International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Computationally Efficient Holistic Approach for Handwritten Urdu Recognition using LRCN Model.

Aejaz Farooq Ganai[1]*, Farida Khursheed[2]

Submitted: 14/11/2022 **Accepted**: 22/02/2023

Abstract: Handwritten Urdu text is difficult to recognize because it poses several challenges such as writer-dependent variations in producing different ligature shapes, irregular positioning of diacritics, and similarity in shape of some Urdu characters in writing. Moreover, the formulation and labeling of a huge database of handwritten Urdu is also a challenging task. Due to these challenges, the handwritten Urdu OCR remained the least explored to date. The goal of a writer's adaptive handwritten recognition system is to build a model that improves the recognition of a generic recognition system for a specific author. Few researchers propose the handwritten Urdudatasets but only UNHD is publicly available. Although UNHD claims to have 10000 text lines, only 700 of those lines are unique in terms of semantic content. Moreover, the UNHD contains ligatures of length only up to five characters and doesn't cover the entire Urduligature corpus. Also, the handwritten Urdu recognition techniques proposed by most studies focus on segmentation-based (implicit or explicit) approaches and have not obtained a significant recognition rate on benchmark datasets. Hence an enriched database and an exhaustive recognition technique are needed that can recognize any unconstrained handwritten Urdu text. In this research, we have proposed a new robust holistic handwritten Urdu text recognition technique evaluated on our new handwritten Urdu database UHLD (Urdu handwritten Ligature Dataset). In our proposed technique, CNN extracts features from UNHD and UHLD datasets, and these features train the multidimensional LSTM based recurrent neural network for the classification and recognition of handwritten Urdu text.The technique is computationally efficient and achieved a remarkable accuracy recognition rate of 94.2 % for UNHD and 96.6% for UHLD datasets.

Keywords: *Handwritten Urdu, Optical Character Recognition, Urdu Nastaliq Handwritten Dataset, Convolutional Neural Networks*

1. Introduction

Urdu Language is a very famous language and is widely spoken and written in most countries like Bangladesh, Afghanistan, UAE, Guyana, Pakistan, and the Indian subcontinents [1,2]. Recognitionof Urdu has gained much importance due to the enormous applications of the Optical Character Recognition (OCR) tool. Urdu OCR has been exhaustively explored for the last few years after the emergence of various machine learning and deep learning techniques, to develop the Urdu OCR tool. The exploration of printed Urdu OCR has achieved some remarkable recognition rates with these deep learning algorithms, but hand written Urdu has been least explored because of several challenges. These challenges include the unavailability of a standard handwritten Urdu dataset, huge variation among writing styles of different Urdu writers, the cursive nature of the script, the similarity in shape of individual handwritten characters, and many more difficulties and complexities [3] in Urdu writing styles.

Only a few researchers have contributed significant achievements toward the development of handwritten Urdu OCR. Furthermore, for printedUrdu text, there are commercially available Urdu OCR systems [4], but there is no OCR system available for recognition of handwritten Urdu to date.

1.1. Urdu Script

Urdu characters/words are written from right to left and Urdu numerals are written from left to right [Table 1], hence Urdu is bidirectional in nature. The handwritten Urdu consists of 39 basicletters [Table 1]. These letters are divided into two classes of character sets, joiner and non-joiner ["و ,ز ,ر ,ذ ,ڈ ,د,ں ,ا, and ے"] character set. Joiners characters of handwritten Urdu can not only be written in four different shapes namely initial, middle, final, and isolated [5], as presented in Fig.1, but the number of possibleshapes of some characters may even approach to 60 [3]. Urdu ligature is formed by joining one or more letters that give a meaningful word [Fig.1].

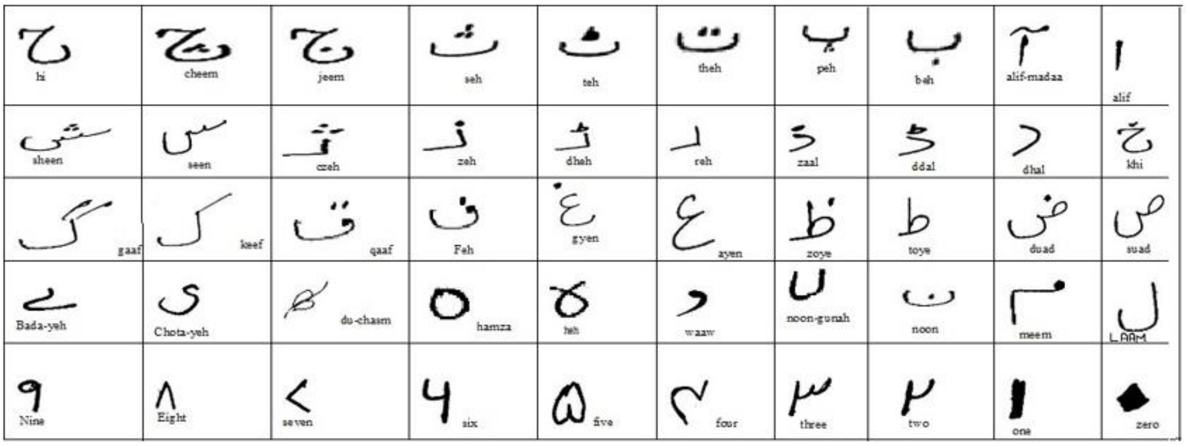
1.2. Challenges in Handwritten Urdu Recognition

The recognition of handwritten cursive scripts is a challenging task. The various challenges encountered in the recognition of handwritten Urdu script are context

[1]*Electronics & Communication Engineering Department, National Institute of Technology Srinagar, J&K India-190006 ORCID ID: 0000-0002-5418-5042,*
Email: gaejaz@gmail.com
[2]*Electronics & Communication Engineering Department, National Institute of Technology Srinagar, J&K India-190006, ORCID ID: 0000-0001-5087-3254*

sensitivity of characters at different positions in a ligature (Fig.1-a), the diagonal nature of Urdu script (Fig. 1-b), ligature overlapping (Fig.1-b) and uneven spacing between ligatures. Besides these challenges [3], handwritten Urdu poses huge variations among the writing styles of different Urdu writers that make researchers difficult to proposea generalized handwritten Urdu dataset. Also, some letters in handwritten Urdu resemble shapes while writing Urdu [Table 1] which enhances the difficulty of recognition. Due to the existenceof these stated challenges, handwritten Urdu has been least explored and hence this research mainly emphasizes on recognition of handwritten Urdu.

Table 1. Basic handwritten Urdu letters and numerals.



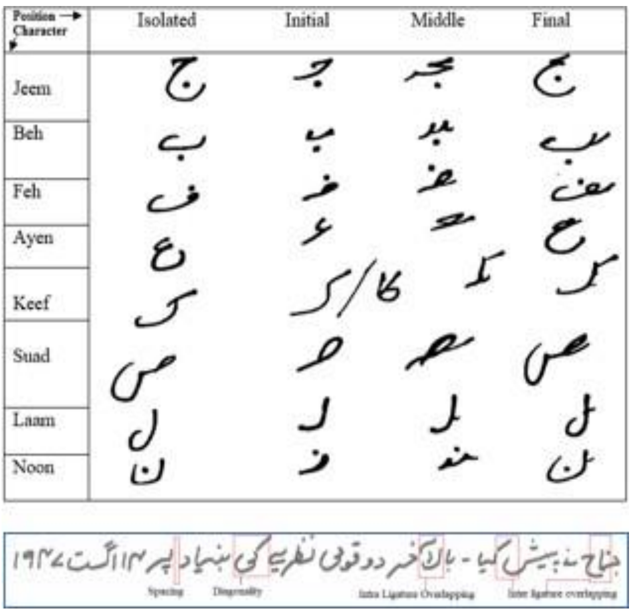


Fig. 1. Challenges in handwritten Urdu recognition **a)** context sensitivityof character ب at different positions (upper) **b)** cursiveness, diagonality, spacing and ligature overlapping (lower)

We are proposing a new holistic approach to ligature recognition of handwritten Urdu using the LRCN model (LR for **L**STM **R**ecurrent and CN for **C**onvolutional **N**etwork). The proposed work relies on a holistic approach where in features are extracted from handwritten Urdu ligature images using Convolutional Neural Networks (**CN**) and these features train multi-dimensional LSTM recurrent neural network (**LR**) [hence name LRCN Model]for classification and recognition of unconstrained handwritten Urdu text. The major motivation behind the idea is that the segmentation of handwritten Urdu ligatures into individual characters is a difficult and challenging task due to overlapping nature of cursive script, and the huge variation in position of diacritics associated with a character in a handwritten ligature. So, recognizing a complete ligature doesn't require segmentation of ligature into characters, which motivates us to execute a holistic approach for recognition of handwritten Urdu text. However, it isvery time consuming to count the number of characters needed for manual classification of the base class of a ligature, hence an algorithm that estimates the number of characters in a ligature hasbeen proposed in this research. The algorithm saves a huge amountof time needed for ligature class construction. In nutshell, our proposed research encompasses the following objectives.

