SLAK: A NEW DATASET FOR SANTHALI HAND-WRITTEN DIGITS AND ITS CLASSIFICATION USING ALEXNET DEEP TRANSFER LEARNING

Abstract

This study explores the use of deep transfer learning with the AlexNet model to recognize hand-drawn digits in the Santhali language. Santhali Language dataset by Arvind and Kamlesh (SLAK Dataset) of 10,000 Santhali digit images was collected and pre-processed for training and testing the model. The AlexNet model was fine-tuned on the Santhali digit dataset. Different experiments were conducted to study the variation of classification accuracy for different hyper-parameters of the pre-trained architecture achieving the best classification accuracy of 99.60% on the test set. This result demonstrates the effectiveness of transfer learning for recognizing handwritten digits in a low-resource language like Santhali. The trained model can be used for various applications, such as optical character recognition systems for digitizing handwritten documents in the Santhali language.

Keywords: Hand digit recognition, Santhali language, AlexNet, Transfer learning, Deep learning, Low-resource Language

Authors

Arvind Kumar
Department of EECE
Gitam University
Bangaluru, India

Kamlesh Kishore
Department of ECE
RTC Institute of Technology
Ranchi, India

Manish Kumar
Department of ECE
RTC Institute of Technology
Ranchi, India

Ankita Kachhap
Department of ECE
RTC Institute of Technology
Ranchi, India

Mahesh Chandra
School of ECE
Reva University
Bengaluru, India

I. INTRODUCTION

Handwritten digit recognition has been a popular research topic in the field of machine learning for many years. Recently, transfer learning has emerged as a powerful technique for improving the performance of machine learning models. In this literature review, we will explore the current state of the art in the application of transfer learning for hand digit recognition. Transfer learning is the process of leveraging the knowledge learned by a machine learning model on one task to improve the performance of another related task [1]. In the case of hand digit recognition, transfer learning can be used to improve the accuracy of a model by training it on a larger dataset and then transferring that knowledge to a smaller, target dataset [2]. One popular approach for transfer learning in hand digit recognition is to use pre-trained convolutional neural networks (CNNs). CNNs are deep learning models that have shown excellent performance in image recognition tasks. The idea is to use a CNN that has been pre-trained on a large dataset, such as ImageNet, and then fine-tune it on the target dataset for hand digit recognition [3]. This approach has been shown to improve the accuracy of the model significantly. In a study by Chen et al. (2019), a pre-trained ResNet-50 CNN was used for hand digit recognition. The ResNet-50 was trained on the ImageNet dataset and then fine-tuned on the MNIST dataset. The results showed that the fine-tuned model achieved an accuracy of 99.31%, which was better than the accuracy achieved by the same model trained from scratch (98.80%). Similarly, in a study by Sabour et al.(2017), a pre-trained Inception-v3 CNN was used for hand digit recognition, achieving an accuracy of 99.7% [4]. Another approach for transfer learning in hand digit recognition is to use domain adaptation techniques. Domain adaptation is the process of adapting a machine learning model to a new domain, typically with a different distribution of data. In the case of hand digit recognition, domain adaptation can be used to transfer knowledge learned from a source dataset, such as MNIST, to a target dataset, such as USPS. In a study by Ganin et al., a domain adaptation technique called adversarial discriminative domain adaptation (ADDA) was used for hand digit recognition [5]. The approach involved training two networks: a feature extractor and a classifier. The feature extractor was trained to learn features that were invariant to the domain, while the classifier was trained to predict the digit label. The results showed that ADDA significantly improved the model's accuracy when transferring knowledge from MNIST to USPS.

