Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 189 (2021) 175–182

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on AI in Computational Linguistics

Handwritten Data Digitization Using an Anchor based Multi-Channel CNN (MCCNN) Trained on a Hybrid Dataset (*h*-EH)

Abhinandan Chiney[a,b], Anwesh Reddy Paduri[a,*], Narayana Darapaneni[a,c], Santosh Kulkarni[a,c], Manish Kadam[a,b], Ishan Kohli[a,d], Malarvizhi Subramaniyan[a]

[a] *Great Learning, Pune, India*
[b] *Sandvik Materials Technology India Pvt. Ltd., Pune, 411012, India*
[c] *Northwestern University, Evanston, Illinois, 60208, USA*
[d] *Avaya India Pvt. Ltd., Pune, 411013, India*
[e] *Biz-Metric India Pvt. Ltd., Pune, 411045, India*

Abstract

To develop a holistic system for handwritten English character recognition for manually filled forms by systematically synthesising a robust handwritten textual character dataset for acceptable representation of handwriting. As part of this study, 572 copies of a form were filled by over 200 different individuals to introduce demographic variation. These forms were then scanned and each handwritten character in the forms was labelled and extracted using standard image processing techniques. The dataset of 84,712 character images created by this method (HW-dataset) comprised of both alphabetical and numerical characters. Three hybrid datasets (*h*-EH) were then formed by combining EMNIST datasets and the HW-dataset based on Digits (*h*-EHd – 329,668 character images), Alphabets (*h*-EHa – 163,085 character images) and a mixture of Digits and Alphabets (*h*-EHm – 189,586 character images). An anchor based image extraction technique was used in conjunction with a Multi-Channel CNN (MCCNN) model which was trained on three versions of *h*-EH, to automate the process of digitization of handwritten forms. The classification accuracies of the MCCNN for *h*-EHa, *h*-EHd and *h*-EHm are 93%, 96% and 93% respectively for test data. Models trained on only the EMNIST dataset perform poorly on test data. An anchor based object detection method used in conjunction with MCCNN trained on *h*-EH produces excellent results in digitising hand filled forms. Touch free solutions will gain prevalence due to the emergence of threat of fomites in the world. In such a space, manual handling of forms for the purpose of data entry, digitization and information handling will be considered as potential health and safety hazards. The solution presented in the current work uses a combination of models which is trained on a hybrid handwritten data set with high demographic variability. The model developed as part of this study is well suited for enabling touch free handling of documents.

Keywords: Hand Written Data Digitization; Anchor based Multi-Channel CNN, OCR, Image Processing, NLP, Language Science

* Corresponding author. Tel.: +91-7019986118
E-mail address: anwesh@greatlearning.in

1877-0509
10.1016/j.procs.2021.05.095

1. Introduction

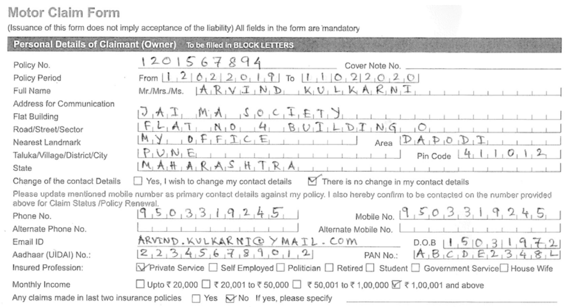
Applications of optical character recognition (OCR) include converting printed and handwritten records into digital text for storing in an electronic database. The OCR systems aim at automating the process of digitization and potentially reduce errors associated with manual entry of data. In the post pandemic world, OCR solutions will potentially lead the way into developing touch free solutions. For OCR models to become practically applicable, they need to be reliable and accurate at recognising textual characters.

Handwriting recognition is a popular problem often encountered in modern day Machine Learning and Computer Vision. The unreliability and laborious nature of manual digitization techniques have made these problems very appealing for researchers in the above mentioned fields. Researchers and investigators have looked at various methods and techniques for resolving this issue. However, it is still unresolved to the satisfaction of the community [1]. Although printed documents have been digitized, handwritten documents pose special problems. Recognising handwritten documents like forms come with unexpected challenges like handwriting type, legibility, and some inherent noise like printed bounding boxes on form documents. These make the handwriting recognition problems more complex [2]. Common variations include characters having very similar shapes, peculiar distortions which are characteristics of specific handwritings and thickness variations among written characters due to use of different writing material and instruments among many others. Even the use of different scanners alters the resolution of the images that are being used for training models [3].