1) The research proposes a new LRCN model for recognition of complete handwritten Urdu ligatures/words up to seven-characterlength.
2) The research proposes a new algorithm that estimates the number of characters in a ligature.
3) The proposed approach is computationally efficient as it takes a small test time complexity of O (1) only, independent of the number of cluster groups used.
4) The proposed approach is transformation invariant i.e.; the model recognizes ligatures irrespective of rotation/shift/scale.

2. Related Work

Recognition methods for cursive writings are generally classified into analytical and holistic approaches. Analytical or segmentation-based approach employ characters as recognition units and these approaches are divided into either implicit segmentation [6,7] or explicit segmentation [8]. Segmenting Urduscript into individual characters is extremely challenging task [3]. Holistic or segmentation free techniques employ partial words or Ligatures (the longest connected component of a word) as units ofrecognition [9]. These approaches are based on recognition of complete ligature that do not require character level segmentation [10], but the number of unique ligature classes to be recognized amounts to be a huge number in case of holistic approaches. Eventhough, Urdu has a total of around 26000 ligatures [10], but most of the ligatures are rarely used. Moreover, many ligatures in Urdushare a common shape and differentiate only in number and position of diacritics. Hence considering only frequent ligatures, the number of classes reduces significantly to only a few thousandunique ligatures that cover almost 99% of Urdu corpus [10].

The exploration of handwritten Urdu started in 2014 by *Saad et al*[11]. The authors propose a new dataset for the Urdu language written in the Nastaliq writing style called UCOM Dataset. The UCOM dataset comprises 6400 text lines written by 100 writers. The authors extend their database up to 300 writers and labeled itwith a unique identity of users. Later in 2018 [12], this dataset was extended to 500 writers and renamed as Urdu Nasta'liq Handwritten Dataset [UNHD]. The authors performed training and classification using a single-dimensional BLSTM classifier and realized an approximate error of 6.04–7.93 %. In early 2019, *Mujtaba et al* [13] employed CNN for the recognition and classification of handwritten Urdu characters. These authors developed a new dataset of handwritten Urdu characters andnumerals and claimed accuracy of 96% on character classification and 98% on numeral classification. Since their work remains confined to the character and numeral recognition only, hence it can't be applied to real-world handwritten Urdu text recognition. In mid-2019, *Hassan et al* [14] developed the Urdu handwritten database of 6000 text lines (not publicly available) and employed seven convolutional layers for feature extraction and 2 BLSTM layers for classification and recognition of Urdu characters. Although the authors explored their work on handwritten characterrecognition from several thousand text lines but the recognition rate obtained was only 83%. In December 2019, *Saad et al* [15] proposed another approach to exploiting the transfer learning experience of similar patterns on handwritten Urdu text examination. The authors employed an MNIST pre-trained network by transferring its learning experience to UNHD samples. The features were extracted using CNN and experimental evaluation was performed using deep multidimensional long short term (MDLSTM) memory networks. Although the authors report a good character accuracy of approximately 93%, their work remains confined to the UNHD database, which doesn't cover thesignificant variation with other datasets and their work doesn't explore recognition of six and seven character ligatures.

After an exhaustive literature survey, it has been revealed that ligature recognition for handwritten Urdu has been least exploredto date and efforts have been made mostly on character recognition (with or without segmentation) of handwritten Urdu with the highest 92% recognition rate. Hence in this research, an attempt has been made for complete ligature recognition of handwritten Urdu. Also, several accuracy parameters such as precision, sensitivity, specificity, F1-Score, ROC curve, and accuracy have been evaluated for different length ligature cluster classes. To determine the impact of various network parameters on system performance, the training accuracy has also been evaluated by varying these parameters to determine the impact on recognition rates. The elaborated methodology to achieve all these objectives has been discussed in the next section.

3. Proposed work

The proposed methodology used for the recognition of handwritten Urdu text has been explained in this section. The methodology comprises the handwritten Urdu dataset collection, the pre-processing of the dataset, the feature extraction phase to train the model, and recognition of handwritten Urdu text employing this trained model. The various phases of methodology are shown in Fig.2 and discussed in the following details.

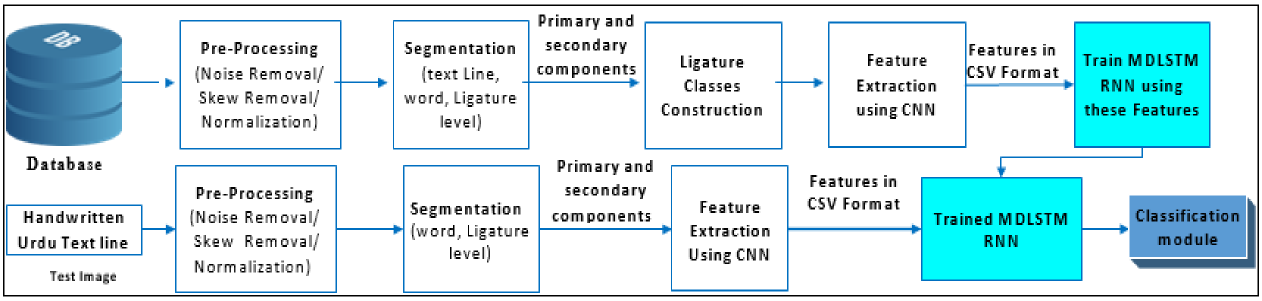


Fig. 2. Proposed Methodology

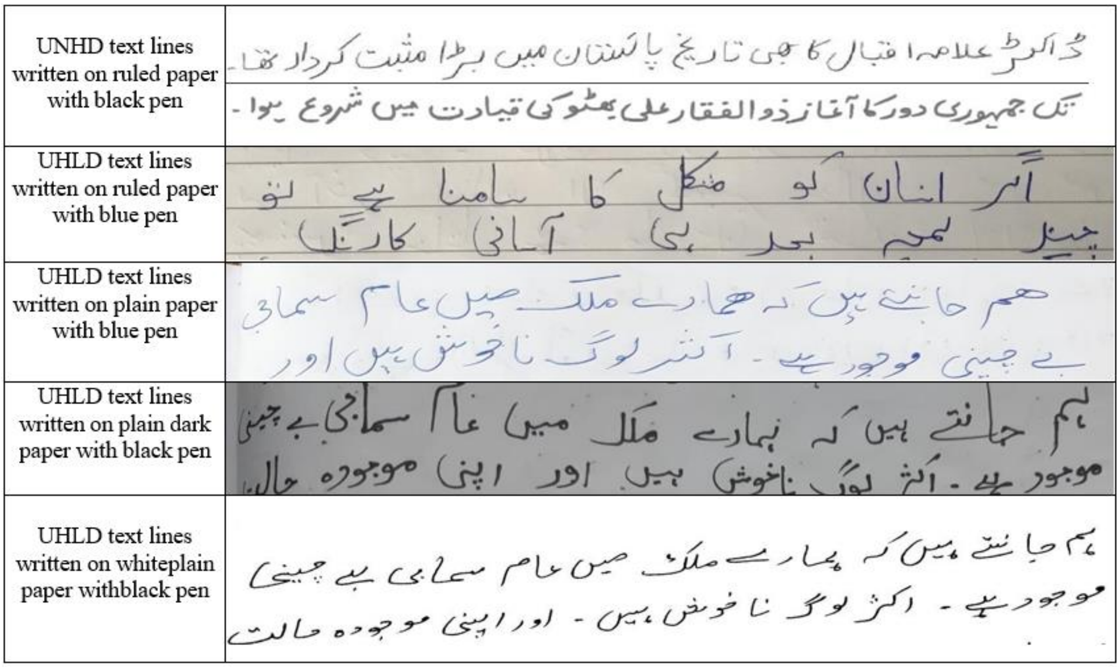
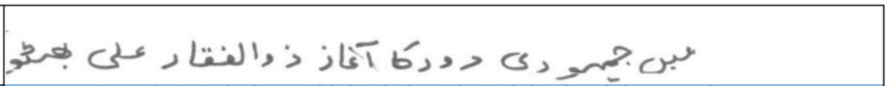
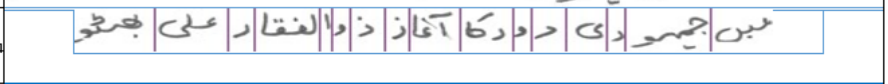
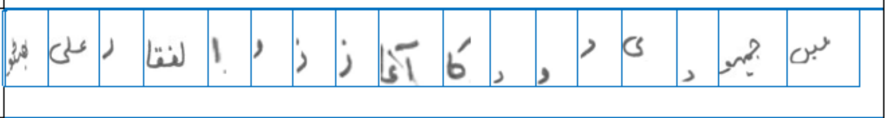
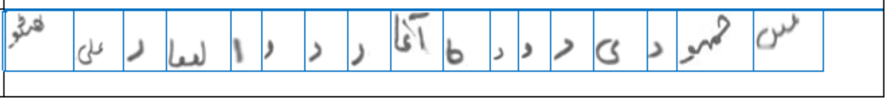


Fig. 3: Few Handwritten Urdu text lines (written by different writers) extracted from UNHD and UHLD datasets

Table 2. Ligature extraction from handwritten Urdu text line

Handwritten Urdu text line	
Application of [17] on handwritten Urdu text.	
Result of [17] (ligatures obtained after Segmentation)	
Primary ligatures extracted using algorithm [19].	