1. **Related Works in Indian Languages:** Hand digit recognition for Indian languages has been an active area of research in recent years, with the aim of developing accurate and reliable methods for recognizing handwritten digits in low-resource languages. In this section, we present an overview of the recent advances in hand digit recognition for Indian languages using machine learning algorithms. In a study conducted by Ghosh et al., a dataset of handwritten digits in Bengali language was collected and used to train a convolutional neural network (CNN) model [6]. The proposed model achieved an accuracy of 98.16%, outperforming other state-of-the-art methods. Similarly, in another study by Singh et al. (2020), a dataset of handwritten digits in Devanagari language was collected and used to train a deep learning model based on the VGG architecture [7]. The proposed model achieved an accuracy of 97.22%, outperforming other methods that were compared in the study. In a recent study, a dataset of handwritten digits in Hindi language was collected and used to train a hybrid model that combined a CNN and a long short-term memory (LSTM) network. The proposed model achieved an accuracy of 99.28%, outperforming other state-of-the-art methods [7]. In addition to deep learning-based

methods, some studies have also explored the use of traditional machine learning algorithms for hand digit recognition in Indian languages. For example, in a study by Das et al., a dataset of handwritten digits in Oriya language was collected and used to train a support vector machine (SVM) classifier [8]. The proposed method achieved an accuracy of 97.8%, outperforming other traditional machine learning methods that were compared in the study. Overall, the recent advances in hand digit recognition for Indian languages using machine learning algorithms have demonstrated promising results, with high accuracy rates achieved using deep learning-based methods as well as traditional machine learning algorithms. These methods have the potential to be used in various applications, such as optical character recognition systems for digitizing handwritten documents in Indian languages.

2. **Santhali Language:** Santhali is an Austro-asiatic language spoken primarily by the Santhal people in India, Bangladesh, and Nepal [9], [10]. It is the most widely spoken Austro-Asiatic language, with over six million speakers in India alone. Santhali has its own script, the Ol Chiki script, which was developed in the early 20th century to promote literacy among the Santhal people. However, the script is still not widely used, and many Santhali speakers write in the Bengali script. The language has a rich oral tradition, with many songs and stories passed down through generations. Santhali culture also includes traditional dances, such as the Chhau dance, which is performed during festivals and other celebrations. Santhali has been recognized as one of the official languages of India since 2003, and efforts have been made to promote and preserve the language and its culture. However, like many indigenous languages, Santhali faces challenges from the dominance of majority languages and the pressures of globalization [11].

Rest of the paper is arranged as follows: Section II briefs the data-preparation process, Section III introduces Alexnet, the transfer learning algorithm. Experimental setup and result and discussions are covered under Section IV and Section V respectively. Section VI presents conclusion and future scope.

II. DATA PREPARATION

In general, data preparation has several steps based on the application [12]. Following steps were involved:

1. **Define the Scope and Purpose of the Dataset:** AK Dataset was prepared in collaboration with graduate students studying at RTC Institute of Technology at Ranchi, Jharkhand (India) to promote digital reservation and research promotion for low-resourced language to prevent its endangerment. Additionally, the result will motivate future works on continuous hand-written text recognition. This dataset will be provided to others for non-commercial activities on request to the authors.

2. **Collect the Handwriting Samples:** Around 100 educated native speakers mostly around Ranchi district, Jharkhand (India) were asked to write each of the ten digits ten times on white paper. The analog samples were digitized using a 4MP camera using a mobile handset (Redmi 9 Phone-Model MIUI version 12.0.18). Fig. 1 displays samples of collected Santhali hand-written digits.

3. **Preprocess the Data:** Preprocessing is necessary to clean up the data and prepare it for use in machine learning. This can include removing noise, cropping images, normalizing image size and resolution, and converting the data into a suitable format such as PNG or JPEG. Pre-processing (cropping and straightening) were done using Picassa Photo editor software. The RGB images were converted using Binary images using MATLAB functions as shown in Figure 2. Thresholding was done using gray threshold command which uses Otsu's method to compute the global threshold [13]. Images were resized to 227x227 pixels to fit the AlexNet model.

4. **Label the Data:** To train a machine learning model, we need to provide labels for the data. Labeling involves identifying the characters or words in the handwritten samples and associating them with their correct values. This was done manually.

5. **Split the Dataset:** To evaluate the performance of your model, we split the dataset into training, validation, and test sets. The training set is used to train the model, while the validation set is used to optimize the model's parameters. The test set is used to evaluate the model's performance on unseen data. A split ratio of 0.7 was used i.e. 7000 images for training and the rest 3000 images for testing.

