International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

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

Leveraging the Swin Transformer for Enhanced Handwritten Urdu Character and Digit Recognition

Shaik Moinuddin Ahmed *[1], Abdul Wahid [2]

Submitted: 27/01/2024 **Revised**: 05/03/2024 **Accepted**: 13/03/2024

Abstract: The Swin Transformer model is used to identify handwritten Urdu characters and digits. The paper also focuses on the challenges associated with the script of Urdu, a language that uses an Arabic-based alphabet. It also explores its ability to read different character types and scripts. This is achieved by training on a large dataset and then testing it for character recognition ability. "MANUU: Handwritten Urdu OCR Dataset" was employed in order to measure the performance of the model in terms of accuracy, precision, recall and F1-score. However, The Swin Transformer outperforms existing CNN and LSTM classifiers in accordance with the higher levels of accuracy (97%), F1-score (89%), recall (91%) as well as precision (83%). With its benefits over traditional classifiers, these results further demonstrate how efficient Swin Transformer performs better than current CNN or LSTM classifiers for achieving higher levels of accuracy (97%), F1-score (89%), recall (91%) as well as precision (83%). So, this research highlights how valuable the Swin transformer is when it comes to recognizing correct Urdu letters besides comparing it with other traditional classifiers that fail in this regard. In addition, there are certain things within this study that show how effective can be Swin transformer while solving problems which are particular to some languages pointing at his importance within bigger framework of language processing.

Keywords: *Swin Transformer, character recognition, handwritten Urdu characters, digit recognition, CNN, LSTM*

1. Introduction

It's difficult to identify and transcribe handwritten texts for computer vision and pattern recognition researchers today. The identification task is complicated by numerous factors including the complexity of cursive scripts us It's difficult to identify and transcribe handwritten texts for computer vision and pattern recognition researchers today. The identification task is complicated by numerous factors including the complexity of cursive scripts used, similarities among letters, and different writing techniques. Even though studies reveal that word and character recognition rates in printed text are better than those in handwriting, there is still a dire need of an effective system capable of producing excellent results when it comes to hand-written recognition. ed, similarities among letters, and different writing techniques. Even though studies reveal that word and character recognition rates in printed text are better than those in handwriting, there is still a dire need of an effective system capable of producing excellent results when it comes to hand-written recognition. [1]. A cursive language like Urdu is very important and commonly spoken in parts of South-East Asia including Pakistan, India, Afghanistan, Bangladesh among others. For example, it should be noted that even though OCR technology for Urdu script started being used in the early 2000s there is a significant gap in research compared to other scripts such as Arabic or English. [2]. Several commercially accessible Urdu OCR solutions exist, however none of them currently support the recognition of handwritten Urdu texts. Within the field of computer vision and pattern recognition, it is important to emphasize that "intelligent character recognition" (ICR) refers particularly to handwritten text recognition, whereas "optical character recognition" (OCR) focuses on analysing printed text found in various media. In this book, we shall use the term ICR [3] to refer to handwritten text recognition. The difficulty may be addressed through the utilization of machine learning (ML) methods. This paper focuses on improving recognition of characters and digits in handwritten Urdu by applying advanced machine learning techniques. The task of identifying handwritten Urdu letters and numbers has various complicated aspects like feature extraction, classification, and preparation for analysis. We therefore aim at utilizing machine learning to not only make the script simpler but also improve the recognition levels to an unprecedented level. In order to find out if these machine learning algorithms can help greatly in enhancing the recognition of handwritten Urdu characters and digits, we would like to carry out empirical analysis as well as extensive testing with a view to determining their effectiveness. Consequently, this research indicates that there is a possibility for better OCR systems and it has implications in many fields such as language processing, preservation of historical documents or multilingual communication channels.

1.1. Urdu Script

[1] *Department of CSIT, Maulana Azad National Urdu University, Hyderabad, India.*
ORCID ID : 0000-0001-5395-7648
[2] *Department of CSIT, Maulana Azad National Urdu University, Hyderabad, India.*
ORCID ID : 0000-0001-6729-7775
** Corresponding Author Email: moinuddinahmed@manuu.edu.in*

In its Eighth Schedule, the Indian constitution recognizes the incorporation of Urdu as a language and highlights its rich cultural heritage. From around the 12th century, Urdu began to be popular in parts of north India especially around Delhi. The dialect spoken in Delhi was heavily influenced by Arabic, Persian and Turkish. ule, the Indian constitution recognizes the incorporation of Urdu as a language and highlights its rich cultural heritage. From around the 12th century, Urdu began to be popular in parts of north India especially around Delhi. The dialect spoken in Delhi was heavily influenced by Arabic, Persian and Turkish. Urdu literature flourished during the fourteenth and fifteenth centuries, with a rich collection of poetry and other written works. Arabic and Hindi have common roots, making them closely related in terms of grammar. Urdu is often considered as a language closely related to Hindi due to their shared grammatical bases. Among the various regions in India, the states of Telangana, Andhra Pradesh, Uttar Pradesh, Bihar, and Jammu & Kashmir, along with Delhi, the National Capital Territory, have officially recognized Urdu language. [4] The exceptional standard The Urdu script has the unique feature of being written in both directions Although the primary text is written in a left-to-right direction, the numerals are written in a right-to-left direction. The Urdu writing system comprises 10 numeral characters and 38 fundamental letters. which are classified as joiners and non-joiners in figure 1 and figure 2, respectively. This character set is notable for being a compilation of various scripts derived from Urdu. It includes the 28 letters of Arabic and 32 characters from Persian, as depicted in Figure 3. Furthermore, the Urdu script incorporates additional letters to represent Hindi phonemes [5]. The only distinction between the phonologies of Hindi and Urdu lies in their written scripts. Arabic and Persian are two scripts based on Urdu that have unique characteristics. Firstly, they are written in cursive style, flowing from right to left. Additionally, these scripts exhibit context sensitivity, resulting in the creation of ligatures that connect multiple letters. Due to various factors such as word position and surrounding characters, many characters can assume different forms. The context awareness has greatly improved the Urdu lexicon by incorporating almost 24,000 ligatures [6]