3.1. Dataset Acquisition

Some researchers have proposed their datasets while working on handwritten Urdu text recognition. *Saad Bin Ahmad et al* [12] proposed Urdu Nastaliq Handwritten Dataset" UNHD" (publicly available). *Shahbaz Hassan* et al [14] also proposed their handwritten Urdu dataset which is not publicly available. We have acquired the UNHD database and analyzed that, though the database is large enough to cover significant variation amongst different writers, the dataset comprises ligatures of only up to five-character length and there are not an adequate number of handwritten Urdu ligatures that can cover most of the Urdu ligature corpus. Although UNHD claims to have 10000 text lines, only 700 of those lines are unique in terms of semantic content. Hence in our proposed work, we have used our proposed handwritten Urdu dataset called "Urdu Handwritten Ligature Dataset [UHLD]" for experimental evaluation. The UHLD consists of 6000 text lines written by 200 writers, with ligatures up to seven-character length while it was only up to five-character ligature length in the benchmark dataset. The UHLD is written independently of paper color, paper type (blank or ruled), ink color, and pen type. Few text lines of dataset UNHD [14] and ourproposed dataset UHLD are shown in Fig. 3.

3.2. Pre-Processing the dataset

Pre-Processing is the initial phase in dataset cleaning. The transformation of Urdu datasets into a noise-free uniform format is necessary to normalize the database. The main steps involved in preprocessing include de-skewing of handwritten Urdu pages, noise reduction, evenness, and normalization of images. The UNHD [14] dataset was acquired as separate handwritten Urdu text line images without skew. The type of noise visible from these handwritten images was usually structural noise that consists of superimposed objects such as lines, marks, etc. We have removed this structural noise using a deep learning process [16].

3.3. Segmentation

Segmentation is the fundamental step in every text recognition system. The handwritten Urdu database collected from UNHD [14] and UHLD datasets is in the form of separate text lines. Eachtext line is divided into words and each word has been segmentedinto ligatures using the algorithm proposed by Rehman et al [17]. A ligature is the longest connected component in a word that can be separated from another ligature using [17] as shown in Table 2.However, it must be noted that segmentation of a ligature into characters is a difficult and challenging task [3] for handwritten Urdu, as characters vary their shapes depending upon their position in a word (context sensitive nature of Urdu script [3]) and irregular position of diacritics associated to a character. Hence segmentation is performed up to ligature level only and ligature forms the basis of recognition in the proposed recognition system. Before formation of classes of different length ligatures, the diacritics associated with a ligature need to be separated as it has been analysed that most of the Urdu ligatures extracted resemble basic shapes and are differentiated in only the number and positionof diacritics [18]. Hence to reduce the number of classes of ligature clusters, the diacritics associated with ligatures are to be separated. To perform this diacritic separation, we have used a connected component labelling algorithm [19] that separates primary ligatures from secondary components as shown in Table 2. The main idea behind this separation is that the primary ligatures whichrepresent the main body of the ligature are naturally larger in sizeas compared to secondary ligatures which represent dots and diacritics. A ligature is identified as a primary or secondary component using statistical features capturing the shape and position (with respect to baseline) information of the component. Table 2 shows the extraction of ligatures from a handwritten Urdu text line. The 'connected component labelling' algorithm doesn't yield great results while separating primary and secondary components of handwritten Urdu ligature as clearly seen in Table2, but for a huge handwritten Urdu dataset, the results are reasonable.

3.4 Ligature Class construction

The primary ligatures extracted after segmentation are input to ourproposed algorithm (Algorithm-I shown in Table 3)) for determining the no. of characters in a ligature, that forms the basisof different length ligature class construction

Table 3: Algorithm-I that estimates the no. of characters in completeligature.

Algorithm-I: Algorithm for determining the numberof characters in the ligature.
Input: Complete Ligatures In
Begin: *For each I in I_n* 1. *Plot vertical intensity profiles* 2. *Count the number of maximum value points P_m=MAXV(I).* 3. *Estimated number of characters in ligature= Pmn=n+1* 4. *End For*
Output: Number of characters in the ligature.

The Algorithm-I shown in Table 3, classifies the ligature into a particular cluster depending upon the number of characters in the ligature. The Algorithm-I saves a lot of time, needed for manual identification of the number of characters in a ligature followed byclassifying this ligature to a particular cluster/group. Using this algorithm, we have created 1500 unique ligature cluster groups (including secondary ligature cluster) in this study and each group contains more or less 100 samples of the ligature class as shown in Table 4. these ligature cluster groups are along with their lengthare shown in Table 5.

Table 4. Ligature class construction from both datasets.

Dataset	No. oftext lines	Approx. no. ofligatures extracted	No. of Ligature Classes constructed
UNHD	5000	50000	500
UHLD	6000	100000	1000

3.5 Normalization

Before applying the feature extraction phase, all the ligatures are normalized because the styles of writers differ concerning skew, slant, height and width of ligatures. As shown in Table 2, the extracted ligatures are not clear and contain noise and gaps within strokes that are caused by loss of data. A noise filtering operation is performed on ligatures using a coordinate logic filter [20] thatfills these gaps within strokes. The process of 'Thresholding' [21] is applied to the ligatures to eliminate speckle noise. After the denoising process, the images of ligatures are normalized using 'linear scaling' [22] because the features are more or less uniformly distributed across a fixed range. All the ligature images are converted to a 28×28 image dataset using 'image rescaling' [22] but keeping the information and aspect ratio intact. This process uniforms all the ligature images to a constant and fixed size and makes the feature extraction process feasible. Thinning process [23] is also performed for ligatures that are considerably thicker. The ligatures obtained after normalization are shown in Fig. 4.

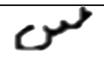
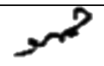
Ligatures extracted from segmentation phase	Normalized Ligature
	
	

Fig. 4. Normalization

Ligature group	Size of cluster
Single Character Ligature	20
Two Character Ligature	390
Three Character Ligature	320
Four Character Ligature	310
Five Character Ligature	210
Six Character Ligature	150
Seven Character Ligature	78
Urdu Numerals	10
Secondary Ligatures (Diacritics)	12
Total Ligature Dataset	**1500**

Table 5. Ligature cluster groups formed and their length

3.6 Feature Extraction

The Convolutional Neural Network (CNN) [24] is used for extracting features from the ligature images. The CNN makes use

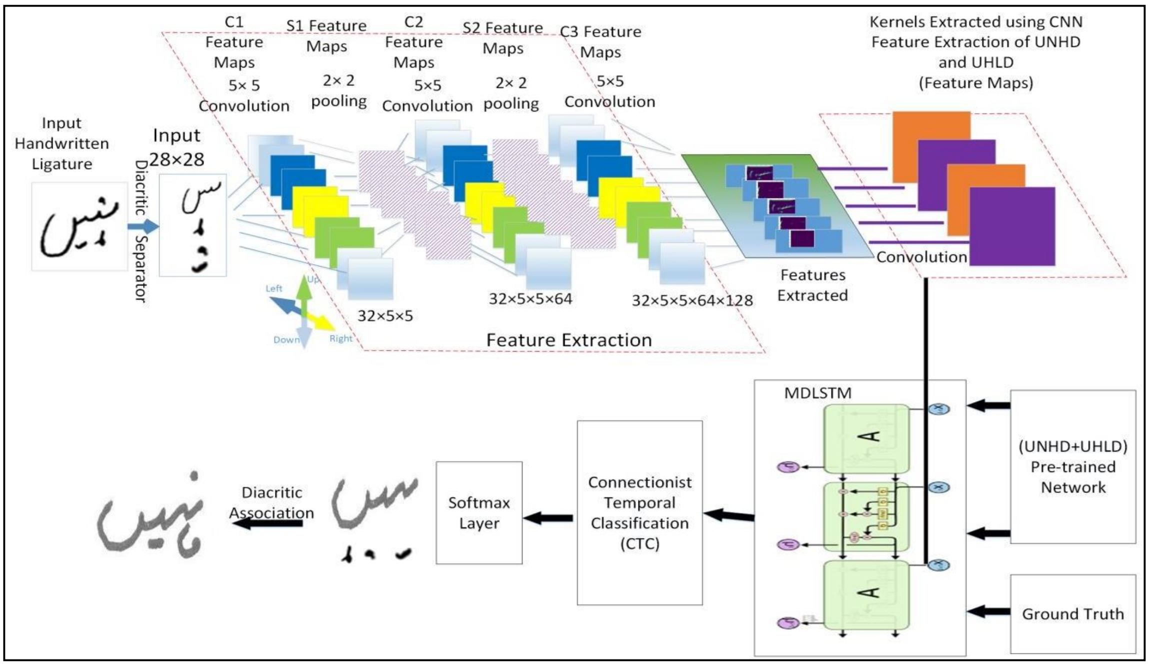


Fig. 5. Feature Extraction using CNN, training and classification using MDLSTM RNN for handwritten Urdu ligatures.

