogleNet	Transfer learning	No
Rizvi et al. [15]	Urdu Nastalique scripting language detected	Discriminative Elastic Matching Learning (DEML)	No
Rizvi et al. [16]	Urdu Nastalique printed form font detection	Supervised learning	No

(Continued)

Table 1 (continued)

Author	Contribution	Technique	Focus on hand written detection
Nasir et al. [17]	MMU OCR introduced for only printed Urdu fonts, characters, and digits	Deep learning VGG16	No
Naseer et al. [18]	OCR system for printed calligraphic Urdu font	Normal to tangent line (NTL)	No
Misgar et al. [19]	OCR introduced for printed Urdu characters and digits	RNN and LSTM	No
Mosbah et al. [20]	ADOCRNet enhances Arabic OCR with CNN-BLSTM, reducing errors	CNN and BLSTM	No
Nabi et al. [21]	Urdu documents detection using deep learning	VGG-16	No
Cheema et al. [22]	ViLanOCR, a bilingual OCR system tailored for Urdu and English	Transformers	No
Hamza et al. [23]	ET-Network enhances Urdu handwriting OCR using EfficientNet and transformers	Vanilla Transformers	No
Mustafa et al. [24]	Proposed PARSeq-based OCR model improving Urdu text recognition and context handling	PARSeq	No
Malik et al. [25]	Proposed transfer learning-based Nastaliq Urdu hate speech detection with fine-tuned transformers	DistilBert	No

3 Materials and Methods

This research presents a CNN-based OCR system for purely handwritten Urdu characters and digits. Implementation is done using TensorFlow and Keras within the Google Colab environment, utilizing Python as the primary programming language. The dataset consists of handwritten Urdu characters and digits contributed by over 900 individuals, ensuring diversity in writing styles. Each image is stored in a standardized 28 × 28 resolution, with an average of 140 images per character and digit in the test set. Preprocessing techniques such as grayscale conversion, binarization, noise removal, and segmentation are applied to enhance the quality of input data. Additionally, data augmentation including rotation, zooming,

and flipping are employed to improve robustness and generalization. The layered architecture for the proposed system is illustrated in Fig. 2.

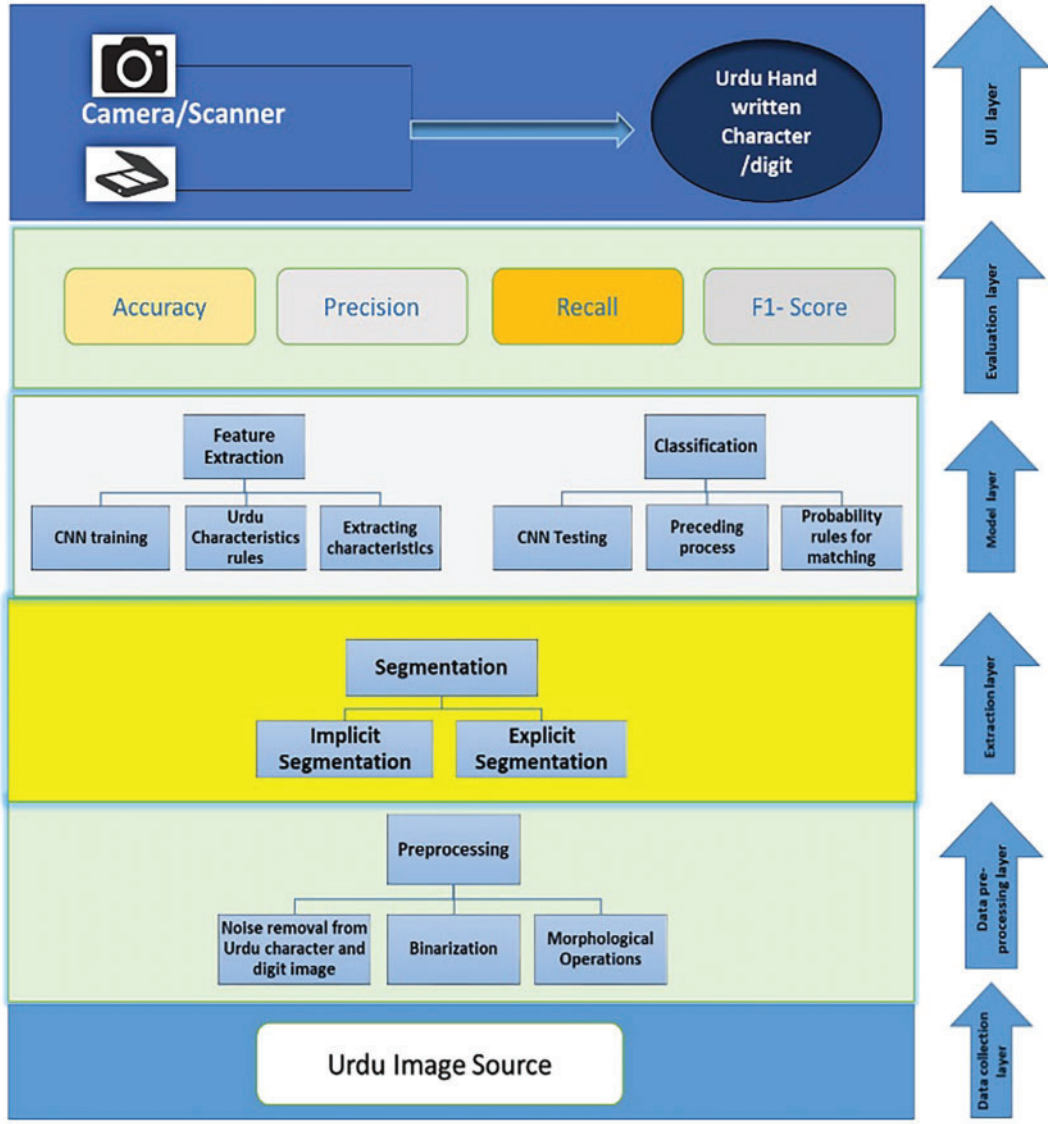

Figure 2: Layered architecture for the proposed system

The CNN architecture comprises six convolutional layers, along with batch normalization and max-pooling layers, which effectively extract meaningful spatial patterns from the handwritten script. The rectified linear unit (ReLU) activation function is used across all layers. The final layer contains the Softmax activation function. The 'Adam' optimizer is used while categorical cross-entropy serves as the loss function. The training process spans 50 epochs and the batch size is 128, ensuring optimal parameter updates. The dataset has an 80% to 20% division for training and testing, allowing a balanced evaluation of the proposed approach.

The experimental setup includes a high-performance computing workstation, an Intel Core i7 7th generation processor, and an NVIDIA GeForce GPU, allowing for efficient training and inference. Python and Keras provide the DL framework, ensuring a smooth implementation. The performance is assessed

using accuracy, precision, recall, and F1 score, demonstrating its effectiveness in recognizing handwritten Urdu text. The findings of this research emphasize the significance of DL in OCR applications, particularly for complex scripts like Urdu. The proposed system holds potential for practical applications including automated document processing, form recognition, and historical manuscript digitization. Future work could involve integrating attention mechanisms and transformer-based models to further refine recognition accuracy and enhance adaptability to highly variable handwriting styles.