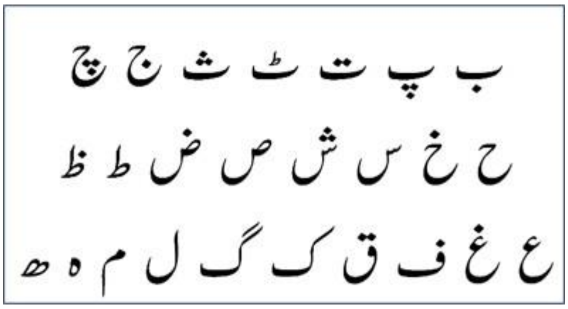

Fig 1: Joiners in Urdu

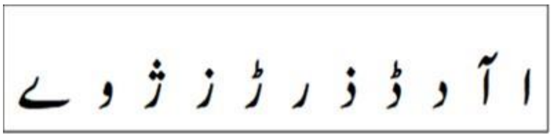

Fig 2: Non-Joiners in Urdu

Recognizing Urdu characters can still be challenging due to the complexity of the language. Despite Urdu being widely spoken, there has been limited research on the handwritten Urdu recognition. Identifying characters in Urdu might be more difficult than in languages such as English and others because of the inclusion of diacritics. There are several factors that contribute to the complexity of Urdu [7]:

- Cursive Style: Urdu is written in a cursive script, which adds complexity to the writing process because of the intricate method in which words are connected.

- Diacritics: Diacritics are used in Urdu, including dots, diagonals, maraca, hamza, etc., which are secondary characters used above or below main characters.

- Script Variety: Urdu has 12 different scripts, making recognition techniques script- specific and challenging to adapt for other scripts

- Bidirectional Writing: Urdu is a bidirectional language, providing additional level of complexity, in contrast to the majority of languages that are unidirectional.

- Strokes: Urdu adheres to the guideline of having zero, one major stroke, or one minor stroke, which makes it much harder to recognize characters.

Language	Script	Writing Direction	Additional Features	Characters	~No. of Characters	Digits
Arabic	Arabic	Right to Left	Connected letters	ا ب ت ث ج ح خ د ذ ر ز س ش ص ض ط ظ ع غ ف ق ك ل م ن ه و ي	28 (Basic)	٠ ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩
Urdu	Arabic	Right to Left	Additional letters and diacritical marks	ا ب پ ت ٹ ث ج چ ح خ د ڈ ذ ر ڑ ز ژ س ش ص ض ط ظ ع غ ف ق ک گ ل م ن و ہ ھ ی ے	39 (Basic)	۰ ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹
Persian	Arabic	Right to Left	Additional letters to represent Persian sounds	ا ب پ ت ث ج چ ح خ د ذ ر ز ژ س ش ص ض ط ظ ع غ ف ق ک گ ل م ن و ه ی	32 (Basic)	۰ ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹
Pashto	Arabic	Right to Left	Additional letters and sounds specific to Pashto	ا ب پ ت ټ ث ج چ ح خ څ ځ د ډ ذ ر ړ ز ژ ږ س ش ښ ص ض ط ظ ع غ ف ق ک ګ ل م ن ڼ و ه ۀ ي ې ی ۍ ئ	44 (Basic)	۰ ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹
Sindhi	Arabic	Right to Left	Additional letters and characters for Sindhi phonetics	ا ب پ ت ٿ ث ج ڄ چ ح خ د ڌ ڏ ذ ر ڙ ز س ش ص ض ط ظ ع غ ف ق ڪ ل م ن ڻ و ۆ ه ھ ي ۾ ۾ ۾ ۾	52 (Basic)	۰ ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹

Fig 3: Comparison of Scripts Similar to Urdu, their characters and Digits

This process utilizes a digital image as input and generates the required character through the recognition of handwritten characters. When it comes to character identification, an automated system outperforms manual identification done by human employees in terms of efficiency. Identifying handwritten characters poses a greater challenge in comparison to other OCR (optical character recognition) methods due to the diverse writing styles of individuals. [8]. Many different image processing jobs use handwritten character recognition (HCR). such as reading postal codes and house numbers, optical character recognition, machine translation from Urdu to English, linguistic connectivity, picture restoration, and robotics. [9].