Convolutional Neural Network (CNN) is one of the most commonly used Deep Learning Architectures for image processing problems. Such methods work by extracting multiple features ranging from simple features like edges and curves to more complex features like textures, automatically [4]. CNN as a result is a very popular technique for recognition of handwritings. Some investigators [5] reported an accuracy rate of *99.59%* on MNIST dataset using a modified CNN. Darmatasia and Fanany [6] proposed a combination of CNN and Support Vector Machines (SVM) to tackle the problem of handwriting recognition on form documents. The authors used CNN to extract the features from handwritten forms and then subsequently passed that information to SVM for classification of the features into alphabets and words. The authors reported an accuracy of *98.85%* on numeral characters, *93.05%* on uppercase characters, *86.21%* on lowercase characters, and *91.37%* on the merger of numeral and uppercase characters. This is reported to be an improvement over the existing stand-alone CNN-Artificial Neural Network based methods. Nasien *et.al.* [7] proposed a Freeman Chain Code with SVM to remove features from a standard NIST dataset consisting of uppercase, lowercase, and merger of uppercase and lowercase. Hussain and Vanlalruata [8] proposed a hybrid approach to extract features. They used a combination of zoning and topological feature to achieve this goal. To evaluate the performance of their model, the authors carried out an experiment using four different types of Artificial Neural Network (ANN) architectures. Multi-Channel CNNs use two or more convolutional channels to prepare different feature maps based on their learnings. The loss is then cumulatively taken into account during the Back Propagation Step (BPS) to modify weights of all the channels. This is especially useful when the production data is expected to have large amount of variations. These types of CNN have been used for textual data with great effect [9]. Al Islam and Khan [10] developed a CNN based classification algorithm for recognising Bengali numerical and mathematical expressions. YOLOv3 was used by the authors for object detection. The model was very successful achieving an end accuracy of over *98%* for numbers and mathematical symbols.

Other researchers have used Long Short Term Memory (LSTM) based methods to identify and isolate handwritten text characters. Breuel *et. al.* [11] applied bidirectional LSTM networks to the problem of machine-printed Latin and Fraktur recognition. The results presented in this paper show that the combination of line normalisation and 1D-LSTM yields excellent OCR results for both Latin/Antiqua OCR and Fraktur OCR. LSTM based methods are not popular for handwriting recognition due to their "black-box" nature.

In the present study, it is aimed to develop a system that will digitize hand filled forms and save them as digital documents for further processing as per business needs. For this purpose, a motor vehicle insurance claim form (available at `https://www.reliancegeneral.co.in/Downloads/Motor_Claim_Form.pdf`) is chosen. The present work considers the Motor Vehicle claim form filled in Upper Case letters and extracts the values against each demographic field such as Name, Address, Date of Birth among others and saves in a database. It is proposed to use a custom Object Detection Algorithm in conjunction with an image processing CNN to achieve this goal. It is additionally proposed that as a part of this work, the efficacy of standard text recognition data sets like Extended MNIST (EMNIST) be



Motor Claim Form

(Issuance of this form does not imply acceptance of the liability) All fields in the form are mandatory

Personal Details of Claimant (Owner) To be filled in BLOCK LETTERS

Policy No. 1201567894 Cover Note No.

Policy Period From 12022019 To 11022020

Full Name Mr./Mrs./Ms. ARVIND KULKARNI

Address for Communication

Flat Building JAI MA SOCIETY

Road/Street/Sector FLAT NO 4 BUILDING O

Nearest Landmark MY OFFICE Area DAPODI

Taluka/Village/District/City PUNE Pin Code 411012

State MAHARASHTRA

Change of the contact Details ☐ Yes, I wish to change my contact details ☑ There is no change in my contact details

Please update mentioned mobile number as primary contact details against my policy. I also hereby confirm to be contacted on the number provided above for Claim Status /Policy Renewal.

Phone No. 9503319245 Mobile No. 9503319245

Alternate Phone No. Alternate Mobile No.