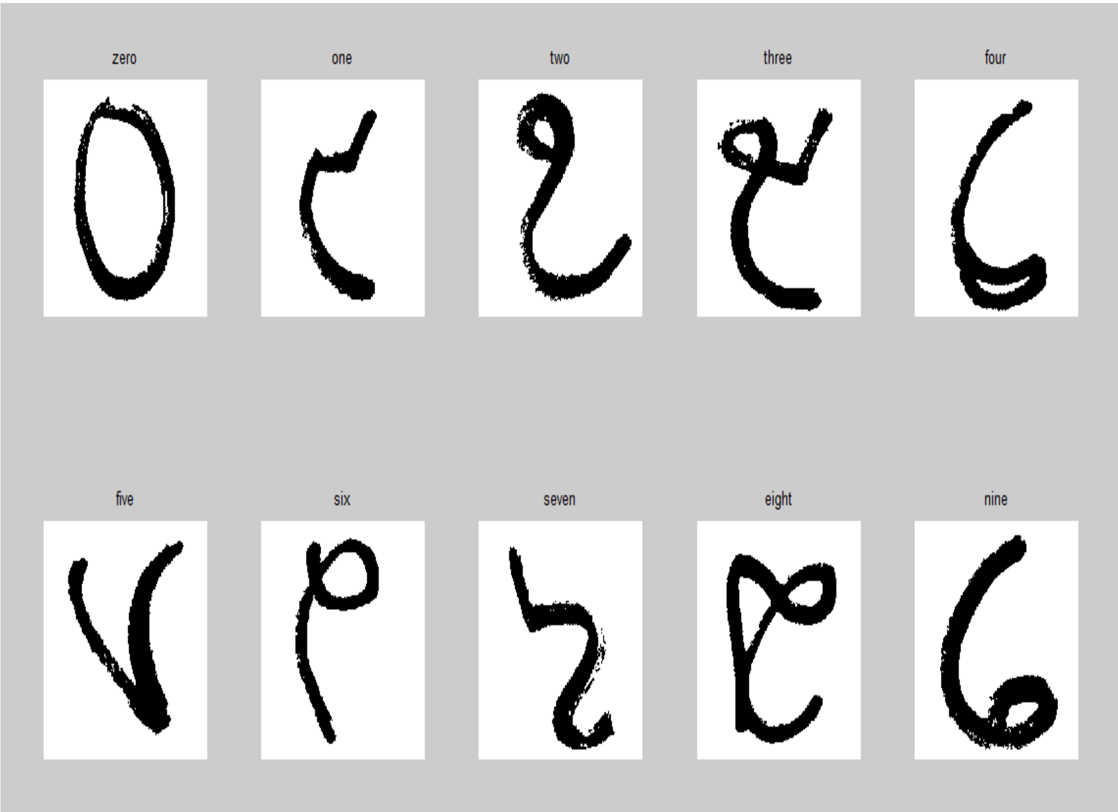

Figure 1: Santhali Hand-written digits

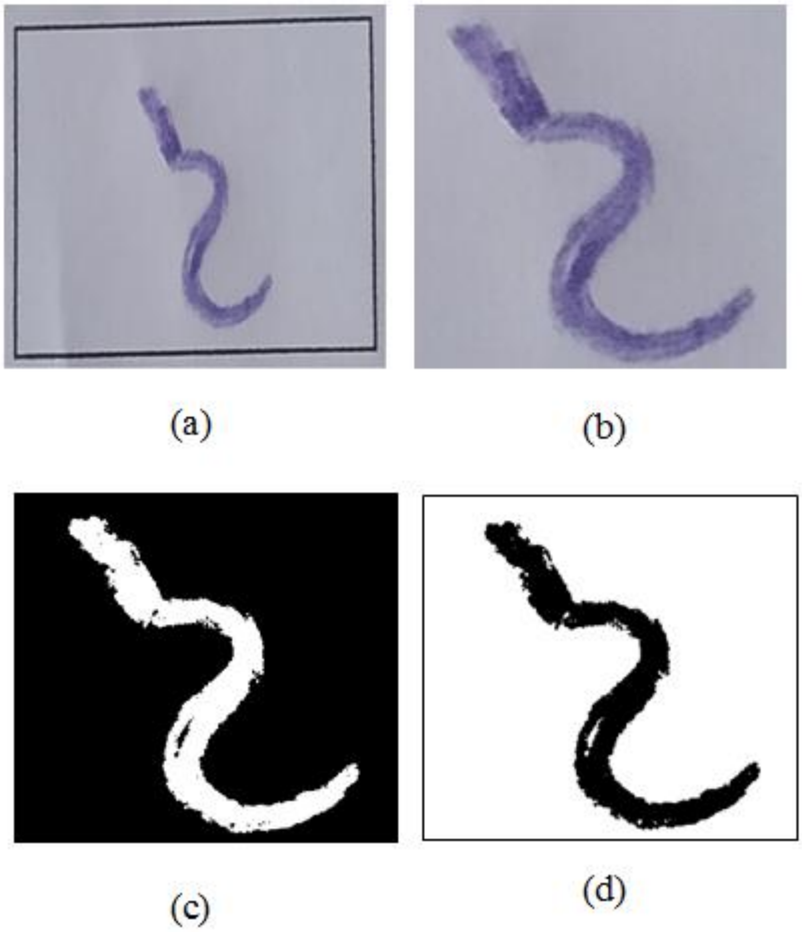

Figure 2: a) Raw Image b) Cropped image with ROI c) Binarized image d) Binarized image

III. ALEXNET ARCHITECTURE

AlexNet is a deep convolutional neural network architecture designed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton [14]. It was the winning entry in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012, which marked a breakthrough in the field of computer vision [15]. The AlexNet architecture consists of eight layers, including five convolutional layers and three fully connected layers. It uses a rectified linear activation function (ReLU) and dropout regularization to prevent overfitting. AlexNet also employs data augmentation techniques, such as random cropping and flipping, to increase the size of the training set.

1. **Alexnet for Hand-Written Digit Classification:** AlexNet can be used for hand-digit recognition by training it on a dataset of hand-written digits, such as the SLAK dataset. The architecture of AlexNet is well-suited for image recognition tasks, including hand-digit recognition [16]. To train AlexNet on the SLAK dataset, the input layer of the network would be adjusted to match the dimensions of the input images, which are 28x28 pixels. The number of output neurons in the final layer would be set to 10 to match the number of possible digit classes (0-9). The rest of the architecture can remain the same.

The training process for AlexNet on the MNIST dataset would involve feeding batches of images and corresponding labels through the network, and adjusting the weights and biases of the network using backpropagation and gradient descent. The goal is to minimize the difference between the predicted outputs of the network and the true labels. Once the network is trained, it can be used to make predictions on new hand-written