3.1 Optical Character Recognition

OCR is used to translate text from images, typed or written characters, and words for automation. For existing OCR systems, typed characters can be identified with an average accuracy of 90% while handwritten characters have lower accuracy of 40% to 50% due to several challenges and complexities. OCR systems do the important job of transforming typed or handwritten text and numerals into computer files that can be further processed easily.

OCR is an essential technology that facilitates the conversion of text images into editable and searchable formats. As the volume of digital documents continues to surge, encompassing a wide array of formats such as scanned images and portable document format (PDF), the significance of OCR has grown exponentially in recent years. However, the task of accurately recognizing handwritten characters and numerals remains a complicated challenge. This difficulty arises from the diversity in individual writing styles, variations in image quality, and the absence of a standardized format for handwritten text. Particularly in the context of the Urdu language, the unique script poses additional challenges, making the recognition of handwritten Urdu numerals and characters exceptionally complex [26–28].

OCR systems comprise scanning, location segmentation, preprocessing, feature extraction, and recognition steps. During scanning, digital images of a document are taken using scanners. Although color images are also captured, images are converted to bilevel images for reduced computational complexity. Since image quality is important for improved final results, the thresholding approach is commonly applied for better contrast. Location segmentation locates the boundaries for the document by identifying regions containing characters or numerals [29]. Afterward, segmentation separates the characters which are later processed and recognized individually. Preprocessing involves several steps to improve the quality of the characters as they might be broken, noisy, tilted, etc., due to the segmentation process. Filling the gaps, smoothing, rotation, and color transformation are the common preprocessing steps for OCR systems. Feature extraction involves capturing important characteristics from characters and numerals that help the system better recognize characters. Selecting an appropriate feature extraction approach is very important and its choice depends on factors like noise, distortions, style variations, etc., found in the images [30]. Various features produce different results for the same image and show different levels of accuracy, robustness, immunity to noise, etc. Finally, the classification involves identifying the characters belonging to a particular class.

3.2 Dataset

We obtained the UHaT: Urdu handwritten text dataset from the Kaggle repository [31]. The dataset contains Urdu characters and numerals, all of which are handwritten. Over 900 people from Comsats University Islamabad, Pakistan contributed to these samples. The images have a resolution of 28 by 28. On average, each character's training set has 700 images. The train set has 811 images for AYN and 697 images are part of the train set for ALIF, for example. Similarly, the average train set has 700 photos. For example, there are 678 train set images for digit one. On average, each character's test set has 140 photos. For example, the letter ALIF has 145 test set images. On average, each digit's test set has 140 photos. For example, digit nine

has 147 test set photos. The dataset comprises four sections, training and test sets each for characters and digits. The original image of Urdu handwritten characters and digits is shown in Figure 3.

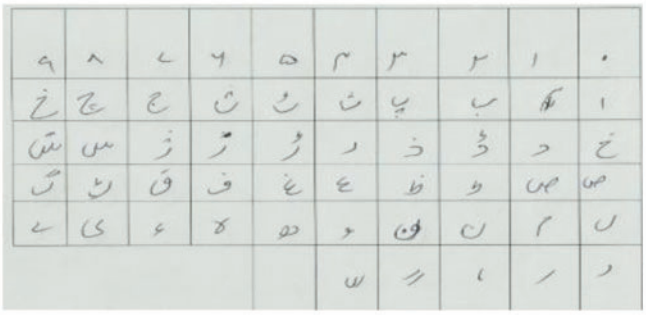

Figure 3: Original image of Urdu handwritten digits and characters

Images for characters and digits vary in number. Detailed information on the number of images for each character and digit is provided in Table 2.

Table 2: Details of the number of training images in the dataset

Character	# of images	Character	# of images	Character	# of images
Alif	697	Alif Mad Aa	748	Baa	525
Paa	764	Taa	734	Aynn	811
Ghayn	742	Faa	765	Jeem	689
Hamza	825	Kaaf	750	Chay	730
Daal	640	Dhaal	621	Gaaf	715
Choti Yaa	819	Bari Yaa	829	Noon	778
Laam	519	Meem	824	Noon Gunnah	723
Sawaad	751	Seen	729	Raa	582
Zaal	537	Zawaad	724	Tawaa	802
Rhraa	512	Ttaa	580	Zero	520
One	678	Two	668	Three	665
Four	670	Five	688	Six	674
Seven	678	Eight	682	Nine	683

3.3 Image Source

Urdu offline hand-written character recognition is introduced in this phase. The image source can be any digitized tool. The images are obtained using a scanner or camera and sent to the next phase.

3.4 Data Preprocessing

Preprocessing plays a crucial role in improving the accuracy of Urdu handwritten digit and character recognition by refining image quality before feature extraction. The first step, noise removal, eliminates unwanted distortions caused by scanning imperfections, ink smudges, or background interference. Techniques such as Gaussian and median filtering are applied to smooth the image while preserving essential

character details (Gaussian filter is applied using a kernel size of 5 × 5 with α set to 1.0). Median filtering is used with a kernel size of 3 × 3. Next, image binarization is performed to convert grayscale images into binary format, distinguishing the foreground text from the background. Otsu's thresholding or adaptive thresholding techniques are commonly used to ensure optimal separation, making character boundaries more distinct (dilation, erosion, opening, and closing with a 3 × 3 kernel). Finally, morphological operations such as dilation, erosion, opening, and closing are applied to refine character structures. These operations help in removing small gaps, connecting broken strokes, and eliminating unwanted artifacts, thereby improving the overall clarity of handwritten Urdu text. By implementing these preprocessing steps systematically, the recognition model becomes more robust and capable of handling variations in handwriting styles, ink intensity, and scanning inconsistencies. This refined input significantly enhances the accuracy and efficiency of Urdu handwritten digit and character OCR systems, paving the way for better document digitization and language processing applications.

3.4.1 Noise Removal

Noise removal is a crucial preprocessing step in Urdu handwritten OCR using CNN. Handwritten text is often affected by noise due to ink smudges, paper texture, scanning artifacts, and background interference, which can hinder the OCR model's ability to accurately recognize characters. Effective noise removal techniques improve image quality, making character boundaries more distinct and enhancing recognition accuracy. Various filtering methods are employed for noise reduction, including Gaussian filtering, which smooths the image by averaging pixel values to reduce random variations. The min-max filtering method helps in enhancing contrast and eliminating unwanted distortions. By applying these techniques systematically, the input image becomes cleaner, allowing the CNN-based OCR system to focus on essential character features, ultimately leading to improved classification performance and robustness in recognizing diverse Urdu handwriting styles. Fig. 4 shows the noise removal process.