Email ID ARVIND.KULKARNI@YMAIL.COM D.O.B 15031972

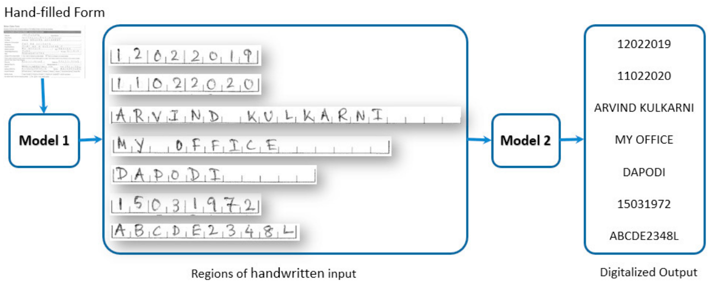
Aadhaar (UIDAI) No.: 223456789012 PAN No.: ABCDE2348L

Insured Profession: ☑ Private Service ☐ Self Employed ☐ Politician ☐ Retired ☐ Student ☐ Government Service ☐ House Wife

Monthly Income ☐ Upto ₹ 20,000 ☐ ₹ 20,001 to ₹ 50,000 ☐ ₹ 50,001 to ₹ 1,00,000 ☑ ₹ 1,00,001 and above

Any claims made in last two insurance policies ☐ Yes ☑ No If yes, please specify

(a) Sample manually filled form for digitization



(b) Information flow for the proposed combination of models

Fig. 1. Sample form and proposed information flow for model development and execution

tested in recognising handwritten data developed from a population of wide demographics in comparison to a human labelled dataset. Section 2 (Materials and Methods)elaborates the data collection process and the methods used to isolate data for training for the models. The development of the models is also described in this section. Section 3 (Results and Discussion) describes the performance of the models in successfully identifying characters and digitizing a form.

2. Materials and Methods

As explained in the Introduction, the objective of this study is to develop a holistic method to digitize a manually filled insurance form. The form type used as demonstration in this study is shown in Fig. 1(a).

Development of the digitization process is subdivided into two major model development stages - (*i.*) A model that identifies which areas in the document are handwritten and (*ii.*) An image classification model which identifies the English alphabets and numerals in the handwritten region. The outputs of these two models are then collated and stored in a separate database for further consumption as per business needs. The information flow for the proposed combination of models is shown in Fig. 1(b).

2.1. Model 1: Isolation of region of handwritten text

The foremost task in digitising a manually filled form is to identify the regions where the text is written. As the primary document to be digitized is a form in standard format, it is seen that there are regions demarcated for writing down the values of each field by hand. This provides an excellent opportunity to mark a unique anchor location on the form and use relative distances to locate the fields and corresponding handwritten input. In this case the printed letter "*M*" is used as an anchor box. This concept is shown in Fig. 2. This anchor is extracted using OpenCV's template matching functionality. This method uses a sliding window technique in which a reference patch (*T*), of size *w X h*, slides over the parent image (*I*). There are several metrics which can be used to evaluate the region of best match. In the present work, Eq. 1 is used to determine the region of best match. *x'* ranges from *0 . . . w-1* and *y'* ranges from

0 . . . h-1. Interested reader can refer OpenCV's documentation for more information on this method [12]. After the coordinates of the anchor box "*M*" are determined, field level regions of handwritten data are extracted as shown in Fig. 2.

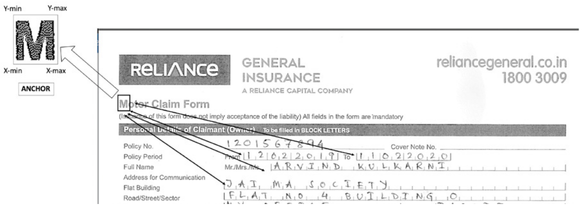


Fig. 2. Anchor based method for extracting field level region of handwritten text

$$R(x.y) = \sum_{x',y'} [T(x', y') - I(x + x', y + y')]^2 \tag{1}$$

These images cannot be easily consumed by an image classification model like Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN). For easy consumption of this extracted data, it is required that individual characters be extracted. Taking advantage of the fact that these forms are all internally consistent, it is determined using a set of 570 such forms that each character in the handwritten region is approximately 78 *X* 78 pixels. Each of the handwritten characters is then extracted. The set of typical resultant character images that is extracted to be fed into the image classification model is shown in Fig. 3.