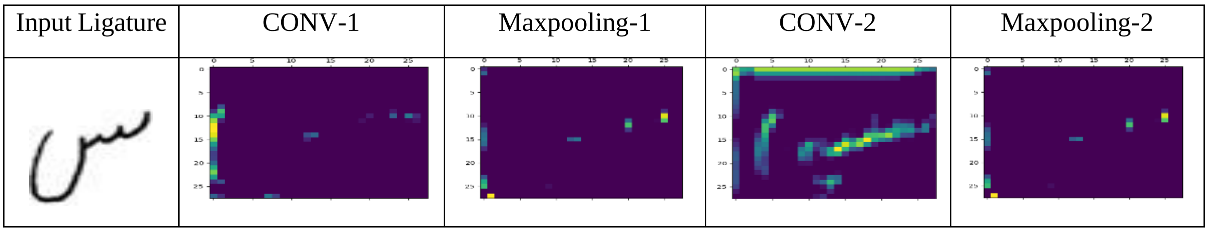


Fig. 6. Fourth channel of the activation in extracting features of given shape.

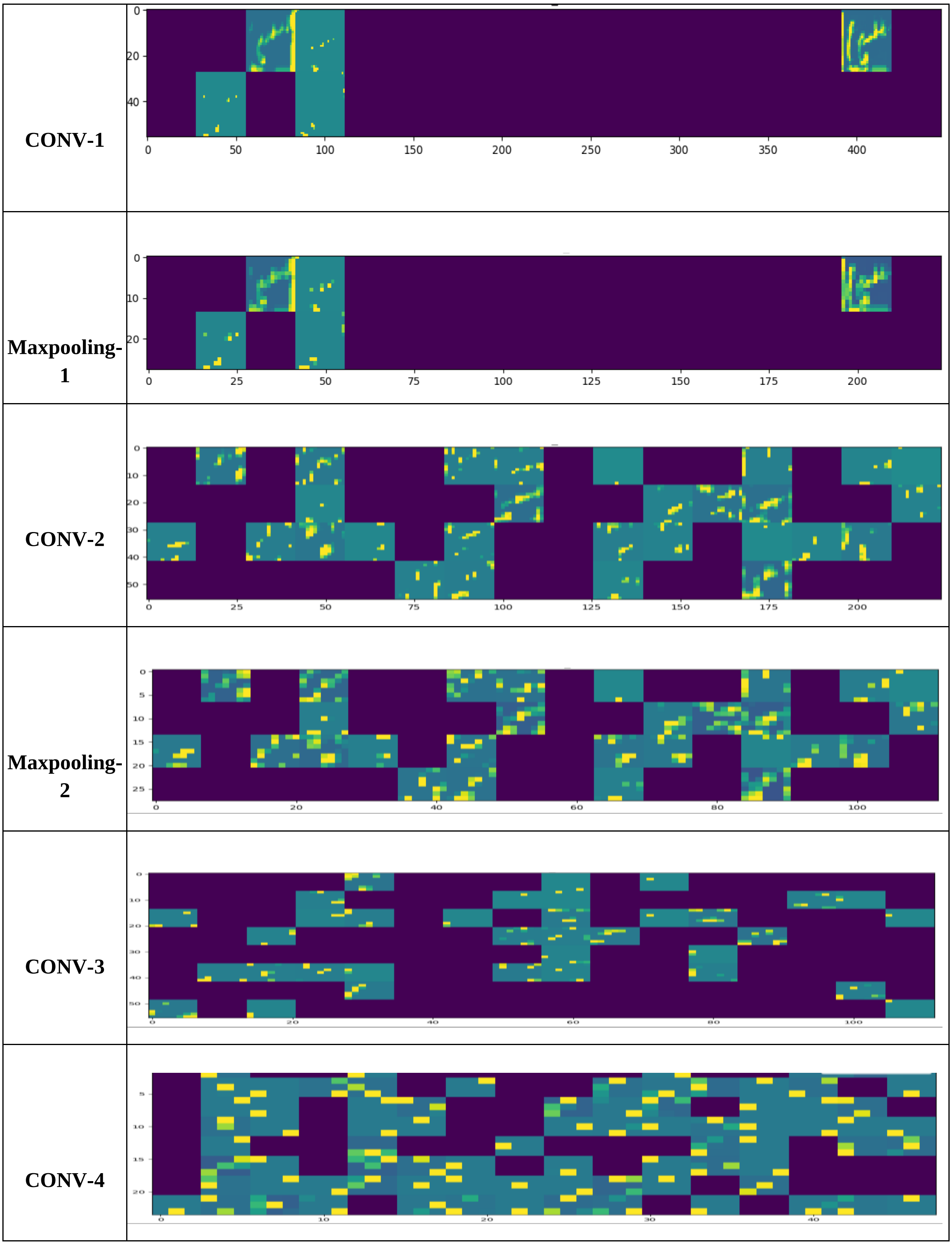


Fig. 7. Activations of all channels of first layer of original model using Four-character ligature dataset.

of convolution of images and filters to generate invariant featureswhich might be handed directly to the subsequent layer. The features in the subsequent layer are convoluted with distinctive filters to generate additional invariant and abstract features and theprocess continues till we get every last feature/output which is invariant to obstructions.

We have engaged a five-layered CNN model (Fig. 5) for the extraction of generic and abstract features from noise-free, normalized ligature image dataset. The first convolution layer C1 of the CNN extracts abstract and generic features such as lines, edges and corner information from the raw pixels of the image. The inner layers are known to extract relatively low-lying features. Therefore, features from the first convolutional layer C1 are selected in the form of convolutional filters. The extracted featuresof the fourth channel of activation for input ligature (س) areshown in Fig. 6. It is clear from the Fig.6 that the first convolutionlayer (CONV-1 or C1) extracts abstract features such as lines, and edges and retains the shape of the ligature image. Max-pooling layer summarizes the most activated presence of a feature and calculates the maximum value for each patch of the feature map as shown in Fig. 6. The second convolutional layer CONV-2 or C2 extracts less visible features such as corner information and information about principal edges in the image.

The set of features extracted from the four-character ligaturedataset are shown in Fig.7. In the output of the feature extraction phase, we get a 28 × 28 feature map with 33 channels. In the wake of plotting every one of the channels of the activation of the primary layer of the CNN model, it has been uncovered that the primary layer is usually holding the full state of the ligatures. Despite the fact that there are a few channels that are not activatedand are left blank. At that stage, the initial layers hold practically the entirety of the data present in the underlying picture. As we go further in the layers, the activations become progressively conceptual and less outwardly interpretable. The activations start to encode more significant level ideas like single boundaries, corners and points. Higher introductions convey progressively fewer data about the visual substance of the picture, and progressively more data identified with the class of the picture. The model construction is exaggeratedly complex to where it verywell may be seen that our last layers not activating by any stretch of the imagination, there's nothing more to learn by then.

The features are extracted from about 150 thousand ligature images belonging to nearly 1500 unique ligature cluster groups (primary as well as secondary). These features are actually the kernels of 5×5 size having specific intensity values that maximallymatch the intensity of Urdu ligature images of a particular class in the recognition phase. The features extracted are used as kernels to train MDLSTM Recurrent neural network discussed in the next section.

3.7 Learning and Training

A 'Recurrent Neural Network' is a densely connected neural network with the property that time-delayed output of the hidden layer is fed back into itself [25]. The RNNs can't learn long term dependencies due to the 'Vanishing Gradient' problem [26] during the backpropagation of error [27] while applying the 'Sigmoid activation function' [28]. This issue was resolved by thepresentation of LSTM–RNN [28] which is fit for holding and corresponding data for longer deferrals. The essential unit of LSTMdesign is a memory block with memory cells and three entryways (input, forget and output) [28]. The standard one-dimensional LSTM organization can likewise be reached out to different measurements by utilizing n self-associations with n forget gates [25].