2. Related work

[10] A modern, large-scale offline database of handwritten Urdu literature is called the Urdu Nastaliq Handwritten Dataset (UNHD). Scholars now lack access to a standard, comprehensive dataset of handwritten Urdu. The gathered dataset consists of commonly used ligatures written in the natural handwriting of 500 writers on A4 paper. In his research, the author employed recurrent neural networks and achieved a significant level of accuracy in handwritten Urdu character detection. [11] This study proposed a methodology for categorizing and mechanically identifying handwritten Urdu numbers and characters. The approach utilized the transfer learning principle and pre-trained convolutional neural networks (CNNs). The study that has been suggested, which uses hybrid datasets, Urdu characters, and numbers and fine-tunes AlexNet, outperforms similar state-of-the-art studies in all three cases with classification accuracy scores of 97.08%, 98.21%, and 94.92%. [12] In order to distinguish offline handwritten Urdu text with many typefaces in an uncontrolled environment, it was suggested to employ a convolution neural network. Since a collection of handwritten Urdu characters is not publically available, the author offered a supplementary dataset of these letters. We perform several tests on our proposed dataset. When compared to findings reported in the literature with the identical task, the accuracy obtained while identifying characters is among the highest [13]. [14] provided a fresh dataset of handwritten Urdu characters and digits from more than 900 individuals. It displays successful recognition outcomes using convolution neural network and deep auto encoder frameworks. The deep auto encoder achieves 97% accuracy for digits, 81% accuracy for characters, and 82% accuracy for both at the same time. The convolutional neural network achieves a precision rate of 96.7 percent for numerals, 86.5% for individuals, and 82.7% for numbers and characters. These frameworks could serve as helpful beginning points for further research on the recognition of handwritten Urdu text. [15] The dataset is openly accessible and contains substantial volumes of accurate information, as well as 600 pages of handwriting Urdu writing in the Nasta'liq script. It is designed specifically for evaluating the performance of handwritten character recognition systems. The collection consists of scanned and segmented text lines that were written on A4 size paper by a small number of people. It contains all ligatures, numeric data variations, and Urdu characters. The UCOM dataset makes writer identification and handwritten character recognition easier. The study also proposes and assesses the effectiveness of the recurrent neural network (RNN) on a sample text line from the UCOM offline database. [16] A sizable Urdu corpus with multiple levels and scripts is MMU-OCR-21. This is the largest compiled set of printed Urdu text ever compiled for deep learning applications. The body of work consists of around 602,472 images in total, which include word pictures and text lines in three prominent typefaces together with the corresponding ground truth. Additionally, we used numerous cutting-edge deep learning techniques to do studies on text-line and word-level images [21-26]. In the upcoming sections, we provided a comprehensive overview of the dataset utilized, the methodologies implemented, the experimental results obtained, and our insights on the future direction. This paper showcases progress in addressing the challenge of handwritten Urdu by leveraging machine learning techniques. The aim is to enhance character and digit recognition in this script, contributing to advancements in the field.

3. Background

This section offers an overview of the Swin transformer architectures we recommend, in addition to other designs like CNN and LSTM.

3.1. CNN

Convolutional neural networks, or CNNs, are a specific kind of artificially generated neural network that is frequently employed in deep learning tasks for the purpose of analyzing visual pictures. The term "convolutional neural networks neural network" (CNN) is also used to describe it. A network of convolutional neural networks, or CNN, consists of three layers: A structured framework comprising of an information layer, a hidden layer, and a final output layer. [17]. Given that the previous Fourier and activation method conceal the inputs and outputs they produce, all of the intermediate stages in a system of lateral connections between neurons can likewise be denoted as hidden layers. The hidden layers of a convolutional neural network are composed of layers of convolution that execute convolutions. Here is the information you requested [18].

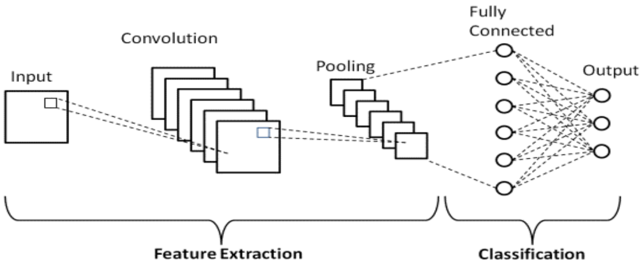

Fig 4: CNN Architecture

3.2. LSTM

An advanced model called LSTM is utilized to detect potential network breachesA modification of a recurrent network of neurons

structure, known as LSTM, is a DL model. When it comes to handling un-segmented data, recognizing relationship patterns, and predicting abnormalities in network traffic, LSTM can be a valuable tool. Voice recognition and IDSs are also areas where LSTM can be applied effectively [19]. An LSTM system consists of four essential components: the input gates, departure gate, memory gate, and cell. Due to the limitations of cell memories in terms of time, three gates were first established to control the movement of information into and out of a cell. The diagram shown depicts the Long Short-Term Memory (LSTM) model for each stage of a cell. The uppermost horizontal line shows the cell status and contains what is done with cell state. Once the data has been completely standardized by the gates that are in a cell nation, the LSTM has no way of eliminating it. A cell comprises essential structural elements. [20].

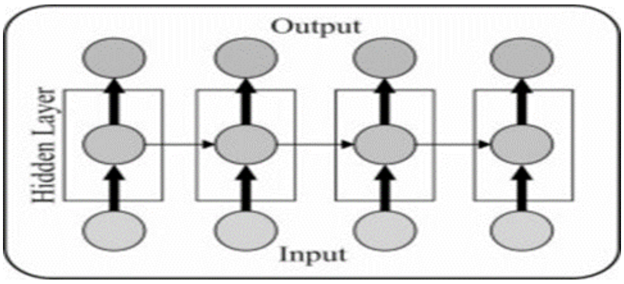

Fig 5: Structure of LSTM cell

3.3. Swin Transformer

The Swim Inverter is a novel design of neural networks that has had a substantial influence in the domain of computer vision. As an extension of the Vision Transformer (vit), the Swin Transformer introduces innovative hierarchical architectures and shifting mechanisms, which effectively address the limitations of traditional transformers for vision tasks. This advanced model surpasses previous state-of-the-art techniques in semantic segmentation, object identification, and image recognition. Conventional convolutional neural networks have long been the go-to for computer vision tasks. However, they do have their drawbacks when it comes to capturing the bigger picture and dealing with long-range connections in images. Vision Transformers provided a compelling alternative by employing self-attention processes that are highly efficient in capturing the overall environment. Nevertheless, they faced challenges in terms of expanding to high-resolution photos because of the intricate computing demands. The Swin Transformer architecture incorporates a patch splitting module to divide an input RGB image into non-overlapping patches, using a similar approach as the Vision Transformer (ViT). The diagram depicts the compact variant known as SwinT. Each patch is considered a "tokens," and its characteristic is determined through the combination of the corresponding RGB values of each of the pixels. The size of each patch is 48 (4x4x3) when the patch size is 4x4. This raw valued feature is subsequently enhanced and mapped to any dimension by means of a linear embedding layer. The first step processing sets the basis for subsequent operations in Swin Transformer architecture.