Fig. 3. Character images obtained as output of Model 1 to be used as input to the Image Classification Model

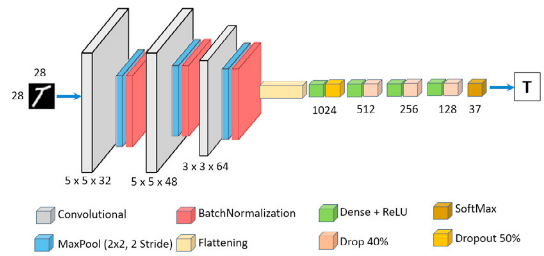
The performance of this model when combined with the model developed in Model 2: Classification of handwritten characters, is discussed in Section Digitization performance.

2.2. *Model 2: Classification of handwritten characters*

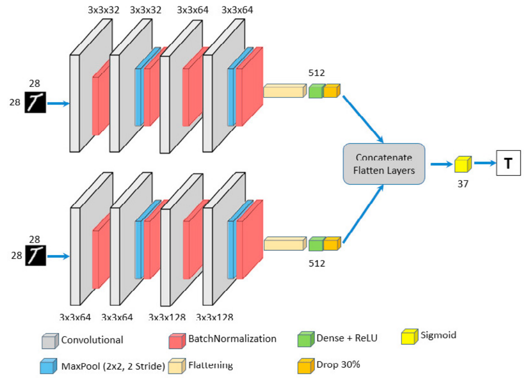
As explained in the Introduction, CNNs have emerged as leading models for image classification. These networks in their most common form are a combination of a series of layers performing a mathematical operation called convolution and dense fully connected layers that are reminiscent of ANN. ANN due to their fully connected layers, are not able to process heavy information like pixel details of images efficiently. The convolutional layers allow the images to be scaled down to manageable proportions by use of convolutions. In the present work, two variations of CNNs have been developed and evaluated – Single-Channel Conventional CNN (SCCNN) and Multi-Channel CNN (MCCNN).

2.2.1. *Single-Channel CNN*

The architecture of the SCCNN developed as a part of this work is shown in Fig. 4(a). This network has a total of $827,510$ trainable parameters and 290 non-trainable parameters. The input to this network is normalised pixel values of the extracted character images. The network contains three convolutional layers, each with their own 2D MaxPooling layer having filter shape $(2,2)$ with stride length of two and Batch Normalisation layer. The output of the convolutional layers is flattened for consumption of the subsequent fully connected layers. There are four fully connected layers in this network. The output layer has thirty-seven neurons to account for ten numerical characters (0–9), twenty-six upper case alphabets (A-Z) and one character for "*BLANK*" entries in the forms. ADAM optimiser [13] is used along with a categorical cross entropy loss function for weight modification as part of the Back Propagation process. This simple network is then trained on three datasets to evaluate its efficacy. The methodology used to create these datasets is explained in the following discussion (Datasets).



(a) Single-Channel Convolutional Neural Network



(b) Multi-Channel Convolutional Neural Network

Fig. 4. Architectures of Convolutional Neural Network used in this study

2.2.2. Multi-Channel CNN

A MCCNN used in this work has two or more "channels" or "streams" of convolutional layers. Each of the channels, take different instances of the same class and produces feature maps of their own. These feature maps are then concatenated and used for loss calculation and optimisation of weights. There are no restrictions on the individual channels in this method. The channels can be identical or different - the only restriction is that the input class should be the same for all the channels. As different instances of the same class is fed to the model simultaneously, it is believed that the feature recognition ability of the model is enhanced greatly when compared to a SCCNN. Since handwritten data has very high variation in style and pattern, MCCNNs are expected to outperform SCCNNs for such data. This is discussed in the Classification performance section.

In the present work, a MCCNN is developed, trained and used with the architecture shown in Fig 4(b). This model has $1,939,013$ trainable parameters and $3,200$ non-trainable parameters. The input to each of the channels of this network is normalised pixel values of the extracted character images. Each channel is like a standard architecture of CNN. The output of the flattening layer of each channel is concatenated and then sent to 37 nodes output layer. ADAM optimiser is used along with a categorical cross entropy loss function for weight modification as part of the Back Propagation process.