The MDLSTM RNN has been used here for learning and training of network because MDLSTM has the property of remembering information and output from previous positions in horizontal, vertical, right and left directions. To train MDLSTM Network we employ the features extracted from handwritten Urdu ligatures. We also perform pre-training of the network using benchmark dataset (UNHD) and our proposed dataset (UHLD) to evaluate various performance parameters of the network. It must be noted that both feature extraction and recognition are performed on the same datasets (handwritten Urdu dataset).

Table 6: Network Parameters for system training

Parameter	Value
Batch Size	128
Number of epochs	10
Momentum	0.9
Weight decay regularization	5×10^{-4}
Base Learning Rate	1×10^{-3}

In the beginning, the network was trained on 10 epochs followed by 20,30, 40 and 50, with network parameters given in Table 6. We have performed a compilation of the model using 'Adams Optimizer' with a learning rate of 0.001. if we pick a huge benefit of the learning rate, it would roll out exceptional improvements tothe weights and bias values, for example, it would be taking colossal leaps to arrive at the minimum value. There is likewise an enormous likelihood that it may overpass the global minima and end up on the opposite side of the training curve rather than the minimum value. With a huge learning rate, we won't ever have the option to meet the global minima and will consistently meander around the global minima. If we pick a small value of learning rate, it will surely drop the danger of overshooting the minima however our algorithm will set aside longer intervals to converge, for example, we make more limited strides yet we need to make more strides. Thus, the algorithm would need to train for an extended epoch of interval. Hence an optimal value of .001 is taken as the learning rate parameter.

3.8 Classification and Recognition using MDLSTM:

In the recognition phase, a query complete ligature or a text line presented to the system is first segmented into words/ligatures by applying *K U Rehman's* Algorithm [17]. The words/ligatures extracted are separated into primary and secondary components using the connected component labelling algorithm [as shown in Table 2]. These component images undergo noise removal, normalization and rescaling to form 28×28 sized images. The noise-free, normalized ligature components are input to the feature extraction phase and the computed feature is fed to a trained Multi-Dimensional LSTM network as shown in Fig.5. The MDLSTMs use a recurrence over both axes of the ligature image allowing it to capture a better context over both axes of the image. This allows for the network model to capture writing variations on both axes and directly work on raw features extracted. The first MDLSTM layer consists of 28 LSTM cells and a batch size of 128. Initially, the image is input to eight MDLSTM layers, two layers for each scanning direction.

The LSTM cell inner state and output are computed in equation (1) from the states and the output of previous positions in the horizontal, vertical, right and left directions.

$(hi, j, qi, j) = LSTM(xi, j , hi , j \pm 1, h\ i \pm 1, j , qi , j \pm 1, qi \pm 1 , j)(1)$

Table 7: Matching comparison and classification results using Kernels.

Query Ligature	Computed Feature	Convolution	Kernel	Matching Comparison	Classification Result	Diacritic Association
		⊛	$\begin{matrix} 0 & 0 & 255 & 0 & 0 \\ 0 & 21 & 75 & 73 & 0 \\ 0 & 85 & 180 & 14 & 0 \\ 162 & 30 & 66 & 26 & 0 \\ 0 & 0 & 0 & 255 & 0 \end{matrix}$		 Maximally matched with Four-CharacterLigature Class	 Recognised and classified as Four-Character Ligature, pronounced in Urdu as "Naheey")
		⊛	$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 140 & 62 & 188 & 73 & 0 \\ 155 & 25 & 127 & 90 & 0 \\ 138 & 55 & 76 & 176 & 0 \\ 0 & 0 & 72 & 0 & 0 \end{matrix}$			
		⊛	$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 210 & 255 & 243 & 0 \\ 65 & 51 & 48 & 206 & 0 \\ 0 & 104 & 161 & 255 & 0 \\ 0 & 0 & 0 & 42 & 0 \end{matrix}$			

Where xi,j is the input feature vector at position (i,j) and h and q represent the output and inner state of the cell, respectively. The ±1 options in this recurrence depend on which of the four scanning directions is considered.

Corresponding to an input image, the image skeleton extracted is convolved with multiple kernels (selected during dataset training, few kernels are shown in Table 7). The output of this convolution is used as a feature and passed to the MDLSTM classifier as shown in Table 7. The MDLSTM compares the features of the query ligature with the kernels extracted, by moving the image on the kernel to extract the feature vector and pass it to the classifier along with the ground truth. Thus the ligature that finds maximal match with the kernel is recognized, as shown in Table 7. The CTC (Connection Temporal Classification) layer in Fig. 5 has a continuous output which is fitted through training to model the probability of output label (ligature here) using soft max layer. The classification of most of our query ligatures yields a maximal match. Since the division of ligature cluster groups is based on the number of characters in the ligature, the ligatures recognized are indexed using the proposed flowchart (Fig. 8) to determine the cluster to which a ligature belongs. In the recognition phase, once the index of a ligature is determined, we can predict the class of the recognized ligature. If the value of the index ranges from 1 to 20, the recognized ligature belongs to the single character ligature class. In case index ranges from 21 to 30, the recognized ligature belongs to Urdu numeral class. In case the index ranges from 31 to 420, the recognized ligature belongs to the two-character- ligature class. In case the index ranges from 421 to 740, the recognized ligature belongs to the three-character-ligature class. In case the index ranges from 741 to 1050, the recognized ligature belongs to the four-character-ligature class. In case the index ranges from 1051 to 1260, the recognized ligature belongs to the five-character-ligature class. In case the index ranges from 1261 to 1410, the recognized ligature belongs to the six-character- ligature class. In case the index ranges from 1411 to 1489, the recognized ligature belongs to seven-character-ligature class. In case the index ranges from 1490 to 1500, the recognized ligature belongs to secondary ligature class. To determine the complete ligature, the ligature is associated with diacritics in the diacritic association phase.

3.9 Diacritic Association:

Once the primary or secondary ligatures are recognized, each secondary ligature needs to be assigned to the parent primary ligature. For this purpose, we will find all 'enveloping' as well as 'overlapping' primary ligatures for a given secondary ligature. 'Envelop' refers to a ligature completely occurring within the bounding box of another ligature while 'overlap' refers to the partial occurrence of a secondary ligature within the bounding box of a primary ligature. This approach is presented by *Israr Uddin et al* [18] for printed Urdu to determine the association of secondary ligatures with their exact primary components. We are applying the same heuristics [18] for handwritten Urdu to associate secondary ligatures with primary components. However, since there can be many characters in a primary component/primary ligature, so the dots/diacritics associated with the primary ligature have to be assigned with their actual characters in that primary component. These dots/diacritics are assigned to their respective characters using the pre-computed position and access-order information [18,9]. A lookup table is maintained that carries information on the allowed set of dots and diacritics with each character class. During the association process, each secondary ligature is picked one by one and analysed for potential assignment to the characters in the primary component after verification from the look-up table. Once a secondary ligature is assigned to a character, it cannot be assigned