digit images. The image is fed through the network, and the output of the final layer represents the probability that the image belongs to each of the 10 digit classes. The digit with the highest probability is chosen as the predicted class. Many researchers in recent years have explored AlexNet for hand-written digit classification. In ”Handwritten Digit Recognition using Convolutional Neural Network with Transfer Learning” (2021), the authors applied transfer learning to adapt the pre-trained AlexNet model to the task of hand-digit recognition on the MNIST dataset. They achieved an accuracy of 98.2% on the test set [17]. In another work, ”Deep Learning for Handwritten Digit Recognition using MNIST Dataset”, the authors compared the performance of several deep learning models, including AlexNet, on the MNIST dataset. They found that AlexNet achieved an accuracy of 98.8%, which was the highest among the models tested [17], [18]. Hence, these study were motivation for us to explore AlexNet for validation of our dataset

IV. EXPERIMENTAL SETUP

All the simulations are carried out on MATLAB 2023 platform using MATLAB Deep Network Designer App [19]. The MATLAB Deep Network Designer App is a graphical user interface (GUI) tool for designing and training deep neural networks. It allows users to interactively design and visualize neural networks, adjust network parameters, and monitor training progress. The app includes a range of pre-built neural network architectures and allows users to modify and customize them or create their own from scratch. Users can choose from a variety of network layers and can also specify activation functions, loss functions, and optimization algorithms. The app provides real-time visualization of the network architecture, training progress, and performance metrics, such as accuracy and loss. The steps to fine-tune the AlexNet model for SLAK dataset is as following:

1. The MATLAB Deep Network Designer App was launched by typing deepNetworkDesigner on the command window.
2. Once the app was launched, the AlexNet model was selected from the existing models as shown in Fig. 3. This will launch the AlexNet architecture with 25 different layers.
3. Two layers i.e. fully connected layer(fc8) and classification layer(output) were replaced as the original model was trained for 1000 classes and the proposed work has only 10 classes (0-9 digits). Fig. 4 illustrates the old and updated structure of the network.
4. The Santhali dataset was further uploaded in the next step. Data were augmented to maintain variation. The scaling factor was kept from 0.5 to 1.5, random rotation parameters were varied from -20deg to +20deg. A split ratio of 0.7 was used.
5. The next step was to set the hyper-parameters for the training process. Different experiments were conducted on varying the solver type like Stochastic Gradient Descent with momentum (sgdm), Adam and Root Mean Square Propagation (rmsprop) . The other parameters are mentioned in Table 1.

Table 1: Training option for AlexNet Architecture		
Sl. No.	**Parameters**	**Values**
1	Solver	sgdm, adam, rmsprop
2	Initial Learning Rate	0.001
3	Mini-batch size	128
4	Maximum Epoch	05
5	Validation Frequency	05

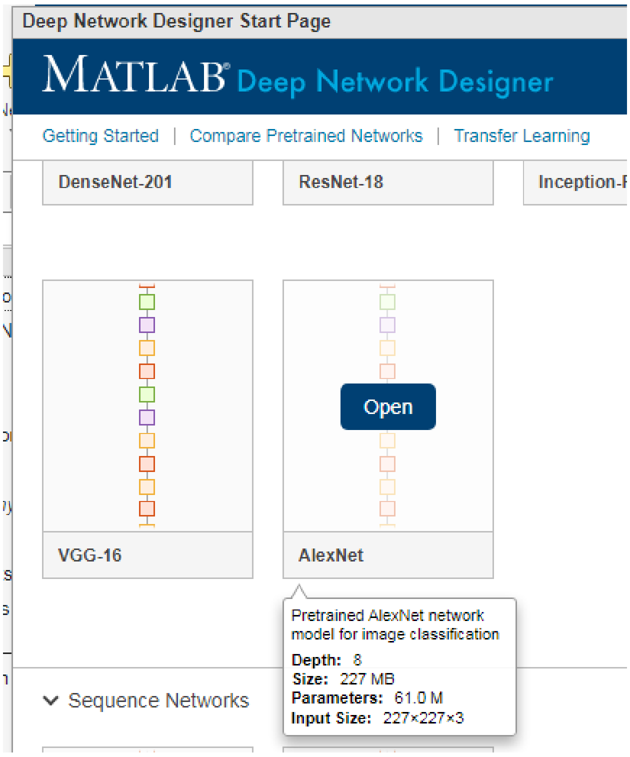

Figure 3: Deep Network Designer App

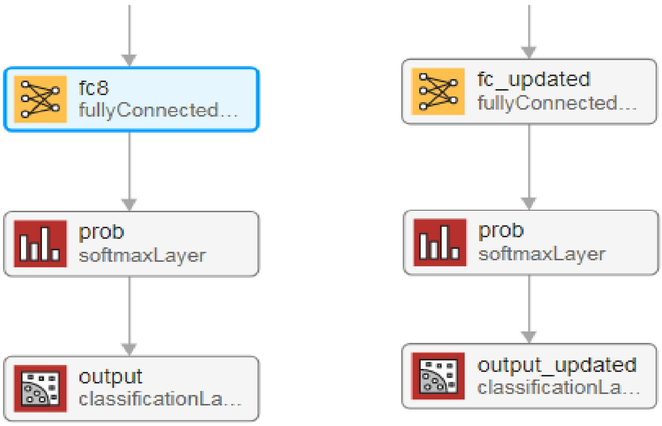

Figure 4: Modified AlexNet architecture

V. RESULT AND DISCUSSION

The focus of this work was to understand the performance of pre-trained AlexNet for a new low-resource language dataset. To evaluate the performance of a hand-written digit classification model, four different metrics were used i.e. accuracy, precision, recall, and F1 score. Accuracy measures the percentage of correctly classified samples, while precision and recall measure the proportion of true positives among the predicted positive samples and the proportion of true positives among all actual positive samples, respectively. The F1 score is a weighted harmonic mean of precision and recall, with values ranging from 0 to 1. Different experiments were carried out to observe the change in the model's performance for different solvers. Table II tabulates the results. The training process is displayed in Fig. 8. The best classification accuracy of 99.61% was seen for sgdm solver. The confusion matrix for the three experiments was illustrated in Fig. 5, Fig. 6 and Fig. 7 respectively. Similar shapes digits like four-nine and three-two reflected some cross-classification in Fig. 7.

Table II Performance Analysis for Different Solvers

Sl. No.	Solver	Accuracy	Precision	Recall	F1 score
1	Sgdm	99.60%	0.99	0.99	0.99
2	Adam	99.20%	0.99	0.99	0.99
3	rmsprop	78.90%	0.80	0.82	0.81

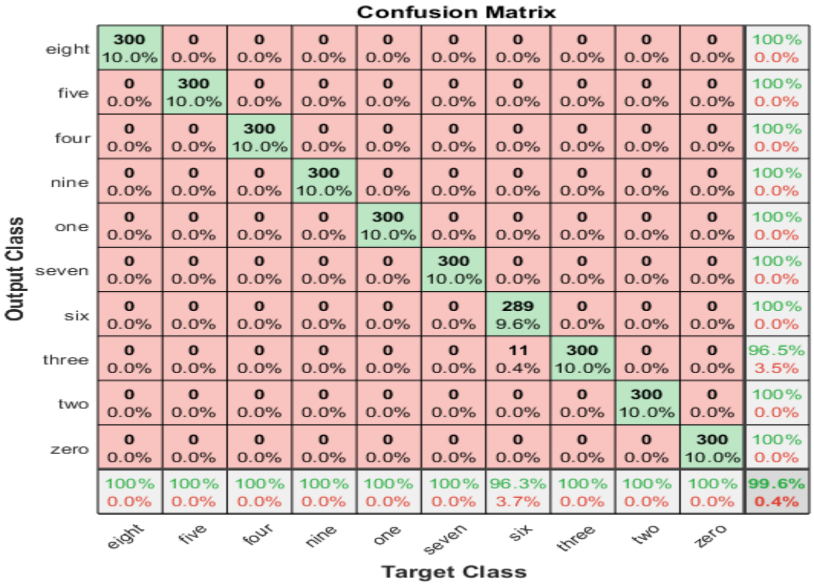

Figure 5: Confusion matrix for sgdm solver

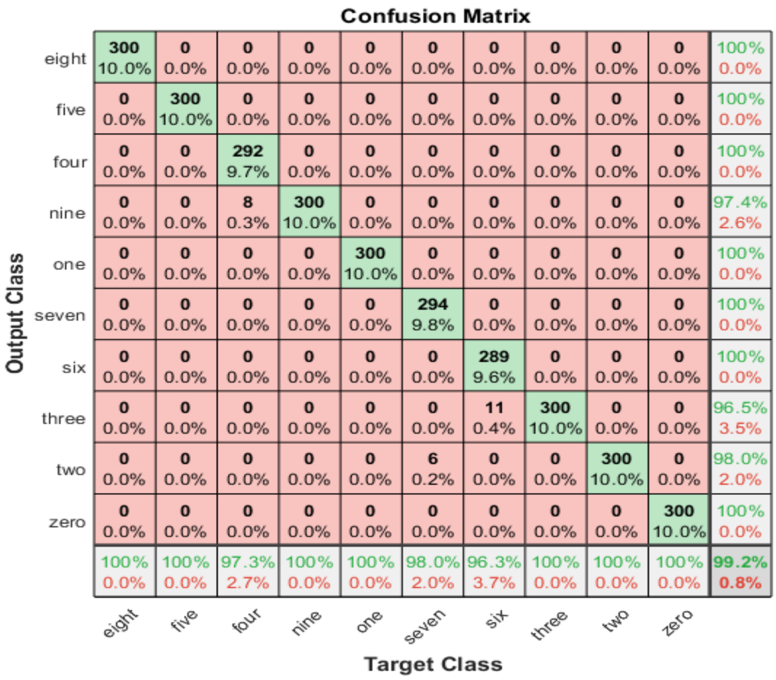

Figure 6: Confusion matrix for Adam solver

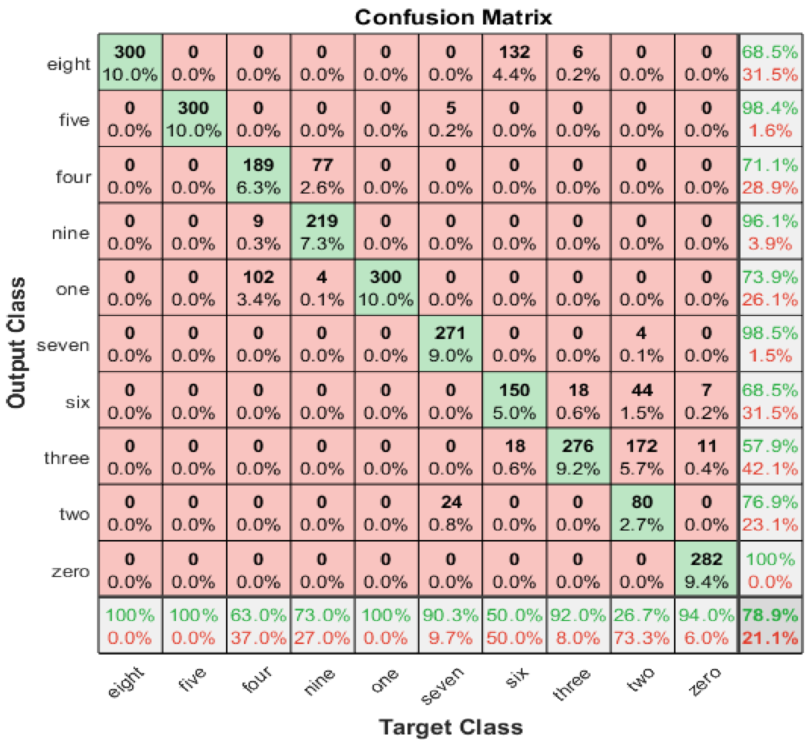

Figure 7: Confusion matrix for rmsprop solver

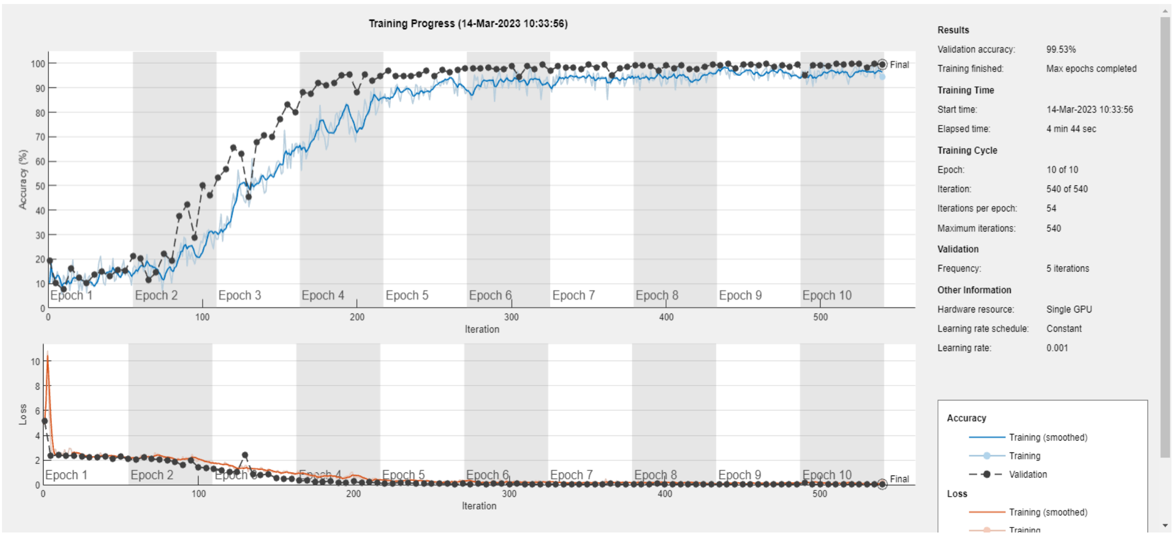

Figure 8: Training Process for Adam solver

VI. CONCLUSION

This study investigated the role of AlexNet pre-trained network for the classification of Santhali hand-digit recognition. Initially, a new dataset ”SLAK Dataset”, was created by collecting hand-written Santhali digits from 100 users with different educational backgrounds. Further, this dataset was pre-processed and separated into training and testing samples. Matlab's Deep Network Design App was then used to modify the layers of AlexNet to fit the model for ten classes. Different experiments were run to observe the classification accuracy of the models. Although the model performed brilliantly for both sgdm and Adam solver with the best classification accuracy of 99.60%, the accuracy fell to around 78.90% for rmsprop. The confusion matrix for rmsprop run reflected the mis-classification of word-pair 'two-three' and 'six-eight' as they were structurally very similar. Future work will focus on building a continuous text recognition model for Santhali digits and characters.

REFERENCES

[1] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," Journal of Big data, vol. 3, no. 1, pp. 1–40, 2016.
[2] N. Azawi, "Handwritten digits recognition using transfer learning," Computers and Electrical Engineering, vol. 106, p. 108604, 2023.
[3] L. Zhang, "A transfer learning approach for handwritten numeral digit recognition," in Proceedings of the 3rd International Conference on Software Engineering and Information Management, 2020, pp. 140–145.
[4] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules (2017)," arXiv preprint arXiv:1710.09829, 2020.
[5] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7167–7176.
[6] T. Ghosh, M.-H.-Z. Abedin, H. Al Banna, N. Mumenin, and M. Abu Yousuf, "Performance analysis of state of the art convolutional neural network architectures in bangla handwritten character recognition," Pattern Recognition and Image Analysis, vol. 31, pp. 60–71, 2021.
[7] M. Yadav, R. K. Purwar, and M. Mittal, "Handwritten hindi character recognition: a review," IET Image Processing, vol. 12, no. 11, pp. 1919–1933, 2018.
[8] A. Das, G. R. Patra, and M. N. Mohanty, "Lstm based odia handwritten numeral recognition," in 2020 international conference on communication and signal processing (ICCSP). IEEE, 2020, pp. 0538–0541.
[9] C. P. Majhi, "Santhal language and culture," Santhal Worldview. New Delhi: Indira Gandhi National Centre for the Arts, pp. 89–99, 2001.
[10] A. Kumar, R. Kumar, and K. Kishore, "Performance analysis of asr model for santhali language on kaldi and matlab toolkit," in 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). IEEE, 2020, pp. 88–92.
[11] A. K. Akhtar, G. Sahoo, and M. Kumar, "Digital corpus of santali language," in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2017, pp. 934–938.
[12] H. Kusetogullari, A. Yavariabdi, J. Hall, and N. Lavesson, "Digitnet: a deep handwritten digit detection and recognition method using a new historical handwritten digit dataset," Big Data Research, vol. 23, p. 100182, 2021.
[13] L. Jianzhuang, L. Wenqing, and T. Yupeng, "Automatic thresholding of gray-level pictures using two-dimension otsu method," in China., 1991 international conference on circuits and systems. IEEE, 1991, pp. 325– 327.
[14] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S.Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, "The history began from alexnet: A comprehensive survey on deep learning approaches," arXiv preprint arXiv:1803.01164, 2018.
[15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "Imagenet large scale visual recognition challenge," International journal of computer vision, vol. 115, pp. 211–252, 2015.
[16] R. Basri, M. R. Haque, M. Akter, and M. S. Uddin, "Bangla handwritten digit recognition using deep convolutional neural network," in Proceedings of the international conference on computing advancements, 2020, pp. 1–7.
[17] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," Artificial intelligence review, vol. 53, pp. 5455–5516, 2020.
[18] [18] D. S. Prashanth, R. V. K. Mehta, K. Ramana, and V. Bhaskar, "Hand-written devanagari character recognition using modified lenet and alexnet convolution neural networks," Wireless Personal Communications, vol. 122, pp. 349–378, 2022.

[19] MATLAB, "Transfer learning with deep network designer," https://in.mathworks.com/help/deeplearning/ug/transfer-learning-with-deep-network-designer.html, 2023 (accessed March, 2023)