2.3. Datasets

As a part of this study, a standard dataset (Extended MNIST - [14]) is tested for its efficacy in training a handwriting based character classification model in comparison with a hand labelled dataset. A combination of the above datasets is also tested. The EMNIST dataset is a set of handwritten character digits derived from the NIST Special Database 19 and converted to a $28X28$ pixels image format. The parent NIST Special Database 19 contains NIST's entire corpus of model training materials for handwritten document and character recognition. Out of the 6 datasets available as a part of EMNIST, for evaluation and benchmarking of the two models, EMNIST Balance dataset is used. This set contains $131,600$ characters with 47 classes. 10 numerical, 26 upper case alphabets and 11 lower case alphabets make up the dataset. In this dataset, some of the lower case alphabets are merged with the upper case alphabets. This is done to address an interesting problem in the classification of handwritten digits, which is the similarity between certain uppercase and lowercase letters. The motivation to choose the balanced dataset over other more populous datasets is to prevent inherent bias. An unbalanced dataset often leads to a biased and overfit model. As such forms are expected to only have upper case alphabets, this dataset is truncated to 36 classes after removing the classes for lower case. One additional class is added to account for "BLANK" cells in the forms.

A second dataset is also created for evaluation and training of the model (HW-dataset). For the creation of this dataset, 570 forms are hand filled by about 200 different individuals to account for natural variation. Each handwritten character in these scanned forms are then labelled using open source software LableImg ([15]). Using the coordinates in the labels, these character images are extracted. The resulting image at this stage is similar in shape and size to the images obtained as the output of Model 2: Classification of handwritten characters. To make these consistent with the standard EMNIST dataset, two changes are performed on them - the colours of the extracted images are inverted to achieve the black background and white text of EMNIST and the images are resized to $28X28$ pixels. It is seen that the form has three category of field classes in which data is hand filled - (*i.*) Alphabet only classes like Name, State, Area, (*ii*) Numerical only classes like Phone numbers, date fields like Date of Birth and (*iii.*) Alphanumeric fields like PAN number, Road/Street/Sector.

To account for these three kinds of data, three different hybrid datasets are created by combining the HW-dataset with the corresponding EMNIST dataset. The set of numerical character images obtained as part of the HW-dataset is combined with the EMNIST Digits dataset. This dataset is named *h*-EHd. The set of character images of alphabets obtained as part of the HW-dataset is combined with the EMNIST Letters dataset.This dataset is named *h*-EHa. To create an alphanumeric dataset, the whole HW-dataset is combined with the EMNIST Balanced dataset (*h*-EHm). Additionally "BLANK" is added to each other three datasets. These three datasets are collectively referred to as *h*-EH dataset.The details of *h*-EH dataset is given in Table. 1.

Dataset for model testing. For evaluation of the two CNN models and the efficacy of the training data, $14,285$ test images from the HW-dataset which is a true representation of the production data is used. This also helps gauge the performance of the three sets of data. For the remainder of this document, this dataset is referred to as HW-test dataset.

Table 1. Details of the *h*-EH dataset

Dataset Description	Dataset Name	Training images	Validation images	Testing images
Alphabets + Numbers	*h*-EHm	$158,864$	$16,437$	$14,285$ (HW-test dataset)
Numbers	*h*-EHd	$281,230$	$42,037$	$6,401$
Alphabets	*h*-EHa	$132,850$	$22,813$	$7,422$

3. Results and Discussion

3.1. Dataset performance

The SCCNN developed as part of this study (Section Single-Channel CNN) is used to benchmark the performances of the different datasets (Section Datasets). The outcome of the study is summarised in Table. 2. It is seen that the model trained on the Balanced EMNIST dataset performed very well on the validation dataset which is comprised of images only from the EMNIST validation dataset. However, when this model is tested on the HW-test dataset which is representative of the true production dataset, the performance of the model drops significantly. This same SCCNN architecture when trained on HW-dataset performs much better, though not to the satisfaction of the authors. However, *h*-EH dataset which is a combination of the EMNIST dataset and the HW-dataset performs the best as a training set. This shows that the EMNIST dataset is not representative of the variabilities encountered in a handwritten dataset spread across different demographics. This also shows that although the EMNIST dataset cannot capture all the variations in the features of a dataset sourced from varying demographics, it does enhance the performance of the model when combined with a more representative dataset.