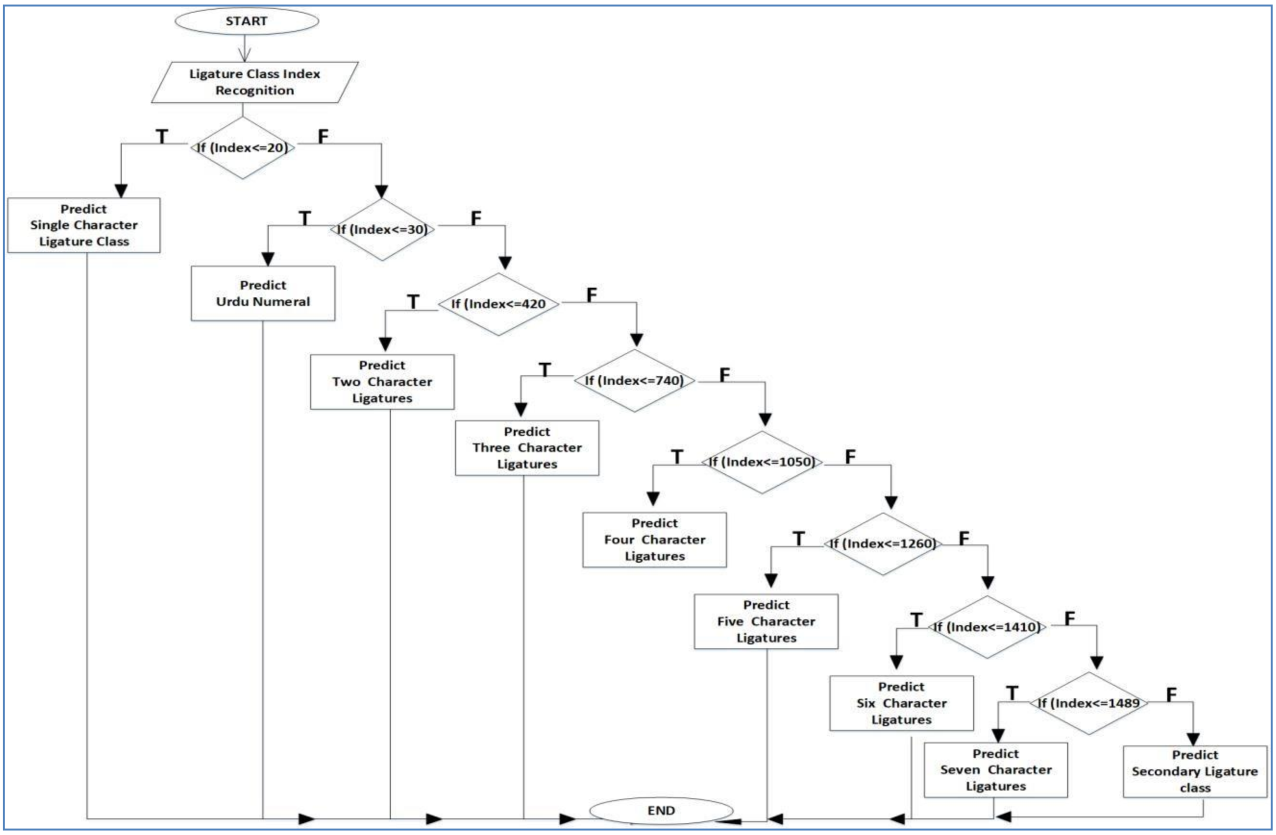


Fig. 8: Flowchart for predicting ligature cluster index

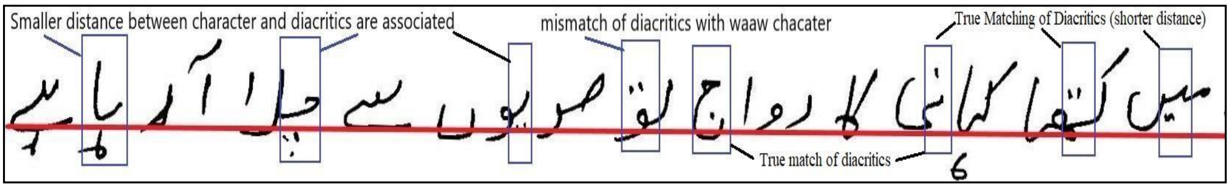


Fig. 9. Diacritic Association Process

to another character and the next secondary ligature in the list is picked for assignment. This complete process of assigning dots/diacritics to their respective character components has been presented by *Israr Uddin et al* [18] for printed Urdu and has been applied here on handwritten Urdu as shown in Fig. 9.

4 Experimental Setup and Results:

The Experiments are carried out on a ligature set of around 150000 images (with repetitions) extracted from both benchmark datasets UNHD [12] (approx. 50,000) and our proposed dataset UHLD (approx. 100,000). After applying the process of diacritic separation (Connected Component Analysis [19]), a large number of ligatures obtained have similarity in shape and hence are reduced to a length of one thousand five hundred unique ligature cluster classes. Also, it is worth mentioning that the ligature dataset extracted whether from UNHD or UHLD is grouped into clusters using our proposed Algorithm-I (described above in Table 3). The two experimental scenarios considered in our study are summarized in Table 8 below.

Table 8. Distribution of datasets into training and testing datasets.

Dataset	**Training size**	**Testing size**	**Ligature cluster size**
UNHD	50000	10000	500
UHLD	100000	20000	1000
Total Dataset	150000	30000	1500

4.1 Results of proposed Algorithm-I:

To save the huge amount of time required for the manual classification of ligatures into different character length clusters, an algorithm (Algorithm-I) has been proposed for determining the number of characters in a ligature as discussed in Table 3. The algorithm has been tested on around 150 thousand ligatures and the results obtained are shown in Table 9. Even if the accuracy of our proposed algorithm is modest, it has saved a significant amount of time needed for manual identification of a class of a ligature.

Table 9. Results of Algorithm-I

Database	No. of Ligatures extracted	No. of Ligatures wherein algorithm successfully determines the actual no. of characters in a ligature.	No. of ligatures wherein algorithm fails to determine no. of characters in a ligature.	Accuracy of Algorithm.
UNHD	50000	37670	12330	**75.2%**
UHLD	100000	82430	17570	**82.4%**

4.2 Classification Report and Confusion Matrix:

The 'Classification Report' visualizer shows the precision, recall, F1, and support scores for the model.

The classification results obtained for different length ligature clusters are shown in Table 10. The overall accuracy of 95.4% obtained on 1500 unique ligature cluster classes demonstrate that the proposed model is highly accurate and outperforms the state- of-the-art for handwritten Urdu recognition [15].

The 'Confusion Matrix' is used to illustrate the overall accuracy of the classifier in addition to truly predicting positive and negative samples. The 'Confusion Matrix' of the Five-Character Ligature class is shown in Fig.10. The matrix indicates the relation between predicted class labels and actual class labels. It is clear from the plot that the matrix has zero values at all false predictions (where predicted class label doesn't match true class label) except diagonal predictions (where Predicted class label exactly equals true class label). So there is 100% sensitivity (True-Positive-Rate of classifier) and 100% True-Negative-Rate of the classifier. The overall accuracy of the classifier is 93.5%. The 'Confusion Matrix' for most of the different length ligature classifiers has shown better accuracy of classification that outperforms the state-of-the-art for handwritten Urdu recognition [15].

Table 10. Summarized classification and accuracy report of all ligature clusters groups(UNHD+UHLD)

LigatureGroup	No. of Clusters	Training Dataset	Test Dataset	Correct Labels	Incorrect Labels	Accuracy
Single Character Ligature	20	11100	2220	2215	05	99.8%
Two Character Ligature	390	24600	4920	4443	477	90.3%
Three Character Ligature	320	23300	4660	4366	294	93.7%
Four Character Ligature	310	22400	4480	4148	332	92.6%
Five CharacterLigature	210	19500	3900	3646	254	93.5%
Six CharacterLigature	150	17700	3540	3377	163	95.4%
Seven Character Ligature	78	13000	2600	2496	104	96.0%
Urdu Numerals	10	8900	1780	1755	25	98.6%
Secondary Ligatures (Diacritics)	12	9500	1900	1858	42	97.8%
Total LigatureDataset	**1500**	**150000**	**30000**	**28304**	**1696**	**95.4%**

The sensitivity, specificity, precision, f1-score and accuracy parameters for different length ligature cluster classes are shown in Table 11 and plotted in Fig.10. The average sensitivity of our model is 93% which shows the predictive model is significantly correct. The average specificity of our model is 97% which showsthe predictive model is significantly specific. The average precision of our model is 94% which shows the predictive model is appreciably precise. The overall average f1-score of our model is 94.5% which demonstrates the predictive model is appreciably correct. The f1-score of the two-character ligature class is plotted w.r.t class labels as shown in Fig. 11.

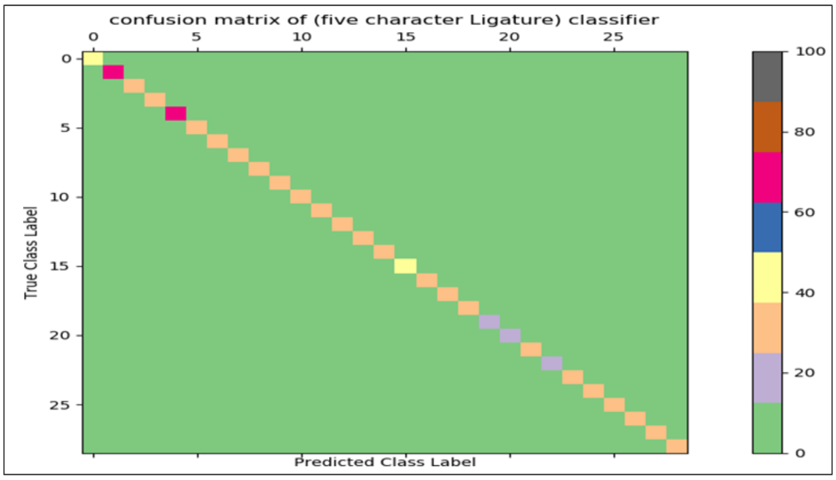


Fig. 10. Confusion Matrix of Five-Character Ligature Classifier

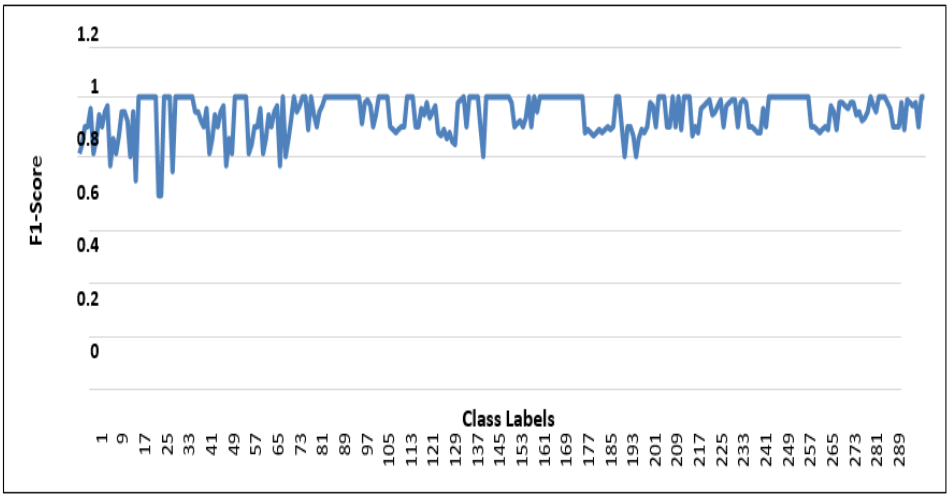


Fig. 11. F1-Score Vs Class Label of Two-Character-Ligature Classes.

It can be visualized from Fig. 11 that most of the classes result in a good f1-score reaching up to 100%, that indicates the high abilityof classification performance of the model.

The accuracy parameters for different length ligature clusters andtotal ligature dataset are shown in Table 11 and plotted in Fig. 12 and it can be visualized that these accuracy parameters are noticeably very high for all ligature cluster classes that outperformthe state-of-the-art for handwritten Urdu recognition [15]. The values of accuracy parameters for two-character-ligature cluster classes are reported to be comparatively less (hence we have plotted its f1-score in Fig. 11) due to the presence of a larger number of cluster classes in two-character-ligature cluster, which increases the chances of resemblance in the shape of one class with another and makes classification prone to errors.

Table 11: Various accuracy parameters of the proposed system

Performance Parameter	Single Character Ligature	Two Character Ligature	Three Character Ligature	Four Character Ligature	Five Character Ligature	Six Character Ligature	Seven Character Ligature	Secondary Ligatures	Urdu numeral Ligature	Total Ligature dataset
Sensitivity	99.4%	88.7%	94.4%	93.8%	94.5%	95.3%	95.7%	98.7%	99.2%	92.88%
Specificity	98.8%	97.5%	95.7%	95.6%	96.4%	96.5%	97.4%	97.6%	98.6%	96.85%
Precision	99.7%	89.6%	93.8%	94.4%	95.7%	97.2%	98.5%	97.9%	98.8%	93.76%
F1-Score	99.6%	91.8%	92.6%	92.7%	94.4%	92.5%	93.8%	98.6%	98.9%	94.53%
Accuracy	99.8%	90.3%	93.7%	92.6%	93.5%	95.4%	96.0%	97.8%	98.6%	95.4%

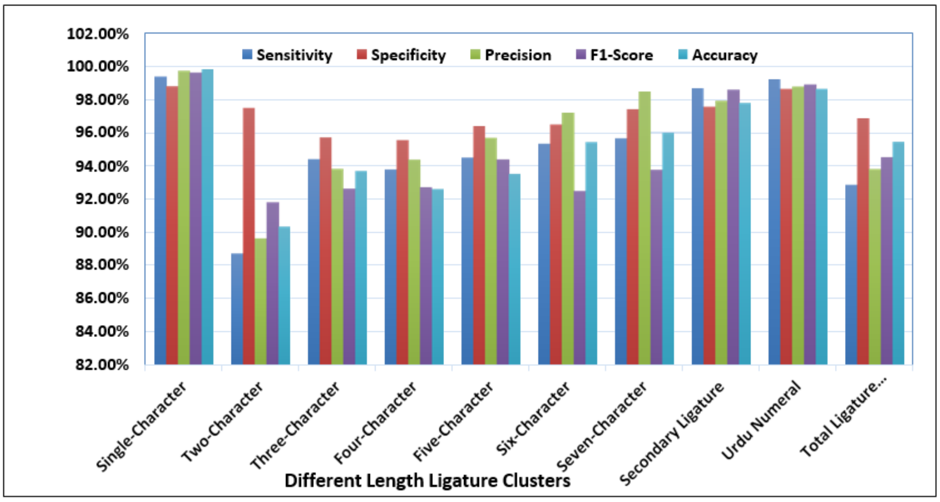


Fig. 12. Accuracy parameters for different length ligature clusters

4.3 Variation of performance of proposed model with systemparameters

In an attempt to study the impact of different network parameters on the training performance, the training accuracies have been testified as a function of these parameters. In the first series of experiments, we vary the number of LSTM layers of RNN from 1 to 7 while keeping other parameters constant. A similar training recognition pattern is observed on both the datasets (UNHD and UHLD) and it is pertinent to mention that the model report maximum training accuracy rates on 8, 10 and 12 layers of LSTM(Fig. 13). However, the accuracy reduces with an increase in the number of LSTM layers because adding more layers will over trainthe dataset.

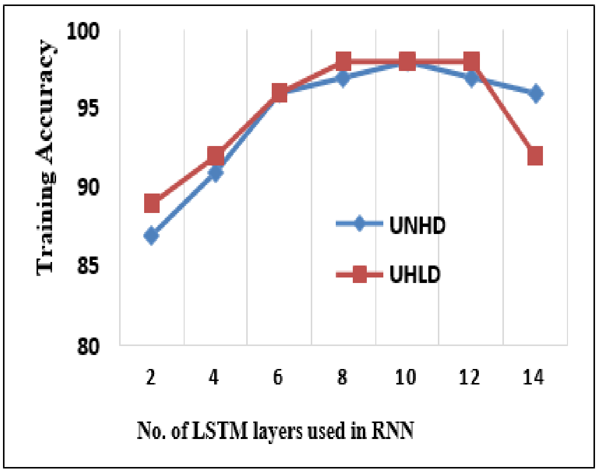


Fig. 13. Variation of accuracy with number of convolutional layers

In an attempt to analyze the scalability of the system, we study theperformance variation concerning the number of ligature clusters i.e. the unique number of ligature classes in the training dataset. The recognition rates are computed by gradually increasing the number of cluster classes and it was found that the accuracy reduceswith an increase in the number of cluster classes as shown in Fig. 3.3.2.3(b). The reason for this variation is that in large number of cluster classes, there are chances that some classes resemble in shape and make classification prone to errors.

However, it is worth to mention here that the proposed model reports a better accuracy of recognition of around 96% whendatasets UNHD and UHLD are trained jointly.

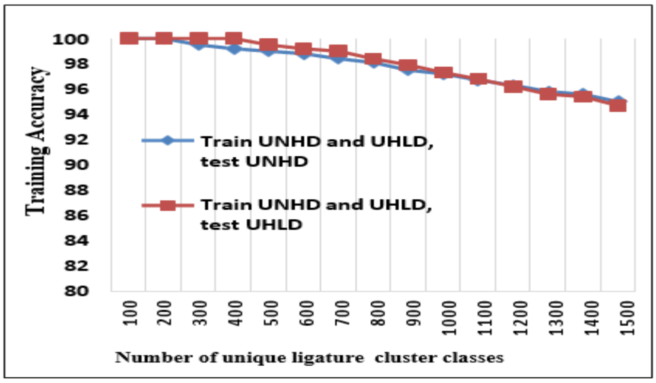


Fig. 14. Accuracy Vs No. of unique ligature classes

4.4 Complexity of the Proposed system

Finally, in the context of the computational complexity of the proposed technique, there are two important phases; extracting theligatures and training/recognition. When a text line is presented to the system, ligatures are extracted by applying the ligature segmentation algorithm [17] and connected component labelling algorithm [19]. This technique is relatively easier than the 'sliding window technique' to recognize characters in a ligature because, in this technique, a complete ligature is recognized at a time.

From a training/recognition complexity point of view, training LSTM RNNs naturally is computationally very expensive. However, in all pattern classification systems, it is the test time complexity that is more critical rather than the training complexity as training is carried out offline. In comparison to other techniques

Table 12. Comparison of proposed work with notable researches in handwritten Urdu recognition.