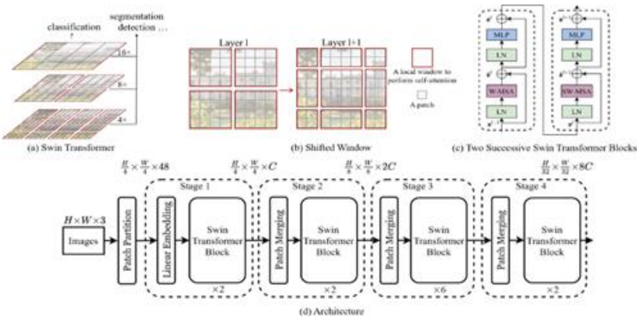

Fig 6: Swin Transformer architecture

4. Methodology

The primary goal for this research was to use a swin transformer for the accurate identification of Urdu characters and numerals in written form. We deployed a highly qualified and experienced model of transformer, which had been trained extensively on huge dataset. In turn, this model was used to precisely identify handwritten Urdu letters and digits. The aim of the study was to test the ability of the proposed model in relation to adapting itself to different challenges of Urdu script including changing letters and writing styles. The purpose of the study was to evaluate whether or not this approach would be effective in terms of character recognition tasks for Urdu language through examining results and accuracy levels from the suggested models on handwriting-based Dataset.

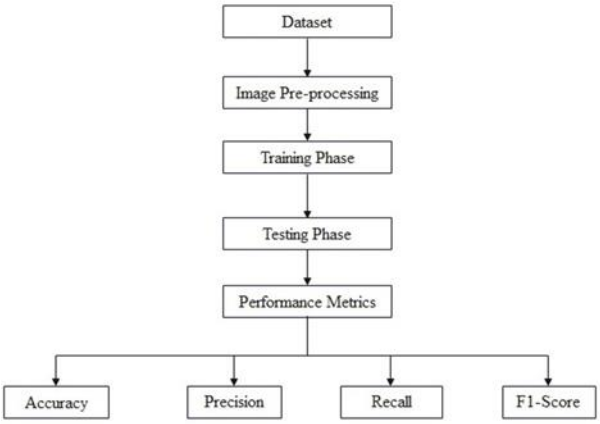

Fig7: Flowchart for proposed methodology

4.1. Dataset description

The "MANUU: Handwritten Urdu OCR Dataset" is a highly detailed and meticulously chosen collection that aims to promote the research in optical character recognition (OCR) field for handwritten Urdu characters, digits, and words. The dataset has been carefully compiled with numerous hand written examples to provide a broad sample from which robust OCR systems can be built and assessed. These systems have to be developed specifically for the complexities of the Urdu script.

Table 1 Dataset Description

<table>
<tr><th colspan="4">School Writers</th></tr>
<tr><td></td><td>Left Handwritten</td><td>Right Handwritten</td><td>Total</td></tr>
<tr><td>Male</td><td>7</td><td>77</td><td>84</td></tr>
<tr><td>Female</td><td>11</td><td>145</td><td>156</td></tr>
<tr><th colspan="4">College Writers</th></tr>
<tr><td></td><td>Left Handwritten</td><td>Right Handwritten</td><td>Total</td></tr>
<tr><td>Male</td><td>13</td><td>117</td><td>130</td></tr>
<tr><td>Female</td><td>22</td><td>257</td><td>279</td></tr>
<tr><td colspan="3">Total Writers</td><td>649</td></tr>
<tr><th colspan="2">Dataset Aspect</th><th colspan="2">Details</th></tr>
<tr><td colspan="2">Pages of Handwritten Examples</td><td colspan="2">2596</td></tr>
<tr><td colspan="2">Total Handwritten Character images</td><td colspan="2">172634</td></tr>
<tr><td colspan="2">Total images of Digits</td><td colspan="2">6,490</td></tr>
<tr><td colspan="2">Total images in Isolated Form of Characters</td><td colspan="2">25,391</td></tr>
<tr><td colspan="2">Total images in Initial Form of Characters</td><td colspan="2">23,364</td></tr>
<tr><td colspan="2">Total images in Medial Form of Characters</td><td colspan="2">16,874</td></tr>
<tr><td colspan="2">Total images in Final Form of Characters</td><td colspan="2">24,662</td></tr>
<tr><td colspan="2">Total images of Special Characters</td><td colspan="2">11,682</td></tr>
<tr><td colspan="2">Total Images of Words</td><td colspan="2">61,006</td></tr>
<tr><td colspan="2">Utility for OCR-related Tasks</td><td colspan="2">Comprehensive resource for OCR development and evaluation</td></tr>
<tr><td colspan="2">Focus</td><td colspan="2">Handwritten Urdu characters, digits, and words</td></tr>
<tr><td colspan="2">Diversity</td><td colspan="2">Represents various handwriting styles and variations</td></tr>
<tr><td colspan="2">Contribution</td><td colspan="2">Collaborative effort with over 649 writers</td></tr>
</table>

The “MANUU” data set is a comprehensive compilation of handwritten Urdu samples, comprising demographic information, education level and hand orientation providing these details to highlight the wide variety of writing styles and features in the real world. This extensive dataset constitutes a solid ground for OCR models that are able to adapt easily to new data. The dataset is neatly organized bringing several advantages for researchers as well as developers. With this tool, model development and benchmarking are possible leading to significant improvements on Urdu OCR accuracy and efficiency. By incorporating different writing styles, it enhances generalizability thus training OCRs with high capability to handle natural variability present in real-world handwritten Urdu contents. A lot of opportunities for many applications including looking into how historical document works and to enable educative resources on boarders of where Urdu manual transcription meets with automated processing are opened up.