Table 2. Performance of EMNIST, HW-datset and h-EH datasets [Note: Validation data consists of images from only EMNIST and Testing data consists of images from only HW-test dataset]

Training Data	Accuracy			Testing		
	Train	Validation	Testing	F1-Score	Recall	Precision
EMNIST (Balanced)	93	92	21	0.29	0.21	0.48
HW-dataset	83	81	72	0.70	0.68	0.72
h-EHm	90	91	82	0.76	0.75	0.78

3.2. *Classification performance*

Three instances of the model architecture (Fig. 4(b)) are trained on h-EHa, h-EHm and h-EHn to account for the productions data types as discussed in Section.Datasets. The test data for evaluating these models are derived from the HW-dataset which is a true representation of the production dataset. The performance of the two CNN based models trained on h-EHm is shown in Fig. 5. It is seen that MCCNN trained on h-EHm outperforms SCCNN trained on the same dataset on every metric tested for. The reason for this is outlined in Section. Multi-Channel CNN. The classification accuracy for the other instances of the MCCNN model trained on hybrid datasets (h-EHa and h-EHn) are given in Table.3. The performance of MCCNN is excellent across all datasets tested for as part of this study.

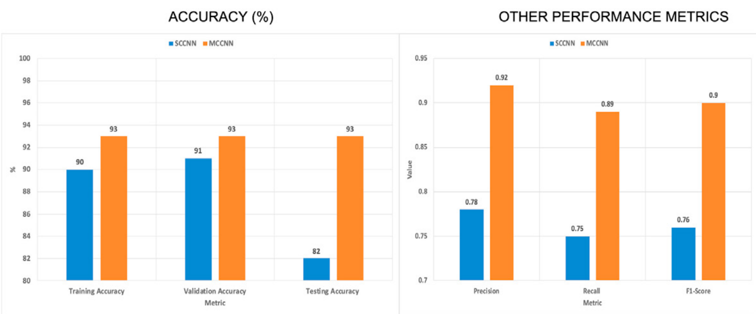


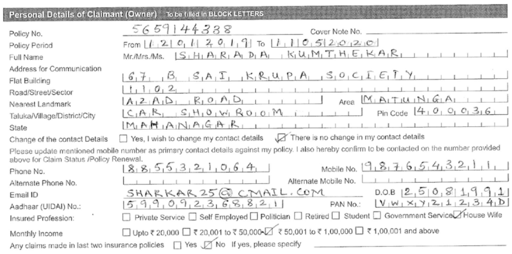
Fig. 5. Comparison of performance of SCCNN and MCCNN on h-EHm test dataset [Note: Validation data consists of images from only EMNIST and Testing data consists of images from only HW-test dataset]

Table 3. Performance of MCCNN for h-EHa and h-EHn datasets [Note: Validation data consists of images from only EMNIST and Testing data consists of images from only HW-dataset]

Training Data	Training accuracy	Validation accuracy	Testing accuracy
h-EHa	96	96	93
h-EHn	99	99	96

3.3. *Digitization performance*

The output of Section. Model 1: Isolation of region of handwritten text, is fed as the input to the Multi-channel CNN based Classification model discussed in Section. Multi-Channel CNN for character identification. The performance of this Anchor based Multi-Channel CNN (MCCNN) trained on a Hybrid Dataset is shown for one of the test forms in Fig. 6. It is seen that this combination of models performs exceedingly well in digitising handwritten forms. The time for execution of this method to digitize one such form on an 9th Generation Intel Core i7 processor is about 30 seconds.



Personal Details of Claimant (Owner) To be filled in BLOCK LETTERS

Policy No. 5659144338 Cover Note No.
Policy Period From 12 01 2019 To 11 05 2020
Full Name Mr./Mrs./Ms. SHARADA KUMTHEKAR
Address for Communication
Flat Building 67 B SAI KRUPA SOCIETY
Road/Street/Sector 1102
Nearest Landmark AZAD ROAD Area MATUNGA
Taluka/Village/District/City CAR SHOWROOM Pin Code 400036
State MAHANAGAR
Change of the contact Details ☐ Yes, I wish to change my contact details ☑ There is no change in my contact details
Please update mentioned mobile number as primary contact details against my policy. I also hereby confirm to be contacted on the number provided above for Claim Status /Policy Renewal.
Phone No. 8855321064 Mobile No. 9876543211
Alternate Phone No. Alternate Mobile No.
Email ID SHARKAR25@CMAIL.COM D.O.B 25081991
Aadhaar (UIDAI) No.: 599092368821 PAN No.: VWXY21234D
Insured Profession: ☐ Private Service ☐ Self Employed ☐ Politician ☐ Retired ☐ Student ☐ Government Service ☑ House Wife
Monthly Income ☐ Upto ₹ 20,000 ☐ ₹ 20,001 to ₹ 50,000 ☑ ₹ 50,001 to ₹ 1,00,000 ☐ ₹ 1,00,001 and above
Any claims made in last two insurance policies ☐ Yes ☑ No If yes, please specify