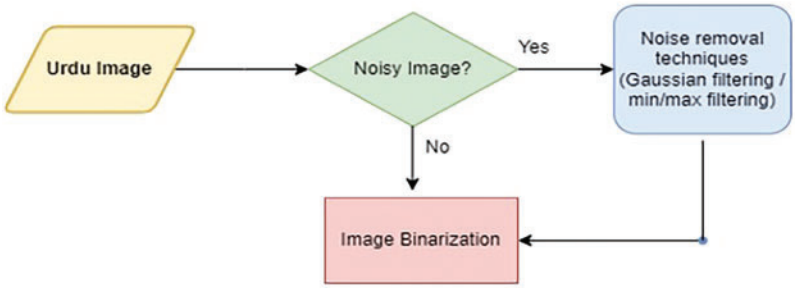

Figure 4: Noise removal from images

3.4.2 Image Binarization

Binarization is a fundamental preprocessing step in Urdu handwritten OCR, converting grayscale images into a binary format, each pixel is represented as either 0 (black) or 1 (white). This transformation simplifies the image by removing unnecessary details and reducing noise, making character recognition more efficient. The process enhances the contrast between the text and background, ensuring clearer character boundaries for better segmentation and feature extraction. Several binarization techniques are commonly used, with Otsu's thresholding and adaptive thresholding being the most widely applied. Otsu's method analyzes the histogram to determine an optimal threshold. Contrarily, adaptive thresholding can be used to adjust the threshold dynamically using the local pixel intensity variations. Adaptive thresholding is ideal for

images with uneven lighting. Binarization plays a crucial role in improving the accuracy of OCR systems, as a well-binarized image minimizes distortions and enhances character visibility. This step ensures that the CNN model focuses on essential features, improving overall recognition performance.

3.4.3 Morphological Operations

This is the process of enlarging or contracting an Urdu handwritten image. This is done primarily because the algorithm anticipates a constant image size. We can add pixels to the image's boundary to increase its size. We can reduce the size of an image by removing pixels from the image's boundary.

3.5 Segmentation

Segmentation is an important step in OCR for Urdu handwritten characters/digits, as it involves dividing the input image into separate, recognizable characters or digits. The goal of segmentation is to isolate individual characters in the image so that they can be recognized and processed individually. There are two types of segmentation: implicit and explicit.

3.5.1 Implicit Segmentation

In the domain of Urdu entirely handwritten numbers and characters, implicit segmentation entails training a CNN model to distinguish unique characters based on their natural properties and relationships. This means that the model learns to distinguish between characters without the use of explicit limits. Strokes and loops, for example, provide individual patterns for each character in cursive Urdu. The CNN layers are designed to detect these subtleties, starting with basic shapes and progressing to finer details. After extensive training on a variety of handwritten Urdu samples, CNN becomes capable of recognizing characters even within the continuous script. This method is especially useful when explicit segmentation is impossible, as it provides a powerful mechanism for precise recognition of both numbers and letters.

3.5.2 Explicit Segmentation

Explicit segmentation in the context of Urdu handwritten numerals and characters requires utilizing specific ways to precisely recognize and separate unique characters within the handwritten text. This strategy is especially useful when the characters are related or overlap. By employing specific segmentation processes, the CNN model is directed to find distinct borders between characters, ensuring that each character is precisely separated before recognition. This combined strategy gives the CNN model a full understanding of Urdu script, increasing its capacity to differentiate and categorize both digits and characters. In this situation, explicit segmentation provides an organized approach to character isolation, effectively supplementing CNN's ability to distinguish finer details. It is especially useful in situations when specific boundaries are required for precise recognition in handwritten Urdu script. Fig. 5 shows the segmentation process of Urdu OCR.

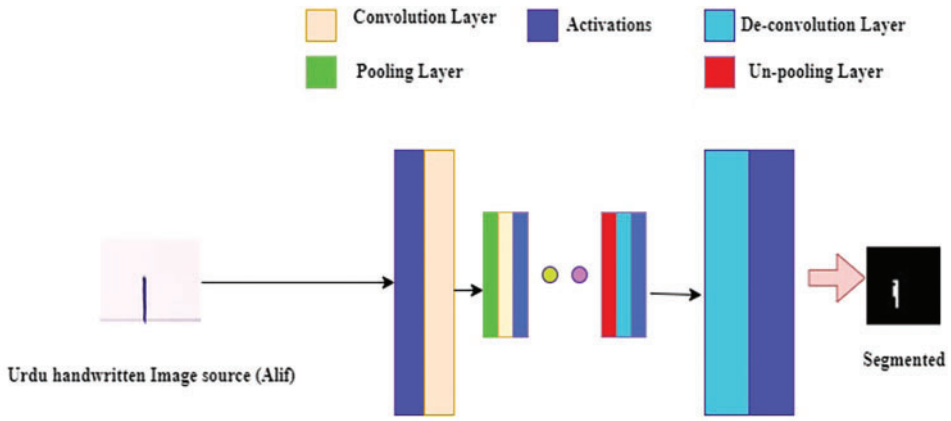

Figure 5: Segmentation process of Urdu OCR

3.6 Feature Extraction

Extracting features is a critical step in Urdu handwritten OCR, as it determines the quality of information fed into the CNN for accurate character classification. The selection of histograms of oriented gradients (HOG) and local binary patterns (LBP) was driven by their proven effectiveness in capturing essential structural and textural characteristics of handwritten text. HOG is widely used for analyzing gradient orientation distributions, making it particularly effective for recognizing complex Urdu script, which contains curves, loops, and varying stroke thicknesses. By computing edge directions and intensity variations, HOG enhances the model's ability to distinguish similar-looking characters, ensuring robust feature extraction. Meanwhile, LBP is a texture descriptor that encodes spatial relationships between neighboring pixels. It captures fine-grained texture patterns, making it highly useful for recognizing different handwriting styles, especially in challenging conditions such as noisy backgrounds and varying illumination levels. Mathematically, the HOG feature extraction process involves computing gradient magnitudes and orientations. The gradient at each pixel (x, y) is determined using the Sobel operator.

$$I_x = I * G_x, \qquad I_y = I * G_y \tag{1}$$

where I_x and I_y represent the image gradients in the horizontal and vertical directions, respectively, computed by convolving the image I with Sobel kernels G_x and G_y. The gradient magnitude and orientation are calculated as:

$$M(x, y) = \sqrt{((I_x(x, y))^2 + (I_y(x, y))^2)} \tag{2}$$

$$\theta(x, y) = tan^{-1}\left(\frac{I_y(x, y)}{I_x(x, y)}\right) \tag{3}$$

These values are then used to construct HoG orientations across spatially divided cells, ensuring robust shape-based feature extraction.

The LBP operator focuses on local texture information by thresholding neighborhood pixel values relative to a central pixel. Given a central pixel I_c and its surrounding pixels I_p LBP is computed as:

$$LBP_p, R = \sum_{p=0}^{p-1} s\left(I_p - I_c\right).2^p \tag{4}$$

where Pp presents the number of neighborhood pixels and, R is the radius and $s(x)$ is a step function.

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{5}$$

By converting local pixel intensity differences into a unique binary code, LBP effectively encodes texture patterns crucial for distinguishing different handwriting styles.

The integration of HOG and LBP into the CNN model offers a hybrid approach that combines both global and local feature representations, leading to improved accuracy and generalization. Rather than processing raw images directly, the extracted HOG and LBP features are converted into numerical vectors before being fed into the CNN. HOG ensures that the CNN captures edge and gradient information crucial for character differentiation, while LBP strengthens texture-based recognition, allowing the model to handle variations in handwriting with greater robustness. Fig. 6 shows the HoG features.

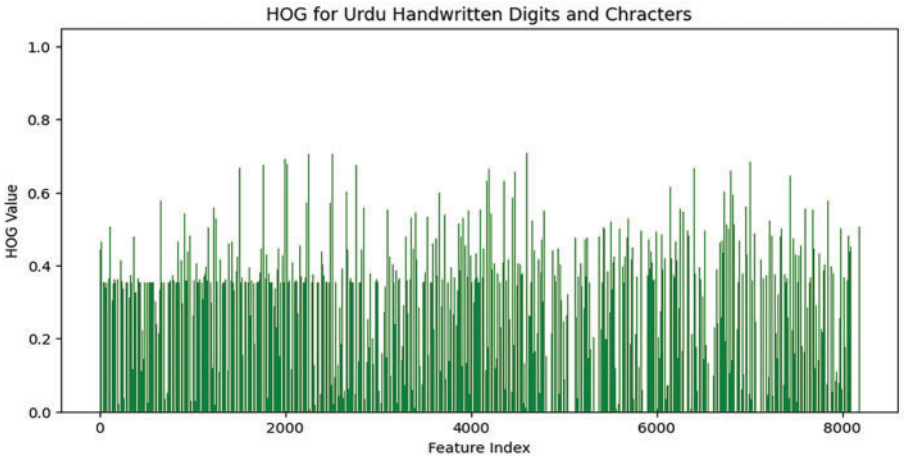

Figure 6: HOG for Urdu Handwritten digits and characters

Furthermore, LBP is particularly effective in dealing with complex backgrounds and handwritten text variations, as illustrated in Figs. 7 and 8. This makes it a valuable addition to the feature extraction pipeline, ensuring improved recognition performance even in real-world scenarios where handwriting styles can differ significantly. By leveraging the complementary strengths of HOG and LBP, the proposed model achieves superior OCR recognition accuracy, demonstrating its effectiveness in processing handwritten Urdu characters and digits. This strategic selection of features enhances CNN's learning capabilities, leading to a more reliable and efficient Urdu OCR system.

This strategic integration enables CNN to use both texture-based and gradient-based characteristics for improved OCR recognition accuracy. HOG computes the gradient orientation histograms of the input image and uses them as features for the CNN. LBP is a feature extraction technique that computes the local binary patterns of the input image and uses them as features for the CNN. LBP is effective for character recognition tasks, especially for text images with complex backgrounds as illustrated in Fig. 9.

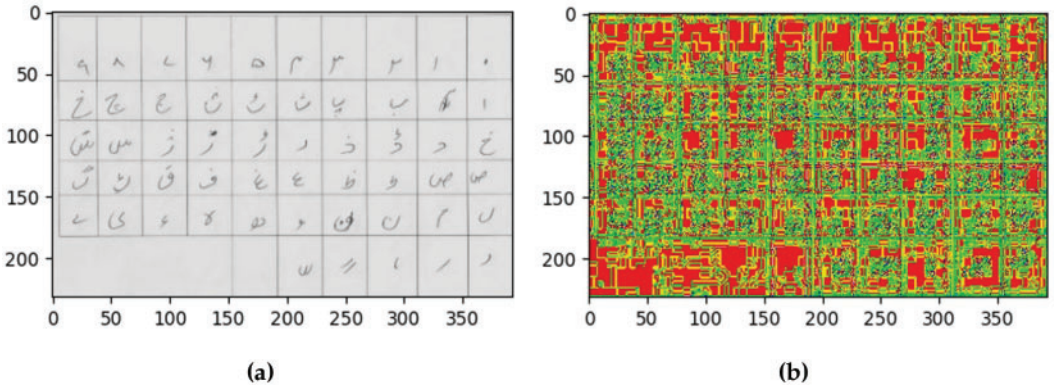

Figure 7: LBP for Urdu handwritten digits and characters, **(a)** Original image, **(b)** LBP

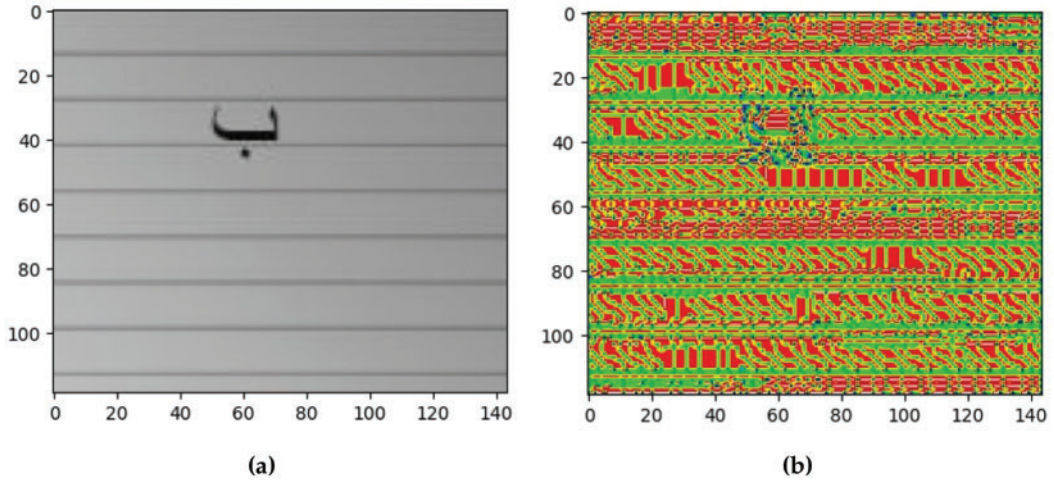

Figure 8: LBP for Single Urdu handwritten character, **(a)** Original image, **(b)** LBP

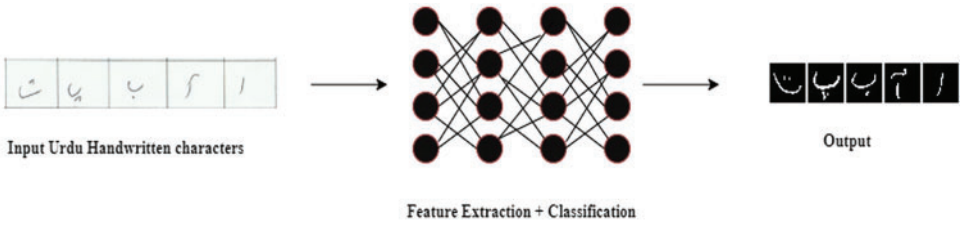

Figure 9: Illustration of feature extraction to classify Urdu characters in the proposed system

3.7 Classification

The classification step is the final step in OCR for Urdu handwritten digits and characters using a CNN. The goal of the classification step is to use the extracted features from the input images to recognize the characters or digits. Multilayer perceptron (MLP) is a feedforward neural network that can be used for character recognition tasks by learning a mapping from the input features to the character classes as illustrated in Fig. 9.