4.2. Image Pre-processing

Image processing is crucial for the recognition of handwritten Urdu characters and numbers. It increases Optical Character Recognition, or OCR, precision effectively. These are the steps followed when applying image processing in this case.

Step 1: Remove Noise from Image

This helps to make letters clearer on any given input image by getting rid of unwanted noises. Typical noise removal methods entail median filtering, Gaussian filtering and morphological operations.

Step 2: Contrast Adjustment

Adjust the contrast of the image to improve the visibility of the characters. This phase helps differentiate between the foreground (the characters) and the background of the image.

Step 3: Binarization

Convert the grayscale picture into a binary image by setting all of the pixels that are over a specified threshold value to white and all of the other pixels to black. This process simplifies the character recognition task by reducing the complexity of the image.

4.3. Two Level Tokenization

This study presents the development of a two-level image for Urdu characters and digits recognition. It encompasses tokenization at both the word level and character level. Unique Urdu tokens were generated to create a vocabulary at a professional level. This was done to ensure that the future phase would not result in any duplicate pictures. At this level of tokenization, word separation was achieved by utilizing white space and punctuation characters as delimiters.

We manually included the various versions of Urdu characters, including the beginning, middle, and final forms, along with Urdu numbers and punctuation marks, at the character level.

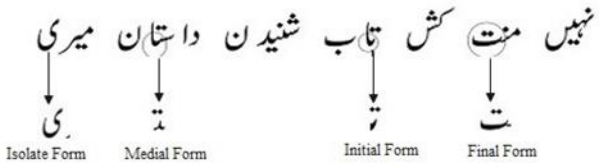

Fig 8: Tokenization

4.4. Train _test split

After completing all pre-processing procedures, The dataset is partitioned into distinct sets to be used for testing and training, determined by the user's specified split ratio. We have collected 37 Urdu characteristics, including digits 0-9, featuring a wide range of handwritten images.

4.5. Training Phase

To accurately ascertain and arrange handwritten Urdu characters and figures within pre-processed images, this study uses Swin Transformer. Through character segmentation, separate the identified word regions into individual characters.

4.6. Testing Phase

Testing phase employs lots of hand-written images which cover thirty-seven Urdu characters and numerals 0-9. Both existing and proposed models were trained on this dataset, they demonstrate our method's outstanding efficiency in correctly identifying and understanding complex handwritten Urdu script. The Swin transformer model has a high level of expertise in detecting and classifying handwritten Urdu characters as well as numbers. Additionally, character segmentation is used to improve it for accurate recognition of isolated characters.

4.7. Character Segmentation

The spacing between characters is determined by carefully examining each word with a view to defining the boundaries of each character. For this purpose, several methods can be utilized, including horizontal projection profiles, connected component analysis or contour detection. The task of extracting individual characters from a word is one that demands accuracy and thoroughness.

Aside from the CNN and LSTM, other established models such as can be used for training the proposed swine transformer model using train data. By incorporating both CNN and LSTM, We assess and contrast the efficacy of the suggested paradigm. The model's effectiveness can be evaluated using the following metrics.

4.8. Performance metrics

The evaluation of a technique's effectiveness involves assessing its reliability, accuracy, sensitivity, and F1-score based on the confusion matrix.

Accuracy: It is the number of subjects from all the subjects that were successfully identified.

$$\boldsymbol{Accuracy = \frac{TP+TN}{TP+TN+FP+FN}} \quad (1)$$

Sensitivity, also known as recall, refers to the percentage of labels that our computer accurately recognises as labels.

$$\boldsymbol{Sensitivity = \frac{TP}{TP+FN}} \quad (2)$$

Precision: Evaluating the total number of accurate projections allows one to assess the level of accuracy of a viewpoint. The notion under discussion is occasionally referred to as value for prediction.

$$\boldsymbol{Pr\,e\,cision = \frac{TP}{TP+FP}} \quad (3)$$

An F1 score is a measure that combines precision and recall into one score.

$$\boldsymbol{F1 - score = 2 * \frac{Pr\,ecisison * Re\,call}{Pr\,ecision + Re\,call}} \quad (4)$$

Accuracy: The algorithm has correctly classified the specificity of negatives.

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

4.9. Statistical Data Analysis

Upon completion of the efficiency assessment of SWIN, LSTM model and CNN models we performed a rigorous statistical analysis to check their predictions for any differences. This crucial step helps us in identifying which model is found to be the best when it comes to character as well as digit recognition in handwritten Urdu text. Utilizing the SPSS software application made ANOVA test possible. ANOVA which is a powerful tool of variance analysis provides insights on how different models perform relatively and helps in understanding more about their effectiveness in real world scenarios.

5. Results

In view of this, it is important to consider such key performance indicators as precision, F1-score, recall and accuracy when assessing the outcomes of new models as well as the models that are currently in use. It is therefore essential to have these metrics computed for a better understanding of how efficient the models are. The confusion matrix illustration gives a clear picture on the performance outcomes of the two models.