policyfrom	12012019
policyto	11052020
name	SHARADA KUMJHEKAR
flat	67 B SAI KRUPA SOCIETY
road	1102
landmark	AZAD ROAD
area	MATWNGA
city	CAR SHOWROOM
state	MAMANAGAR
pin	400036
phoneno	8855321064
mobileno	9876543211
alternatephoneno	
alternatemobileno	
adhar	599092368821
dob	25081991
pan	VNXY212340

Fig. 6. Performance of the combination of models in digitization of hand-filled forms

4. Conclusions

In the present work, a holistic Anchor based text isolation method followed by Multi-Channel CNN model is developed to digitize hand-filled forms. A representative hybrid dataset of character images is created for the purpose of training the models. This dataset is found to be more representative of varied demographics as compared to a standard dataset like EMNIST. It is also demonstrated that a Multi-Channel CNN outperforms a conventional Single-Channel CNN on all evaluation parameters. The combination model developed as part of this study does a very good job in digitization of the hand-filled forms and is well suited to enable touch free handling of documents.

References

[1] Shruthi. and M S Patel. Offline handwritten word recognition using multiple features with svm classifier for holistic approach. 2015.

[2] A. Jindal and M. Amir. Automatic classification of handwritten and printed text in icr boxes. In *2014 IEEE International Advance Computing Conference (IACC)*, pages 1028–1032, 2014.

[3] Nisha Sharma, Tushar Patnaik, and Bhupendra Kumar. Recognition for handwritten english letters : A review. 2013.

[4] Yann Lecun and Yoshua Bengio. *Convolutional networks for images, speech, and time-series*. MIT Press, 1995.

[5] L. Chen, S. Wang, W. Fan, J. Sun, and N. Satoshi. Reconstruction combined training for convolutional neural networks on character recognition. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 431–435, 2015.

[6] Darmatasia and M. I. Fanany. Handwriting recognition on form document using convolutional neural network and support vector machines (cnn-svm). In *2017 5th International Conference on Information and Communication Technology (ICoIC7)*, pages 1–6, 2017.

[7] D. Nasien, H. Haron, and S. S. Yuhaniz. Support vector machine (svm) for english handwritten character recognition. In *2010 Second International Conference on Computer Engineering and Applications*, volume 1, pages 249–252, 2010.

[8] J. Hussain and Vanlalruata. A hybrid approach handwritten character recognition for mizo using artificial neural network. In *2018 International Conference on Advanced Computation and Telecommunication (ICACAT)*, pages 1–6, 2018.

[9] Bao Guo, Chunxia Zhang, Junmin Liu, and Xiaoyi Ma. Improving text classification with weighted word embeddings via a multi-channel textcnn model. *Neurocomputing*, 363:366 – 374, 2019.

[10] Md Nafee Al Islam and Siamul Karim Khan. HishabNet: Detection, Localization and Calculation of Handwritten Bengali Mathematical Expressions. *arXiv e-prints*, art. arXiv:1909.00823, September 2019.

[11] T. M. Breuel, A. Ul-Hasan, M. A. Al-Azawi, and F. Shafait. High-performance ocr for printed english and fraktur using lstm networks. In *2013 12th International Conference on Document Analysis and Recognition*, pages 683–687, 2013.

[12] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[13] D.P. Kingma and L.J. Ba. Adam: A method for stochastic optimization. In *2015 International Conference on Learning Representations (ICLR)*, 2015.

[14] G. Cohen, S. Afshar, J. Tapson, and A. van Schaik. *EMNIST: an extension of MNIST to handwritten letters*, 2017 (accessed March 23, 2020). Retrieved from http://arxiv.org/abs/1702.05373.

[15] Tzutalin. Labelimg. Free Software: MIT License, 2015. URL https://github.com/tzutalin/labelImg.