Study	Database	Training Set	Test Set	Proposed Approach	Accuracy	True Ligature Recognitio n	Computationa lComplexity
Ahmad S.B. et al [11]	UCOM	50 text lines	20 text lines	Not Available	Not Available	Not Available	Not Available
Ahmad S.B. et al [12]	UNHD	6400 Text Lines	1840 text lines	BLSTM Classifier	92%	NO	O(N)
Husnain M. et al [13]	Proposed a database of Urdu Characters.	Information unavailable	Information unavailable	CNN classifier	96% Characters 98% Numerals	NO	O(N)
Hassan S et al, 2019 [14]	Claimed their own Database, not publicly available	4000 Hand written Lines	1000 text lines	7 CNN layers and2 BLSTM layers	83.6%	NO	O(N)
Ahmad S.B.et al [15]	UNHD, MNist Pre-trained Network	10000 Text Lines	2000 Text lines	CNN and MDLSTM	93%	NO	O(N)

Proposed Work	UNHD [12]	5000 text lines	1000 text lines	LRCN model	**94.2 %**	**YES**	**O(1)**
	UHLD [our proposed dataset]	6000 text Lines	1000 text lines	LRCN model	**96.6 %**	**YES**	**O(1)**
	Both UNHD and UHLD	**11000 text lines**	**2000 text lines**	LRCN model	**95.4%**	**YES**	**O(1)**

where a separate model is trained for each ligature class (forexample a separate HMM for each class in [9]), a single LSTM RNN model is required in our system to recognize all ligature classes. Modelling a separate classifier for each class has a test timecomplexity of O(N) as a query ligature is fed to all N trained models(N is the number of unique classes). On the other hand, irrespective of the number of classes, in our technique, a query ligature is fed toonly one model and hence has a complexity of O (1) making it efficient in terms of test time computational complexity.

4.5 Comparison of results with notable studies:

As discussed earlier, the Handwritten Urdu recognition has been the least explored to date and our proposed work outperforms thestate-of-the-art [15] in terms of accuracy parameters and computational complexity reported and forms the new state-of- the-art for handwritten Urdu recognition. A comparison of recognition rates of notable research in handwritten Urdu and thatof proposed work is summarized in Table 12.

5. Conclusion and future scope

Handwritten Urdu recognition is a challenging area in natural language processing. In this chapter, we have proposed another holistic approach of handwritten Urdu text recognition. In this approach, we construct the ligature classes from benchmark datasets UNHD and UHLD. The features are extracted from theseclasses using CNNs, followed by training MDLSTM RNN using these features to recognize any unconstrained handwritten Urdu text. In this research, several benchmark methods and metrics suchas confusion matrix, specificity, sensitivity, accuracy (ACC) and precision were computed to accurately assess the proposed recognition model. The results of these parameters demonstrate that the predictive model depicts better recognition rates. The individual recognition rate of a few ligature clusters like four-character ligature class, five-character ligature class, six-characterligature class and seven-character ligature class and even Urdu numerals reports100% accuracy.

The proposed technique is computationally efficient as it offers a constant time complexity of O (1) only (a constant number) irrespective of the number of classes used. Simultaneous training of all of the ligature classes (primary, secondary and Urdu numerals) has shown a good recognition rate for a large number of ligature clusters that demonstrate the success of the proposed technique for recognition of handwritten Urdu ligatures and formsthe new state-of-the-art for handwritten Urdu recognition. This study can be improved more by performing a semantic and contextual study of handwritten cursive scripts. Through semanticanalysis, we can predict a ligature's most appropriate class and wecan check the meaning of that ligature in the dictionary. Also, we can textually characterize a text and can infer knowledge from it by doing contextual analysis.

Acknowledgements

We thank Dr. Saad Bin Ahmad for providing UNHD database.

References

[1]. Naz S, Umar AI, Shirazi SH, Khan SA, Ahmed I, Khan AA. Challenges of Urdu named entity recognition: a scarce resourced language. Research Journal of Applied Sciences, Engineering and Technology. 2014 Sep 15;8(10):1272-8.

[2]. Daud A, Khan W, Che D. Urdu language processing: a survey. Artificial Intelligence Review. 2017 Mar 1;47(3):279-311.

[3]. Satti DA, Saleem K. Complexities and implementation challenges in offline urdu Nastaliq OCR. In Proceedings of the Conference on Language & Technology 2012 Nov (pp. 85-91).

[4]. Khan NH, Adnan A. Urdu optical character recognition systems: Present contributions and future directions. IEEE Access. 2018 Aug 16;6:46019-46.

[5]. Dawood, Hassan, Hussain Dawood, and Ping Guo. "Improved Arabic Word Classification using Spatial Pyramid Matching Method."

[6]. Wang, L., Chen, K., Long, Y., Mao, X. and Wang, H., 2015, September. A modified efficient certificateless signature scheme without bilinear pairings. In 2015 International Conference on Intelligent Networking and Collaborative Systems

(pp. 82-85). IEEE.
[7]. Naz, S., Umar, A.I., Ahmed, R., Razzak, M.I., Rashid, S.F. and Shafait, F., 2016. Urdu Nasta'liq text recognition using implicit segmentation based on multi-dimensional long short term memory neural networks. Springer Plus, 5(1), p.2010.
[8]. Naz, S. et al., 2015. Segmentation techniques for recognition of Arabic-like scripts: A comprehensive survey. Education and Information Technologies, 21(5), pp.1225–1241. Available at: http://dx.doi.org/10.1007/s10639-015-9377-5.
[9]. Din, I.U., Siddiqi, I., Khalid, S. and Azam, T., 2017. Segmentation-free optical character recognition for printed Urdu text. EURASIP Journal on Image and Video Processing, 2017(1), p.62.
[10]. Lehal, G.S., 2012, December. Choice of recognizable units for Urdu OCR. In Proceeding of the workshop on document analysis and recognition (pp. 79-85).
[11]. Ahmed, S.B., Naz, S., Swati, S., Razzak, I., Umar, A.I. and Khan, A.A., 2017. UCOM Offline Dataset-An Urdu Handwritten Dataset Generation. International Arab Journal of Information Technology (IAJIT), 14(2).
[12]. Ahmed, S.B., Naz, S., Swati, S. and Razzak, M.I., 2019. Handwritten Urdu character recognition using one-dimensional BLSTM classifier. Neural Computing and Applications, 31(4), pp.1143-1151.
[13]. Husnain, M., Saad Missen, M.M., Mumtaz, S., Jhanidr, M.Z., Coustaty, M., Muzzamil Luqman, M., Ogier, J.M. and Sang Choi, G., 2019. Recognition of Urdu handwritten characters using convolutional neural network. Applied Sciences, 9(13), p.2758.
[14]. Hassan, S., Irfan, A., Mirza, A. and Siddiqi, I., 2019, August. 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) (pp. 67-72). IEEE.
[15]. Ahmed, S.B., Hameed, I.A., Naz, S., Razzak, M.I. and Yusof, R., 2019. Evaluation of handwritten Urdu text by integration of MNIST dataset learning experience. IEEE access, 7, pp.153566-153578.
[16]. Gang Chen and Sargur N. Srihari. 2014. Removing Structural Noise in Handwriting Images using Deep Learning. In Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing (ICVGIP '14). Association for Computing Machinery, New York, NY, USA, Article 28, 1–8.
DOI: https://doi.org/10.1145/2683483.2683511.
[17]. Rehman KU, Khan YD. A Scale and Rotation Invariant Urdu Nastalique Ligature Recognition Using Cascade Forward Backpropagation Neural Network. IEEE Access. 2019 Aug 19; 7:120648-69.
[18]. Uddin I, Javed N, Siddiqi IA, Khalid S, Khurshid K. Recognition of printed Urdu ligatures using convolutional neural networks. Journal of Electronic Imaging. 2019 May;28(3):033004.
[19]. Singh, Y.K., 2016. Finding Connected Components in a Gray Scale Image. *ADBU Journal of Engineering Technology*, *5*(2).
[20]. Mostafavi SM, Kazerouni IA, Haddadnia J. Noise removal from printed text and handwriting images using coordinate logic filters. In2010 International Conference on Computer Applications and Industrial Electronics 2010 Dec 5 (pp. 160-164). IEEE.
[21]. Devi H. Thresholding: A Pixel-Level image processing methodology preprocessing technique for an OCR system for the Brahmi script. Ancient Asia. 2006 Dec 1;1.
[22]. Singh TR, Singh OI. Image Resizing with Enhancement Technique on DCT Domain.
[23]. Pavlidis T. A thinning algorithm for discrete binary images. Computer graphics and image processing. 1980 Jan 1;13(2):142-57.
[24]. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In2017 International Conference on Engineering and Technology (ICET) 2017 Aug 21 (pp. 1-6). Ieee.
[25]. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena. 2020 Mar 1;404:132306.
[26]. Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 1998 Apr;6(02):107-16.
[27]. Hecht-Nielsen R. Theory of the backpropagation neural network. In Neural networks for perception 1992 Jan 1 (pp. 65-93). Academic Press.
[28]. Narayan S. The generalized sigmoid activation function: Competitive supervised learning. Information sciences. 1997 Jun 1;99(1-2):69-82.
[29]. Ganai AF, Koul A. Projection profile based ligature segmentation of Nastaleeq Urdu OCR. In2016 4th International Symposium on Computational and Business Intelligence (ISCBI) 2016 Sep 5 (pp. 170-175). IEEE.