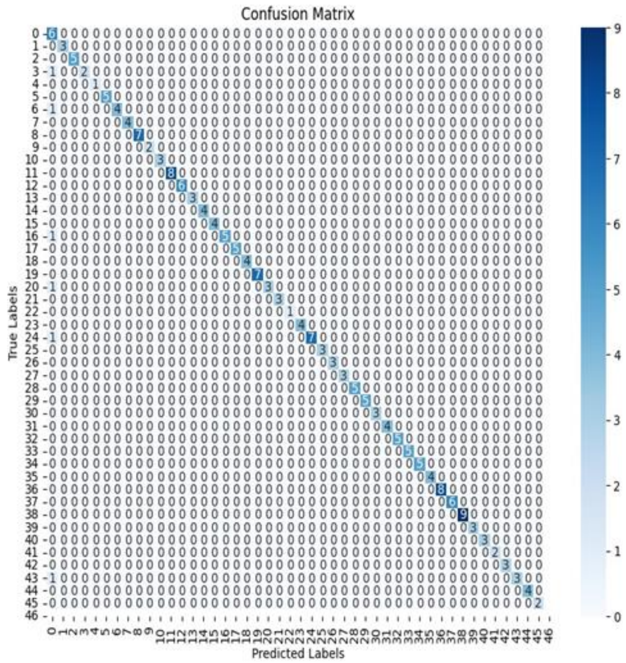

Fig9: Confusion matrix for Swin transformer

When Character and digit classification is well captured by the confusion matrix of Swin Transformer. It accurately differentiates between digits 0-9 and Urdu characters that are represented in separate classes. This way, it recognized Urdu characters as well as digits with high accuracy. Nevertheless, some misclassifications were made whereby the model incorrectly classified characters as numbers and numbers as characters. The visualization provides a summary of how well the model performed pinpointing accurate identifications and areas to be worked on.

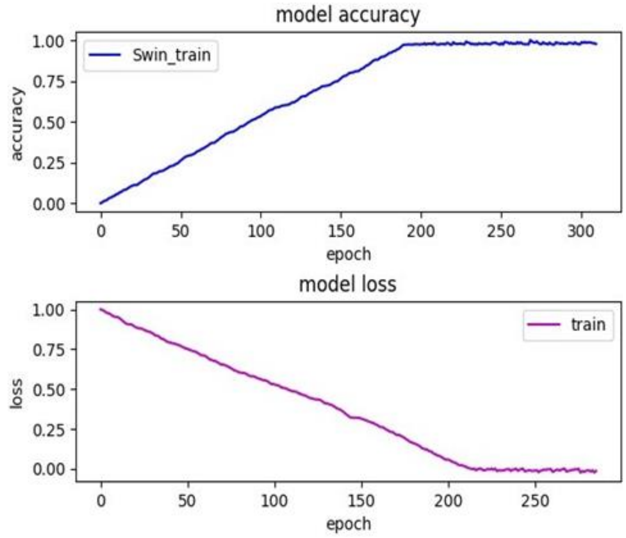

Fig10: Model Loss & Accuracy

An accuracy plot shows how good the model becomes in predicting; a loss plot, on the other hand, indicates how close predicted and actual values become. These visualizations help researchers better understand learning curves of their models to improve hyper parameters and training strategies. The graphs support the evaluation of the Swin Transformer's performance by this paper and its position in wider research context.

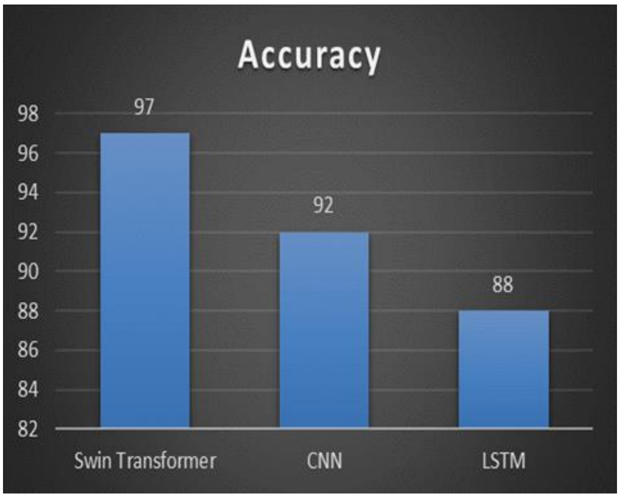

Fig11: Accuracy

The diagram above shows a plot that compares the accuracies of different classifiers, which includes Swin Transformer and some existing models like CNN and LSTM. There is an accuracy of 97% in classification by the proposed swin transformer while CNN and LSTM have an accuracy of 92% and 88% respectively. From the results, it can be seen that the proposed Swin transformer performs better than CNN and LSTM in character recognition. Compared to the present classifiers, CNN and LSTM, the provided figure shows a comparison of precision levels with regards to the suggested Swin transformer. For proper character classification, Swin Transformer obtains 83% precision rate while it is 69% for CNN and 74% for LSTM. The image clearly indicates that when identifying characters with high accuracy, proposed Swin Transformer is better than both other models cnn as well lstm in terms of this.

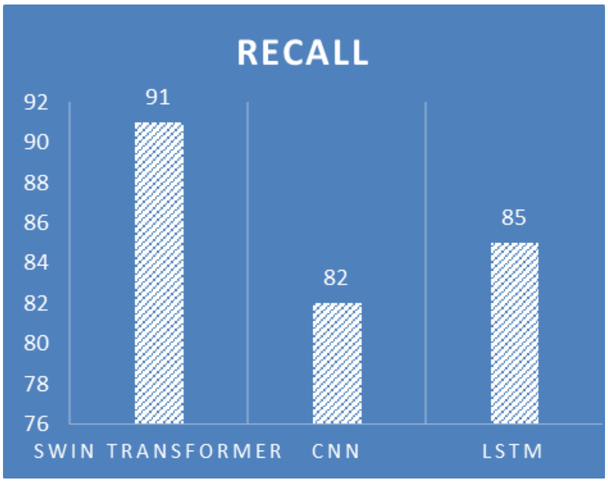

Figure 13 Recall

The plot above compares the recall of the proposed Swin Transformer and the existing classifiers CNN, LSTM. The proposed Swin Transformer has a recall of 91% in correctly identifying letters while CNN and LSTM have a recall of 82% and 85% respectively. From the figure it is clear that Swin transformer performs better than CNN and LSTM in accurately categorizing characters with higher recall.