3.8 CNN Algorithm for Proposed Technique

The CNN-based OCR system for pure handwritten Urdu digits and characters is evaluated. The model was implemented using Python and the TensorFlow DL framework. We have used the dataset of Urdu digits and characters, previously discussed. The 'Adam' Optimizer and categorical cross-entropy loss function were used to optimize the model during the training phase. The model was trained for 50 epochs with 128 batches. The following steps are carried out for the proposed approach.

BEGIN:

i. IMPORT necessary libraries and dataset
ii. PREPROCESS DATASET
    - Load handwritten Urdu digit and character images
    - Convert images to grayscale and normalize pixel values
    - Make subsets for training and testing
iii. DEFINE CNN Model
    - Input layer matching image dimensions
    - Convolutional layers with filters, kernel sizes, activation functions
    - Pooling layers for downsampling
    - Fully connected layers with activation functions
    - Output layer based on number of classes
iv. COMPILE MODEL
    - Loss function: Categorical Cross-Entropy
    - Optimizer: Adam
v. TRAIN MODEL
    - Feed training data in batches (batch size: 64)
    - Optimize parameters using backpropagation
    - Monitor training loss and accuracy
vi. EVALUATE MODEL
    - Test dataset used for evaluation
    - Compute accuracy and performance metrics
    - Save trained model
vii. ANALYZE RESULTS
    - Report overall accuracy and class-wise performance

END

3.9 CNN Working for Proposed Technique

CNN is a well-known DL model that has proven its potential in several classification tasks, particularly in image processing. CNN is primarily employed in image classification. In the proposed CNN-based OCR system for Urdu handwritten digits and characters, the architecture is structured with multiple layers to ensure effective feature extraction and classification. A total of 32 filters with 3×3 size are used in the first layer with a stride of 1, along with the ReLU. The following max-pooling layer has a size of 2×2 to manage spatial dimensions. The second layer contains 64 filters with a 3×3 kernel size, improving the ability to detect intricate Urdu script patterns. This layer is also followed by a 2×2 max-pooling operation to downsample the feature maps. The third convolutional layer further enhances feature extraction with 128 filters of size 3×3, enabling the model to capture detailed structural variations in handwritten Urdu characters.

A flatten layer is then used for the extracted features which are passed through two fully connected layers, with 256 and 128 neurons respectively, using ReLU activation to refine learned representations. The Softmax function determines the probability distribution over the character classes. The learning rate for training is 0.001, ensuring stable and efficient convergence. Training is done using 50 epochs with a batch size of 64, balancing training efficiency and performance. This structured CNN architecture ensures high accuracy in recognizing handwritten Urdu characters while effectively handling variations in writing styles and image quality. The custom structure of the CNN model enables it to generalize well across different handwriting samples, making it robust for real-world OCR applications. Depending on the requirements, CNN can have many layers [32]. The details of the model's layer are given here.

3.9.1 Convolutional Layer

In the proposed CNN-based OCR system for pure handwritten Urdu digits and characters, convolutional layers play a critical role in extracting relevant features. The filters in the layers help detect patterns or features in the input images, such as edges, curves, or shapes. The model may learn more complicated and abstract properties from the input photos by stacking many convolutional layers, allowing it to distinguish between distinct Urdu characters and numerals. The proposed OCR system uses convolutional layers to improve character and digit recognition accuracy by allowing the model to learn discriminative features from input images.

The convolution layer employs a matrix known as the kernel or filter, as shown in Fig. 10. This filter specifies the pattern to be recognized. The following filter is used to recognize the top horizontal edge. The character image's matrix representation is extracted. The image is pre-processed to remove noise and other unwanted data in the image. As part of the pre-processing, we can also reduce the image's dimensionality. The filter is multiplied by an image sub-matrix of the same size. The multiplication begins in the upper left corner [33]. Dot product multiplication is used instead of matrix multiplication. The dot product result is saved in the top left corner, and the filter is moved to the next sub-matrix. Every possible sub-matrix in the image is multiplied by the filter [34].

-1	-1	-1
1	1	1
0	0	0

Figure 10: Matrix example for CNN filter layer

3.9.2 Pooling Layer

Pooling layers downsample the feature maps by summarizing a rectangular neighborhood of the input image. This summarization can be achieved by max-pooling or average-pooling. Pooling layers reduce the number of parameters and control overfitting. It also makes the feature map invariant to small translations, rotations, and distortions in the input image [35].

3.9.3 Fully Connected Layer

A fully connected layer in a CNN is a traditional neural network layer that is used at the end of the CNN architecture to make the final prediction. The fully connected layer takes features from previous layers as

input, and the neurons in this layer are connected to all neurons in the previous layer. The role of this layer is to perform classification by using the features learned by previous layers and make the final prediction based on these features. The output of the fully connected layer is fed to an activation function, such as softmax, to produce the final prediction [33,35].

Multiple layers of perceptron are used to flatten the value. The inputs are multiplied by weights before being fed into the activation function. The majority of this activation function is ReLU. Negative values in the input are typically removed by this function. The definition of ReLU is generalized as follows.

$$F(a) = max(0, a) \tag{6}$$

According to the equation, if the input to ReLU is negative, it returns 0, otherwise it returns the input. This is how CNN works in a single iteration [35]. CNN typically goes through several iterations of this type. Epoch is the name given to these iterations [33]. The accuracy rate increases to a certain extent as the number of epochs increases. When the threshold is exceeded, the accuracy rate decreases.

4 Experiments and Results

For the CNN-based OCR system designed for purely handwritten Urdu digits and characters, comprehensive experiments were conducted to evaluate its performance and analyze potential sources of error. The model was implemented using Python with the TensorFlow DL framework, ensuring efficient training and evaluation. The dataset comprises handwritten Urdu digits and characters collected from over 900 individuals, providing a diverse range of handwriting styles. Each image had a resolution of 28 × 28 pixels, with over 700 samples available for each digit and character, allowing the model to generalize effectively across different variations in handwriting.

To optimize model training, the 'Adam' optimizer was employed ensuring stable convergence. The categorical cross-entropy loss function was used to minimize classification errors. The model was trained for 50 epochs with a batch size of 128, achieving a state-of-the-art accuracy of 98.3% and an F1 score of 88.6% on the testing set. A detailed analysis of the sources of error reveals certain limitations. Misclassification primarily occurred in characters with high structural similarity, such as closely resembling Urdu letters with minor stroke variations. Digits and characters with overlapping strokes or inconsistent spacing were also more prone to errors. Additionally, variations in handwriting styles, including different writing speeds, pen pressures, and ink inconsistencies, contributed to occasional misclassification.

Furthermore, noise in the dataset, such as blurred or faint strokes due to poor penmanship or scanning artifacts, impacted recognition accuracy. Some errors were attributed to the model's sensitivity to fine-grained textural details, where characters written with unconventional styles deviated from typical training samples. Despite these challenges, the model demonstrated strong resilience against common variations in handwriting and noise. The findings suggest that incorporating additional preprocessing techniques, such as advanced denoising algorithms and adaptive thresholding, could further enhance recognition performance. The proposed OCR system shows significant potential for real-world applications, including digitizing historical manuscripts, automating Urdu handwriting recognition, and facilitating digital archiving of handwritten documents.

4.1 Experimental Setup

The experiments are done on a workstation Intel Core i7 7th Generation processor and an NVIDIA GeForce. The proposed system is implemented using the Python programming language and the Keras DL library. For this research, we have used CNN tensor flow and Keras libraries have been implemented over the

Urdu handwritten characters and digits dataset using Google Colabs and Python programming language. The proposed OCR system is evaluated on a comprehensive dataset of pure handwritten Urdu digits and characters. The dataset consists of 1000+ images for each digit and character, split into a training set of 80% and a testing set of 20%.

4.2 Results for Character Recognition

Experiments are performed on the Urdu handwritten dataset for characters and numerals separately and results are reported for test sets only. Table 3 shows the accuracy results of character-wise recognition using the proposed approach. Contrary to previous research works where average accuracy is reported, character-wise accuracy is reported.

Table 3: Urdu handwritten characters experiment results

Urdu character	Character name	Accuracy	Urdu character	Character name	Accuracy
	Alif	73.79%		Alif Mad Aa	75.89%
	Baa	89.50%		Paa	93.06%
	Taa	93.06%		Aynn	95.80%
	Ghayn	96.65%		Faa	97.24%
	Jeem	97.88%		Hamza	96.50%
	Kaaf	98.12%		Chay	97.34%
	Daal	95.80%		Khaa	98.20%
	Gaaf	95.80%		Choti Yaa	98.90%
	Noon	98.40%		Laam	97.40%
	Meem	95.43%		Noon Gunnah	70.70%
	Sawaad	89.50%		Seen	98.60%
	Raa	76.70%		Zaal	97.00%

(Continued)

Table 3 (continued)

Urdu character	Character name	Accuracy	Urdu character	Character name	Accuracy
	Zawaad	98.40%		Tawaa	95.00%
	Rhraa	80.20%		Ttaa	94.00%

Results show that the accuracy of various characters varies significantly. It is observed that the accuracy of round-shaped characters is exceptional well surpassing 95% on average, compared to 'Alif' and 'Alif Mad Aa' where the accuracy is 73.79% and 75.89%, respectively. The proposed CNN can attain a 98.30% accuracy for handwritten character recognition.

4.3 Results for Numerals Recognition

Experimental results for digits detection using the proposed CNN model are depicted in Table 4 indicating super performance of the approach. Besides the handwritten digit seven, all digits have exceptionally good accuracy showing the potential of the approach for varied style digit detection.

Table 4: Urdu digit experiment results

Urdu character	Character name	Accuracy	Urdu character	Character name	Accuracy
	Zero	98.2%		One	93.45%
	Two	89.53%		Three	96.64%
	Four	97.60%		Five	95.80%
	Siz	97.88%		Seven	74.50%
	Eight	92.06%		Nine	94.75%

4.4 Performance Comparison with State-of-the-Art Approaches

A performance comparison with existing state-of-the-art approaches (SOTA) for Urdu handwritten character recognition is desirable to show the model's reliability. For this purpose, several recent works have been considered to analyze their accuracy against the proposed approach. Kashif et al. [13] used a ResNet18 model for Urdu font detection and reported a 94.4% accuracy. Similarly, KO and Porunan [14] reported 96.3% using the AlexNet mode and 94.7% accuracy using the GoogleNet model for Urdu language detection. Nastalique, a particular type of Urdu script is detected with 97.3% accuracy by Saqib et al. [16].

Along the same direction, Nasir et al. [17] and Naseer et al. [18] perform OCR for printed and calligraphic Urdu characters and reported 96% and 90% accuracy respectively. Misgar et al. [19] used recurrent neural network and LSTM for Urdu character and digits detection and reported a 96.9% accuracy for characters while the accuracy for digits identification is 85%. A similar recent work on the handwritten Arabic language is carried out by Hamdani et al. [20] who used a hybrid model comprising CNN, BiLSTM and CTC models reported accuracy of 93%, and 95%, for handwritten and printed Arabic characters. Cheema et al. [22] adopted ViLanOCR, a bilingual system to identify Urdu and English characters, and reported good results with a 1.1% CER.

Table 5 shows the comparative analysis of these SOTA works with the proposed approaches indicating the better results obtained in the current study. Along with the accuracy, limitations of SOTA works are also pointed out to better highlight the potential of the proposed model in this study. From extensive experimentation, it is observed that the proposed approach is better in handwritten character recognition in comparison to existing approaches.

Table 5: Comparison of detection accuracy with SOTA works

Ref.	Year	Model	Accuracy	Limitations
[13]	2021	ResNet18	94.4%	Limited to characters, lacks numeral recognition
[14]	2020	AlexNet, GoogleNet	96.3%	Focused on printed & semi-handwritten text
[16]	2019	–	97.3%	Ineffective for handwritten text
[17]	2021	Deep learning	96.00%	No handwritten evaluation
[18]	2022	NTL model	90.00%	Struggles with handwritten variations
[19]	2023	LSTM	96.90%	Needs better digit recognition
[20]	2024	CNN+BiLSTM+CTC	95.00%	Struggles with Urdu complexities
[22]	2024	ViLanOCR	1.1% CER	Requires fine-tuning for handwritten Urdu
[36]	2019	CNN-BiLSTM	83.69%	Lower accuracy, high computational complexity
[37]	2019	CNN	96.04%	High fine-tuning needed
[38]	2019	BiLSTM	94.00%	Focuses on characters only
[39]	2022	CNN-RNN	94.72%	Urdu digits are not covered and model is computationally complex
Proposed	2024	CNN	98.30%	F1 score needs improvements

4.5 Discussion

The results show the effectiveness of the proposed CNN-based OCR system for Urdu handwritten characters and digit recognition. The high accuracy of 98.3% and an F1 score of 88.6% demonstrate the model's capability to handle the inherent complexities of Urdu script, including variations in stroke thickness, writing styles, and structural differences between similar-looking characters. The recall score of 0.65 indicates that while the model successfully identifies a significant portion of handwritten Urdu characters and digits, there is room for improvement in capturing less frequently occurring patterns. This can be attributed to the intricate nature of Urdu handwriting, where subtle differences in character formation can lead to misclassification. The model's strong precision score of 0.89 further highlights its ability to make

highly accurate predictions, minimizing the likelihood of false positives, which is particularly crucial for applications in document recognition and automated data entry.

The confusion matrix shows correct and wrong predictions from the model. The matrix reveals that certain Urdu characters and digits, especially those with similar structural patterns, are occasionally misclassified. For instance, handwritten digits such as "2" and "3" or characters like (Bay) and (Tay) can exhibit minimal visual differences, leading to occasional misidentifications. The presence of such errors underscores the challenges posed by handwritten script, where individual writing styles, pen pressure, and ink spread can significantly alter character appearance. Additionally, external noise factors, such as smudges or incomplete strokes, contribute to these classification errors. Despite these challenges, the model exhibits a strong ability to generalize across diverse handwriting styles, thanks to the comprehensive dataset used for training, which includes over 900 contributors with varied writing habits. Fig. 11 indicates that the model has very few wrong predictions.

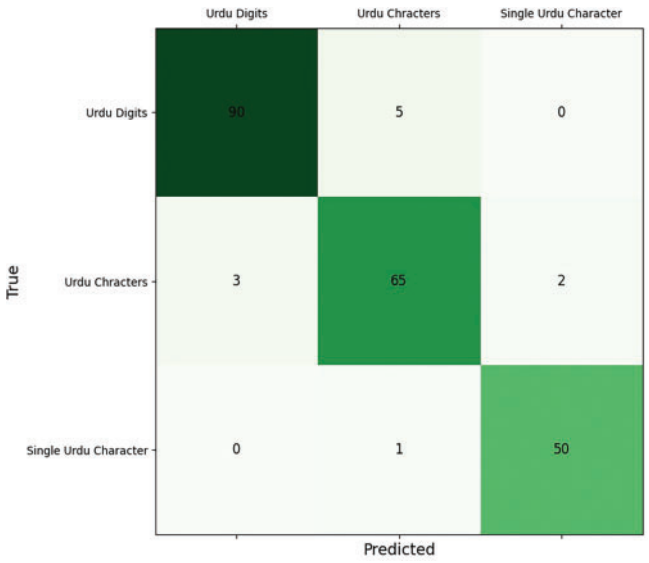

Figure 11: Confusion matrix for the proposed model

The physical reasoning behind the model's success lies in the effectiveness of convolutional layers in capturing key spatial features of Urdu handwritten script. The CNN architecture allows for hierarchical feature extraction, where lower layers detect basic edges and shapes, while deeper layers learn complex structural patterns unique to Urdu characters and digits. The application of max pooling further enhances feature representation by reducing redundant information and focusing on the most salient aspects of character formation. The use of the Softmax activation function in the final layer ensures precise probability distribution among character classes, optimizing classification performance. The robust performance of the proposed system suggests its potential applicability in real-world scenarios such as postal services, banking automation, and historical manuscript digitization. Future improvements, such as integrating attention mechanisms and expanding the dataset with additional handwriting variations, could further enhance the model's performance and adaptability.

4.6 Limitations

While the proposed Urdu handwritten OCR system demonstrates high accuracy and robustness, several limitations must be acknowledged. First, the model relies on a predefined dataset, which may not fully capture the diversity of real-world handwriting variations, including different writing styles, ink intensities, and paper textures. Additionally, the preprocessing techniques, such as binarization and noise removal, may not perform optimally on highly degraded or low-resolution images, potentially impacting recognition accuracy. Another limitation is the misclassification of visually similar characters due to overlapping strokes or inconsistent handwriting, which remains a challenge in Urdu script recognition. Furthermore, while the model integrates HOG and LBP features to enhance recognition, it may not generalize well to unseen handwriting patterns outside the dataset. Lastly, the study does not extensively explore real-time deployment challenges, such as computational efficiency on mobile or embedded devices. Future work should focus on addressing these limitations to improve practical applicability.

5 Conclusion and Future Work

This study introduced a CNN-based OCR system for pure handwritten Urdu numerals and characters, which achieved superior performance on the testing set with an accuracy of more than 98.3% with an F1 score of 88.6%. Results reveal that the suggested approach can effectively recognize the complexity and diversity of Urdu script, as well as its robustness to variations in handwriting styles and noise. The proposed system could be used in a variety of practical applications, such as digit recognition in banking, character recognition in postal services, and text recognition in document analysis. Furthermore, the proposed method can be expanded to other languages with comparable scripts, helping to design more accurate and efficient OCR systems. The success of this study emphasizes the relevance of deep learning approaches in tackling difficult problems in the field of image recognition, and it shows that additional research in this area could lead to considerable advances in OCR systems for languages with complicated and unique scripts. For future work, the proposed OCR system can be enhanced by integrating attention mechanisms to focus on key features, incorporating transformer-based architectures for improved contextual understanding, and expanding the dataset with diverse handwriting styles. Additionally, real-time deployment on mobile devices, multi-script recognition, self-supervised learning, and domain adaptation techniques can further improve accuracy, robustness, and practical usability.

Acknowledgement: The authors are thankful to the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R192), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R192), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Syed Atir Raza: conceptualization, data curation, wiring—the original manuscript; Muhammd Shoaib Farooq: formal analysis, conceptualization, wiring—the original manuscript; Uzma Farooq: methodology, data curation, formal analysis; Hanen Karamti: funding acquisition, visualization, investigation; Tahir Khurshaid: software, investigation, project administration; Imran Ashraf: supervision, validation, writing—review and editing. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The dataset used in this study is publicly available on the following link. https://www.kaggle.com/datasets/hazrat/uhat-urdu-handwritten-text-dataset (accessed on 15 May 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Cheng LS, Kramer MA, Upreti R, Namboodiripad S. Moving past indirect proxies for language experience: 'Native speaker' and residential history are poor predictors of language behavior. Proc Annual Meet Cognit Sci Soc. 2022;44:1682–9.
2. Qazi MH, Javid CZ, Ullah I. Representation of indigenous languages employing a religious screen for the discursive construction of students' postcolonial national identities: a curious case of 'internal colonisation' and 'cultural invasion' in Pakistani schools. British Educat Res J. 2023;49(1):35–52. doi:10.1002/berj.3827.
3. Chirimilla R, Vardhan V. A survey of optical character recognition techniques on indic script. ECS Trans. 2022;107(1):6507–14. doi:10.1149/10701.6507ecst.
4. Srivastava S, Verma A, Sharma S. Optical character recognition techniques: a review. In: 2022 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS); 2022 Feb 19–20; Bhopal, India. p. 1–6. doi:10.1109/SCEECS54111.2022.9740911.
5. Li Y. Research and application of deep learning in image recognition. In: 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA); 2022 Jan 21–23; Shenyang, China. p. 994–9.
6. Sun X, Liu Y, Xie L, He X, Ma X. Single check method of relay protection fixed value based on OCR image recognition. In: 2022 Power System and Green Energy Conference (PSGEC); 2022 Aug 25–27; Shanghai, China. p. 1101–5.
7. Amudha J, Thakur MS, Shrivastava A, Gupta S, Gupta D, Sharma K. Wild OCR: deep learning architecture for text recognition in images. In: Proceedings of International Conference on Computing and Communication Networks: ICCCN 2021; 2021 Jul 19–22; Athens, Greece. p. 499–506.
8. Nikoghosyan K. Optical character recognition (OCR) using tensorflow. In: Main trends in the development of science in the XXI century. Current problems of the world scientific space Russia federation (Elibrary.ru); 2022. p. 35–41.
9. Verma P, Foomani G. Improvement in OCR technologies in postal industry using CNN-RNN architecture: literature review. Int J Mach Learn Comput. 2022;12(5). doi:10.18178/ijmlc.2022.12.5.1095.
10. Albattah W, Albahli S. Intelligent arabic handwriting recognition using different standalone and hybrid CNN architectures. Appl Sci. 2022;12(19):10155. doi:10.3390/app121910155.
11. Ghosh MMA, Maghari AY. A comparative study on handwriting digit recognition using neural networks. In: 2017 International Conference on Promising Electronic Technologies (ICPET); 2017 Oct 16–17; Deir El-Balah, Palestine. p. 77–81.
12. Arafat SY, Ashraf N, Iqbal MJ, Ahmad I, Khan S, Rodrigues JJ. Urdu signboard detection and recognition using deep learning. Multimed Tools Appl. 2022;81(9):11965–87. doi:10.1007/s11042-020-10175-2.
13. Kashif M. Urdu handwritten text recognition using ResNet18. arXiv:2103.05105. 2021.
14. KO MA, Poruran S. OCR-nets: variants of pre-trained CNN for Urdu handwritten character recognition via transfer learning. Procedia Comput Sci. 2020;171(2):2294–301. doi:10.1016/j.procs.2020.04.248.
15. Rizvi SSR, Khan MA, Abbas S, Asadullah M, Anwer N, Fatima A. Deep extreme learning machine-based optical character recognition system for nastalique Urdu-like script languages. Comput J. 2022;65(2):331–44. doi:10.1093/comjnl/bxaa042.
16. Rizvi S, Sagheer A, Adnan K, Muhammad A. Optical character recognition system for Nastalique Urdu-like script languages using supervised learning. Int J Pattern Recognit Artif Intell. 2019;33(10):1953004. doi:10.1142/s0218001419530045.
17. Nasir T, Malik MK, Shahzad K. MMU-OCR-21: towards end-to-end Urdu text recognition using deep learning. IEEE Access. 2021;9:124945–62. doi:10.1109/access.2021.3110787.
18. Naseer A, Hussain S, Zafar K, Khan A. A novel normal to tangent line (NTL) algorithm for scale invariant feature extraction for Urdu OCR. Int J Document Anal Recognit (IJDAR). 2022;25(1):51–66. doi:10.1007/s10032-021-00389-x.
19. Misgar MM, Mushtaq F, Khurana SS, Kumar M. Recognition of offline handwritten Urdu characters using RNN and LSTM models. Multim Tools Applicat. 2023;82(2):2053–76. doi:10.1007/s11042-022-13320-1.

20. Mosbah L, Moalla I, Hamdani TM, Neji B, Beyrouthy T, Alimi AM. ADOCRNet: a deep learning OCR for Arabic documents recognition. IEEE Access. 2024;12(1):55620–31. doi:10.1109/access.2024.3379530.
21. Nabi ST, Kumar M, Singh P. DeepNet-WI: a deep-net model for offline Urdu writer identification. Evolving Systems. 2024;15(3):759–69. doi:10.1007/s12530-023-09504-1.
22. Cheema MDA, Shaiq MD, Mirza F, Kamal A, Naeem MA. Adapting multilingual vision language transformers for low-resource Urdu optical character recognition (OCR). PeerJ Comput Sci. 2024;10(6):e1964. doi:10.7717/peerj-cs.1964.
23. Hamza A, Ren S, Saeed U. ET-Network: a novel efficient transformer deep learning model for automated Urdu handwritten text recognition. PLoS One. 2024;19(5):e0302590. doi:10.1371/journal.pone.0302590.
24. Mustafa A, Rafique MT, Baig MI, Sajid H, Khan MJ, Kallu KD. A permuted autoregressive approach to word-level recognition for Urdu digital text. arXiv:2408.15119. 2024.
25. Malik MSI, Nawaz A, Jamjoom MM. Hate speech and target community detection in nastaliq urdu using transfer learning techniques. IEEE Access. 2024;12(2):116875–90. doi:10.1109/access.2024.3444188.
26. Abolghasemi V, Falahati R. L2 handwritten assignments for automated writing evaluation: a text recognition study. In: Technology in second language writing. Oxfordshire, UK: Routledge; 2023. p. 152–67.
27. Pande SD, Jadhav PP, Joshi R, Sawant AD, Muddebihalkar V, Rathod S, et al. Digitization of handwritten Devanagari text using CNN transfer learning-A better customer service support. Neurosci Informat. 2022;2(3):100016. doi:10.1016/j.neuri.2021.100016.
28. Sankara Babu B, Nalajala S, Sarada K, Muniraju Naidu V, Yamsani N, Saikumar K. Machine learning based online handwritten Telugu letters recognition for different domains. In: A fusion of artificial intelligence and internet of things for emerging cyber systems. Cham, Switzerland: Springer; 2022. p. 227–41. doi:10.1007/978-3-030-76653-5_12.
29. Chaudhuri A, Mandaviya K, Badelia P, Ghosh SK. Optical character recognition systems. 1st ed. Cham, Switzerland: Springer; 2017.
30. Thorat C, Bhat A, Sawant P, Bartakke I, Shirsath S. A detailed review on text extraction using optical character recognition. In: ICT analysis and applications. Singapore: Springer; 2022. p. 719–28. doi:10.1007/978-981-16-5655-2_69.
31. Ali H. UHaT: urdu handwritten text dataset; 2020 [Internet]. [cited 2025 May 15]. Available from: https://www.kaggle.com/datasets/hazrat/uhat-urdu-handwritten-text-dataset.
32. Yao Y, Zhang S, Yang S, Gui G. Learning attention representation with a multi-scale CNN for gear fault diagnosis under different working conditions. Sensors. 2020;20(4):1233. doi:10.3390/s20041233.
33. Li Y, Hao Z, Lei H. Review of convolutional neural networks. Comp Appl. 2016;36:2508–15.
34. Li Y, Gu S, Gool LV, Timofte R. Learning filter basis for convolutional neural network compression. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision; 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 5623–32.
35. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET); 2017 Aug 21–23; Antalya, Turkey. p. 1–6.
36. Hassan S, Irfan A, Mirza A, Siddiqi I. Cursive handwritten text recognition using bi-directional LSTMs: a case study on Urdu handwriting. In: 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML); 2019 Aug 26–28; Istanbul, Turkey. p. 67–72.
37. Husnain M, Saad Missen MM, Mumtaz S, Jhanidr MZ, Coustaty M, Muzzamil Luqman M, et al. Recognition of Urdu handwritten characters using convolutional neural network. Appl Sci. 2019;9(13):2758. doi:10.3390/app9132758.
38. Ahmed SB, Naz S, Swati S, Razzak MI. Handwritten Urdu character recognition using one-dimensional BLSTM classifier. Neural Comput Applicat. 2019;31(6):1143–51. doi:10.1007/s00521-017-3146-x.
39. ul Sehr Zia N, Naeem MF, Raza SMK, Khan MM, Ul-Hasan A, Shafait F. A convolutional recursive deep architecture for unconstrained Urdu handwriting recognition. Neural Comput Appl. 2022;34(2):1635–48. doi:10.1007/s00521-021-06498-2.