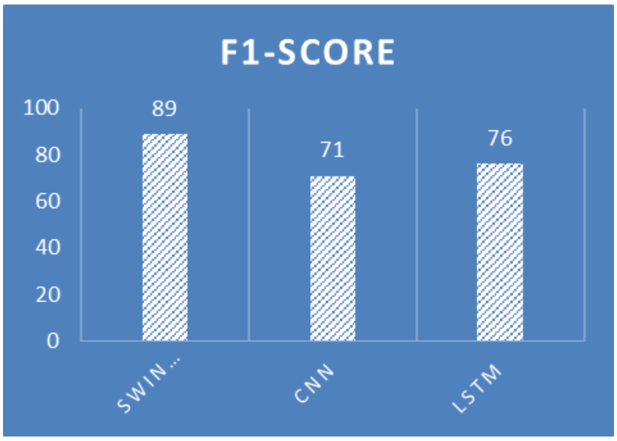

Figure 12 : F1-SCORE

The figure given compares the F1-Score plot of the Swin Transformer model and other well-known classifiers, for example, CNN and LSTM. The Swin transformer is noteworthy in terms of performance having an F1-Score of 89% in truly characterizing Characters. When comparing the results, the CNN achieves an F1-score of 71%, while the LSTM achieves 76%. This visual analysis demonstrates that the proposed swin transformer outperforms the LSTM and CNN classifiers in accuracy and consistency in classifying characters, with significantly higher F1-score values.

5.1. Statistical Analysis

H0: There is no significant difference among methods (SWIN, CNN and LSTM) in predicting the characters and digits.

For this hypothesis testing, we have considered ANOVA test, because there are more than two methods, so we utilized this test.

Table 2 Standard Deviation and Means Values

Samples data				
	N	Mean	Std. Deviation	Std. Error
SWIN	42	1.4524	.50376	.06652
CNN	42	1.7619	.43108	.07666
LSTM	42	1.5952	.49680	.07773
Total	126	1.6032	.49119	.04376

This table displays the standard deviation and means values for the SWIN, CNN, and LSTM. The CNN mean & standard deviation is 1.5952 & 0.49680, the LSTM mean & standard deviation is 1.4524 and 0.43108, the SWIN technique mean & standard deviation is 1.7619 & 0.43108, and 0.50376. The mean value is SWIN is high compared to the remaining methods.

Table 2 ANOVA

Samples data					
	N Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.016	2	1.008	4.405	.014
Within Groups	28.143	123	.229		
Total	30.159	125			

Table 3 Robust Tests of Equality of Means

Samples data				
	Statisxtic[a]	df1	df2	Sig.
Welch	4.615	2	81.560	.013

a. Asymptotically F distributed.

Welch's test is commonly used to assess whether the means of two populations are equal. It is a reliable two-sample location test. The test in Table 3 provides professional results. The test findings suggest that there is a significant variation in the average values between the projected data techniques. This difference is significant at the level of significance of 5%, with a p-value of less than 0.05.

The tests done indicate that the statistically significant value in Bonferroni and the significant test for average equal is below 0.05. By conducting meticulous analysis of strict measures for equality of means, it becomes evident that there exists a distinct disparity in the methods. Utilizing the ANOVA test, if the significance value falls below 0.05, we can confidently assert the presence of a significant difference. The null hypothesis, which posits that there are no noteworthy disparities in the ability of the SWIN, CNN, and LSTM algorithms to predict characters and digits, has been refuted. Conversely, the other theory, which proposes that there is a major difference among these methodologies, has been approved.

6. Conclusion

The study sought to assess the effectiveness of the Swin Transformer model in recognizing handwritten Urdu letters and numerals. The model underwent rigorous training using a comprehensive dataset and was then evaluated on the "MANUU: Handwritten Urdu OCR Dataset" to determine its ability to accurately handle the complexities of the Urdu script. The Swin Transformer has been shown to be outstandingly successful in multiple metrics with an accuracy of 97%, higher precision (83%) and recalls (91%) as compared to CNN and LSTM classifiers. Such performance differences have been confirmed by statistical tests such as ANOVA and Welch's test making the Swin transformer stand out as the best model in this niche. The study highlights outstanding recognition ability of hand written Urdu characters and digits in Swin Transformer – which makes it applicable in very many practical contexts while underlining its significance within the area of computational intelligence and machine learning.

7. References and Footnotes

7.1. References

Acknowledgements

We are grateful to the research team, MANUU dataset producers and reviewers for indispensably contributing to this study.

Author contributions

Shaik Moinuddin Ahmed conceptualized, did the methodology, developed software, curated data and wrote the first draft of the project. Abdul Wahid: Monitoring, authenticating, distributing funds for, appraising and revising

Conflicts of interest

The authors declare no conflicts of interest.

Dataset Availability

The dataset is available at http://ieee-dataport.org/12726

References

[1] M. A. Kilvisharam Oziuddeen, S. Poruran, and M. Y. Caffiyar, "A novel deep convolutional neural network architecture based on transfer learning for handwritten urdu character recognition," Teh. Vjesn., vol. 27, no. 4, pp. 1160–1165, 2020, doi: 10.17559/TV-20190319095323.

[2] Q. A. Safdar, K. U. Khan, and L. Peng, "A novel similar character discrimination method for online handwritten Urdu character recognition in half forms," Sci. Iran., vol. 29, no. 5, pp. 2419–2436, 2022.

[3] M. N. AlJarrah, M. Z. Mo'ath, and R. Duwairi, "Arabic handwritten characters recognition using convolutional neural network," in 2021 12th International Conference on Information and Communication Systems (ICICS), IEEE, 2021, pp. 182–188.

[4] H. Zargar, R. Almahasneh, and L. T. Kóczy, "Automatic recognition of handwritten Urdu characters," Comput. Intell. Math. Tackling Complex Probl. 3, pp. 165–175, 2022.

[5] I. Uddin et al., "Benchmark pashto handwritten character dataset and pashto object character recognition (OCR) using deep neural network with rule activation function," Complexity, vol. 2021, pp. 1–16, 2021.

[6] S. Khan and S. Nazir, "Deep learning based Pashto characters recognition: LSTM-based handwritten Pashto characters recognition system," Proc. Pakistan Acad. Sci. A. Phys. Comput. Sci., vol. 58, no. 3, pp. 49–58, 2021.

[7] R. Guha, N. Das, M. Kundu, M. Nasipuri, and K. C. Santosh, "Devnet: an efficient cnn architecture for handwritten devanagari character recognition," Int. J. Pattern Recognit. Artif. Intell., vol. 34, no. 12, p. 2052009, 2020.

[8] W. Jiang, "Evaluation of deep learning models for Urdu handwritten characters recognition," in Journal of Physics: Conference Series, IOP Publishing, 2020, p. 12016.

[9] V. K. Chauhan, S. Singh, and A. Sharma, "HCR-Net:

A deep learning based script independent handwritten character recognition network," arXiv Prepr. arXiv2108.06663, 2021.

[10] R. M. Ahmed et al., "Kurdish Handwritten character recognition using deep learning techniques," Gene Expr. Patterns, vol. 46, p. 119278, 2022.

[11] F. Anwar, "Online urdu handwritten text recognition for mobile devices using intelligent techniques." International Islamic University, Islamabad, 2019.

[12] S. S. R. Rizvi, A. Sagheer, K. Adnan, and A. Muhammad, "Optical character recognition system for Nastalique Urdu-like script languages using supervised learning," Int. J. Pattern Recognit. Artif. Intell., vol. 33, no. 10, p. 1953004, 2019.

[13] K. O. Aarif and P. Sivakumar, "Multi-Domain Deep Convolutional Neural Network for Ancient Urdu Text Recognition System.," Intell. Autom. Soft Comput., vol. 33, no. 1, 2022.

[14] H. Ali, A. Ullah, T. Iqbal, and S. Khattak, "Pioneer dataset and automatic recognition of Urdu handwritten characters using a deep autoencoder and convolutional neural network," SN Appl. Sci., vol. 2, pp. 1–12, 2020.

[15] M. M. Misgar, F. Mushtaq, S. S. Khurana, and M. Kumar, "Recognition of offline handwritten Urdu characters using RNN and LSTM models," Multimed. Tools Appl., vol. 82, no. 2, pp. 2053–2076, 2023.

[16] M. S. Amin, S. M. Yasir, and H. Ahn, "Recognition of pashto handwritten characters based on deep learning," Sensors, vol. 20, no. 20, p. 5884, 2020.

[17] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on Convolutional Neural Networks (CNN) in vegetation remote sensing," ISPRS J. Photogramm. Remote Sens., vol. 173, no. July 2020, pp. 24–49, 2021, doi: 10.1016/j.isprsjprs.2020.12.010.

[18] L. Alzubaidi et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.

[19] H. Salman, J. Grover, and T. Shankar, "Hierarchical Reinforcement Learning for Sequencing Behaviors," vol. 2733, no. March, pp. 2709–2733, 2018, doi: 10.1162/NECO.

[20] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," Eur. Phys. J. Spec. Top., vol. 228, no. 10, pp. 2313–2324, 2019, doi: 10.1140/epjst/e2019-900046-x.

[21] Sanjay Kumar Suman, Dhananjay Kumar and L. Bhagyalakshmi, "SINR pricing in non- cooperative power control game for wireless ad hoc network", KSII Transactions on Internet and Information Systems, KSII TIIS, vol. 8, no. 7, pp. 2281-2301, 2014. https://dio.org/10.3837/tiis.2014.07.005

[22] L. Bhagyalakshmi, Sanjay Kumar Suman, Sujeetha Devi, "Joint Routing and Resource Allocation for Cluster Based Isolated Nodes in Cognitive Radio Wireless Sensor Networks", Wireless Personal Communication, Springer, vol. 114, issue 4, pp. 3477-3488, Oct. 2020. https://doi.org/10.1007/s11277-020-07543-4

[23] K. Mahalakshmi, K. Kousalya, Himanshu Shekhar, Aby K. Thomas, L. Bhagyalakshmi, Sanjay Kumar Suman et. al., "Public Auditing Scheme for Integrity Verification in Distributed Cloud Storage System", Scientific Programming, Hindawi, vol. 2021, Article ID 8533995, Dec. 2021. https://doi.org/10.1155/2021/8533995

[24] Sanjay Kumar Suman et. al., Detection and prediction of HMS from drinking water by analysing the adsorbents from residuals using deep learning, Hindawi (SAGE Journal) Adsorption Science & Technology, vol. 2022, Article id 3265366, March 2022. https://doi.org/10.1155/2022/3265366.

[25] Bhagyalakshmi and K. Murugan, "Avoiding Energy Holes Problem using Load Balancing Approach in Wireless Sensor Network", KSII Transaction on Internet and Information Systems, vol. 8, no. 5, pp. 1618- 1637, 2014. https://dio.org/10.3837/tiis.2014.05.007.

[26] Satyanand Singh, Sajai Vir Singh, Dinesh Yadav, Sanjay Kumar Suman, Bhagyalakshmi Lakshminarayanan, Ghanshyam Singh, "Discrete interferences optimum beamformer in correlated signal and interfering noise, International Journal of Electrical and Computer Engineering, vol 12, no. 2, pp. 1732-1743, April 2022. http://doi.org/10.11591/ijece.v12i2.pp1732-1743: