Convolutional Neural Networks for Handwritten Text Recognition of Medical Prescription

Makarand Shahade, Mayuri Kulkarni, Vivek Pawar, Jatin Chaudhari, Yash Lakade, Darshan Kotkar
SVKM's Institute of Technology
Dhule, India
{manu1509.shahade@gmail.com}
{mayuridkulkarni@gmail.com }
{vivekpawar189@gmail.com}
{jatinchaudhari006@gmail.com}
{yashlakade2002@gmail.com}
{darshankotkar123@gmail.com}

Journal of Digital
Information Management

ABSTRACT: *Converting handwritten prescriptions into electronic format offers several advantages and is crucial for modern healthcare systems. It is essential nowadays because of some factors such as – Legibility and Accuracy: Handwritten prescriptions can be challenging to read and interpret; accessibility and Portability: Electronic prescriptions can be easily stored; Decision Support Systems: By digitising prescriptions, healthcare systems can integrate them with electronic health records (EHRs) and utilise decision support systems. Convolutional neural networks (CNNs) are a class of deep learning algorithms that have proven effective in extracting handwritten text from various documents, including medical prescriptions. By leveraging CNNs for handwritten text extraction, healthcare systems can automate the process of digitising prescriptions, reducing manual effort and potential human errors. This enables seamless integration with electronic systems, facilitating better patient care and overall healthcare management. In this paper, we have trained the CNN model for different parameters and observed the accuracy and loss for various parameters. We got a maximum training accuracy of 89% and a maximum testing accuracy of 70%.*

Subject Categories and Descriptors: J 3[Medical Information Systems]; I.7.5 [Document Capture]. I.5 [PATTERN RECOGNITION]: Neural Nets

General Terms: Handwritten Characters, OCR, Text recognition, Medical Records, Electronic Health Records, Neural Networks

Keywords: Handwritten Characters, Text recognition, Neural Networks, Electronic Health Records, Medical Prescription

Received: 28 April 2023, Revised 5 July 2023, Accepted 18 July 2023

Review Metrics: Review Scale: 0/6, Review Score: 5.02, Inter-reviewers Consistency: 87.5%

DOI: 10.6025/jdim/2023/21/4/117-124

1. Introduction

Electronic Health Records (EHRs) [1] are computerised representations of patients' medical records, including in-depth details about their medical history, diagnoses, treatments, medications, and other pertinent information. EHRs attempt to increase healthcare information efficiency, accuracy, and accessibility by switching to electronic systems from the more common paper-based ones [2]. Here are a few factors emphasising the significance of EHRs:

Accessibility and Continuity of Care: Healthcare professionals may securely and swiftly access patient information thanks to EHRs, no matter where or when they need to. This makes it easier for the various healthcare professionals involved in a patient's treatment to coordinate and provide continuity of care.

Patient Safety and Quality of Care: EHRs offer a centralised and comprehensive view of a patient's health data, lowering the possibility of medical errors, such as drug conflicts or overprescribing. It helps healthcare professionals give individualised care based on the patient's needs and medical history.

Efficiency and Cost-Effectiveness: EHRs reduce paperwork and administrative stress by streamlining administrative chores, including record-keeping, appointment scheduling, and billing [3]. As a result, efficiency is increased while manual record management expenses are decreased.

The digitalisation of medical prescriptions plays a crucial role in maintaining EHRs. By converting handwritten prescriptions into electronic format, the prescription information can be easily integrated into the patient's EHR [2]. This ensures accurate and up-to-date medication records, facilitates medication reconciliation, and enhances patient safety.

Deep learning [4], a subset of machine learning, has revolutionised various fields, including healthcare. It involves training neural networks with multiple layers to learn and extract complex patterns from data automatically. Deep learning techniques, particularly convolutional neural networks (CNNs), are vital in extracting handwritten text from medical prescriptions. Here's why:

Accuracy in Text Extraction: CNNs excel in image recognition tasks and can accurately identify and extract handwritten text from prescription images. By training CNN models on large datasets of handwritten samples, they can learn handwritten text's intricate patterns and structures, improving text extraction accuracy.

Flexibility and Adaptability: CNNs can handle variations in handwriting styles, sizes, and orientations. They can generalise well to different prescription images, making them suitable for processing a wide range of handwritten prescriptions.

Speed and Efficiency: CNNs can rapidly process prescription images, enabling real-time or near-real-time text extraction. This expedites the conversion of handwritten prescriptions into electronic format, enhancing the overall efficiency of maintaining EHRs.

The future scope of extracting handwritten text from medical prescriptions using CNNs involves continuous advancements in deep learning techniques and increasing the availability of large, diverse, and well-annotated datasets. This can further improve the accuracy and generalisation capabilities of the models.

Additionally, advancements in natural language processing (NLP) techniques can be integrated to extract more comprehensive information from the prescription text, such as dosage instructions, medication names, and patient-specific details. This would enhance the utility of EHRs and improve the automation of clinical workflows. Furthermore, integrating computer vision techniques with CNNs can extract additional information from prescriptions, such as prescription signatures, stamps, or other relevant annotations.

Overall, the future lies in developing robust and versatile deep learning models, leveraging larger datasets, and integrating complementary technologies to enhance the extraction of handwritten text and other relevant information from medical prescriptions, thus further advancing the maintenance and utility of EHRs.

In this paper, we have defined a Convolutional Neural Network for recognising Medicine names from the paperbased prescription for which we have collected real-world data from doctors and clinicians. The training and testing on the CNN model are performed on the collected image samples, and the model's performance is evaluated based on loss and accuracy. In the future section, we discuss our work during training, testing, and evaluation of the model.

2. Early Work

Optical character recognition is a significant topic for machine learning and deep learning study (OCR) [4]. With the help of machine learning (ML) and deep learning (DL) algorithms like SVMs, HMMs, CNNs, RNNs, and LSTMs, it is possible to identify text and extract useful information accurately. This is known as handwritten text recognition (HTR) and handwritten word recognition (HWR).

The ongoing effort in OCR has many applications. Still, this paper focuses on how we incorporate HWR and HTR techniques in the healthcare application, i.e., to maintain the electronic health record (EHR), as well as how precisely and effectively we can extract the information in a prescription. To maintain EHR, improve the accuracy, and increase the efficiency with which we can extract information from the medical prescription. This paper discusses how OCR techniques can extract text from paperbased medical prescriptions successfully.

The current section refers to earlier work in HTR and HWR for prescriptions. The compilation that follows includes some of the significant recent work.

The proposed solution [5] offers an efficient method to extract handwritten text separately from printed text in one of the latest works by the paper's author that concentrates on techniques to classify or separate printed and handwritten texts. The handwritten text extracted here offers useful tools for future research because it includes details about the medications and the patient's dosages.

The author of the paper [6] proposes a three-layer method that uses convolution layers to translate each character in the recognised sequence into an alphabetic character using a CTC loss function. To determine the end probabilities of words for HWR, the bi-LSTM layer is followed by a linear layer. In the third layer, string matching is used to find the closest term in the corpus. The corpus mainly consists of medical terminologies to improve the precision of text extraction.

One of the latest systems proposed in [7] captures text through a mobile application using CNN and OCR methods. Although the system is implementable, its accuracy, which is in the range of 70%, prevents it from achieving greater accuracy.

Deep learning-based methods have shown to be significant for OCR techniques, implying that they might additionally be effective for HWR and HTR. The text from the input picture can be divided into 64 predefined sets of classes for text recognition using artificial neural networks, according to the author of one of the most recent papers [8]. Deep Convolutional Recurrent Neural Networks train the supervised machine learning model in the proposed system. The proposed system's output displays impressive outcomes with 98% accuracy.

The paper's [9] author suggested utilising machine learning to recognise doctors' handwriting to generate digital prescriptions. "Handwritten Medical Term Corpus" is a dataset that includes 17431 samples of 480 medical terms. A Bidirectional Long Short-Term Memory (LSTM) network is the foundation of the proposed system, and eight alternative combinations are used to assess the effectiveness of the suggested approach. Utilising Rotating, Shifting, and Stretching (RSS) and Bidirectional LSTM for data augmentation, the system achieved an average accuracy of about 93.0%.

Balci et al. used two approaches: classifying words directly and character segmentation. using CNN and LSTM. [10] Khandokar, M et al. investigated CNN's capability to recognise the characters from the image dataset and the recognition accuracy with training and testing. [11]. Combining CNN for classification and Error error-correcting output Code Classifier yields better accuracy than the conventional CNN. [12] deep recurrent neural networks with LSTM, for which the authors used Gated Fully Convolutional Network architecture. [13]

We can infer from the papers above that deep learning and machine learning techniques efficiently create a system to recognise the text from prescriptions. The research mentioned above offers HTR and HWR implementation methods. The papers also outline the areas that could use improvement and how the suggested systems might soon prove helpful.

3. Methods and Materials

In the proposed Model discussed in this paper, we will train the Convolution Neural Network (CNN) model on the images of medicine names extracted from the paper-based prescription. The image pre-processing is performed on the collected sample images that get converted to a format suitable for training. Image Pre-processing is followed by the processing phase, which includes defining the CNN model and model training.

Out of the sourced data, 80% is collected for training and 20% for testing.

3.1. Data Collection

Data is collected from various clinics in the city. Doctors of a particular clinic are requested to write prescriptions three times, each containing 15 medicines that physicians nationwide must prescribe. The number of doctors we reached was 10, and the number of collected prescriptions for training was around 30. Also, each doctor writes a new prescription as they prescribe to the patient, which is used for model evaluation.

The prescriptions were written using a pen of variable tips so the trained model would be more generalised. After collecting the required data, some manual pre-processing is done on this data, like cropping the individual medicine names and renaming the image file. The image format used here is '.PNG'. In total, 30 prescriptions contained 15 medicine names, as shown in Figure 1. In total, the number of samples is equal to 450, of which 360 samples are used for training & remaining 90 samples are used for testing.

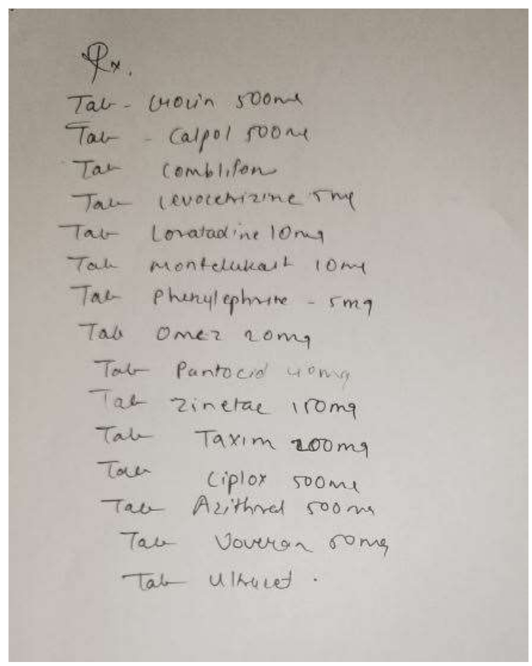

Figure 1. Sample Prescription for data collection

3.2. Training Data and Testing Data Preparation

Training data: The sample data images were collected in the medicine data folder. The sample image is given a name according to sequence from 1 – 450, for example, 1.PNG, 2.PNG, and so on, as shown in Figure 2(a).

After storing the images in the folder, the image labels with corresponding medicine names separated by a comma (,) are written in an 'annotation.txt' file. This file is used while training the CNN model to map an image to the corresponding labels, as shown in Figure 2(b).

Testing Data: The images selected for testing are added to the folder named 'test-data' with image names given by sequence numbers from 1 to 90. The test images are also stored in '.PNG' format, as shown in Figure 2(c). Like the containing test image label and respective medicine name, as shown in figure 2(d), this 'test-data.txt' file will be used to compare the performance of the CNN model based on the predicted labels and actual labels, which are mentioned in the 'test-labels.txt' file.

After performing the above process on the training and testing data set, we must import that data into a Python program to perform further actions. A dictionary data structure is used to store the image filenames.

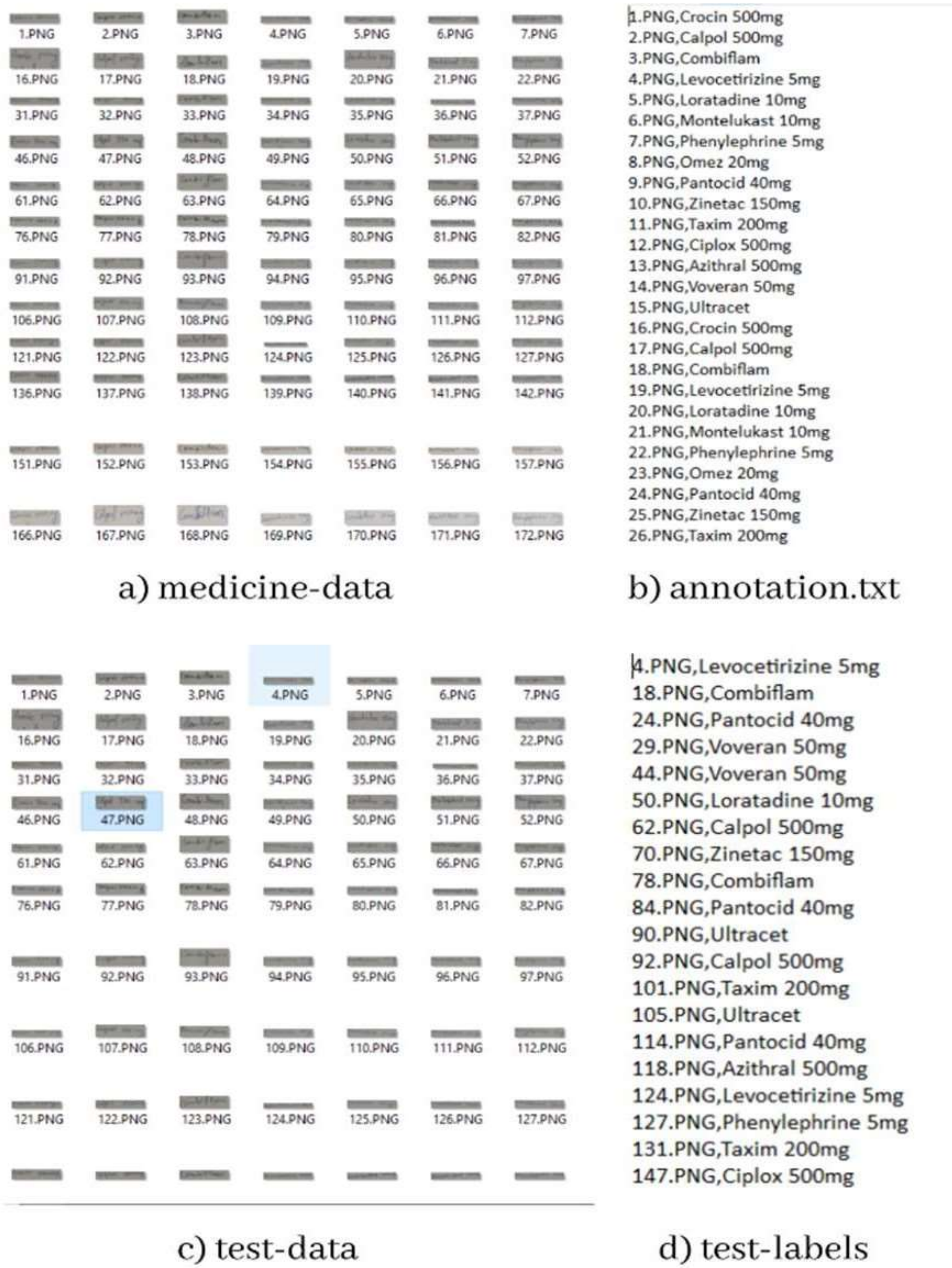

Figure 2. Training and Testing data

3.3. Data Pre-processing

Data pre-processing is crucial in training convolutional neural network (CNN) models for handwritten text recognition. The pre-processing techniques (14) during the CNN model discussed in this paper are as follows:

Image normalisation: The images are converted to grey-scaled images to improve image quality. Handwritten text images often have lighting conditions, contrast, and background noise variations. Normalisation techniques such as histogram equalisation or adaptive histogram equalisation can be applied to standardise the image intensities and improve the overall image quality.

Image resizing: For the CNN model in this paper, the images are converted into 64 X 64 squared shape images. Resizing the input images to a fixed size is necessary to ensure consistency during training. Typically, the images are resized to square while maintaining the aspect ratio to fit the CNN architecture.

Image Binarization: Converting the grayscale or colour images to binary images (black and white) is often done to enhance the contrast between the foreground text and the background. Various binarisation techniques like thresholding or adaptive thresholding can be employed to achieve this.

3.4. Processing Phase

The Convolution Neural Network is used to recognise the medicine names. The CNN model consists of various layers. The Keras Sequential API implemented a convolutional neural network (CNN) model architecture. Here's a breakdown of the model:

Convolutional Layers: The first Conv2D layer has 32 filters of size 3x3 and uses the activation function, which can be ReLU, exponential linear unit (ELU), tanh or Leaky ReLU. It takes an input shape defined by the variable 'input_shape'. The second Conv2D layer has 64 filters of size 3x3 and uses the ReLU activation function.

Pooling Layer: MaxPooling2D is applied with a pool size of 2x2. This operation downsamples the feature maps obtained from the convolutional layers, reducing their spatial dimensions.

Dropout: Dropout is applied after the pooling layer with a rate of 0.5. Dropout randomly sets a fraction of input units to 0 at each training step, which helps prevent overfitting.

Flattening: The Flatten layer converts the 2D feature maps into a 1D vector. It flattens the previous layer's output, preparing it for the subsequent fully connected layers.

Fully Connected Layers: The first Dense layer consists of 128 neurons with the activation function. It takes the flattened vector as input. Dropout is applied again with a rate of 0.5 to prevent overfitting further.

The final Dense layer has 'num_classes' neurons, the number of classes in the classification task. It uses the softmax activation function to output the predicted probabilities for each class.

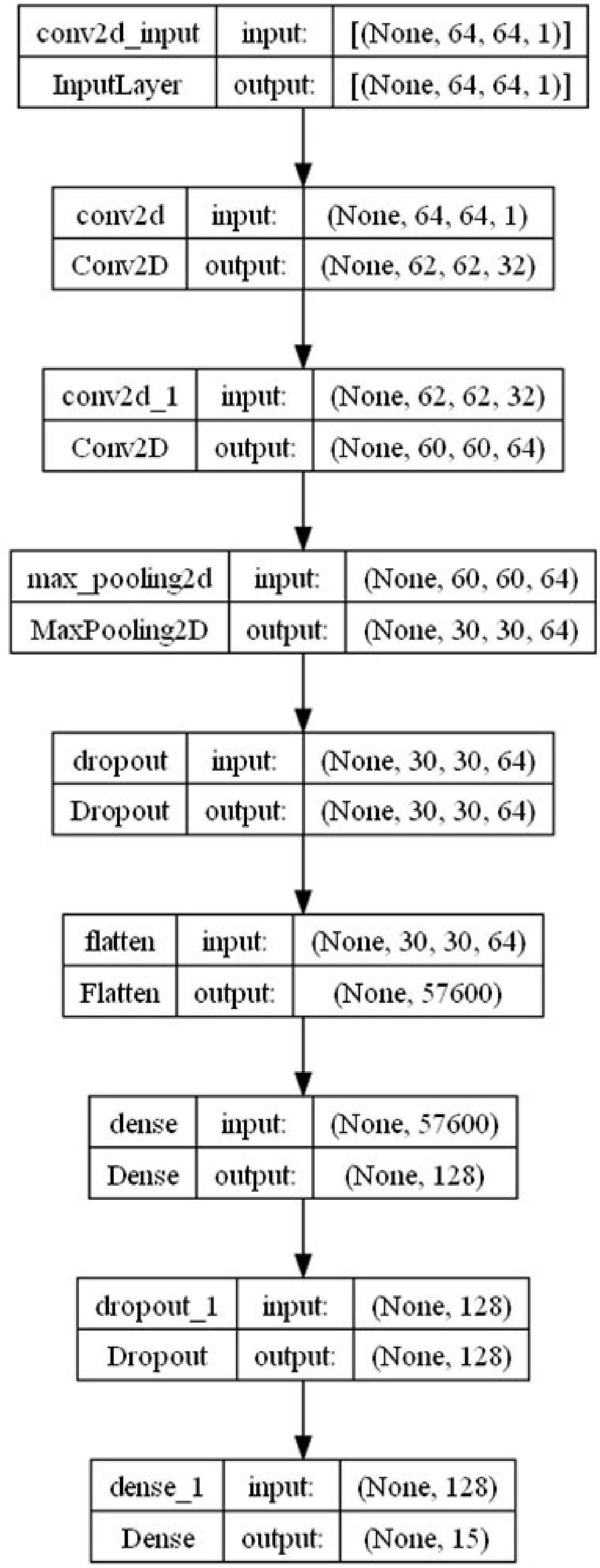

Figure 3. Defined CNN Model

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 32)	320
conv2d_1 (Conv2D)	(None, 60, 60, 64)	18496
max_pooling2d (MaxPooling2D)	(None, 30, 30, 64)	0
dropout (Dropout)	(None, 30, 30, 64)	0
flatten (Flatten)	(None, 57600)	0
dense (Dense)	(None, 128)	7372928
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 50)	6450

Total params: 7,398,194
Trainable params: 7,398,194
Non-trainable params: 0

Figure 4. CNN Model summary

Overall, this CNN model consists of multiple convolutional layers followed by pooling and dropout layers to extract hierarchical features from the input images. The flattened features are passed through fully connected layers, ultimately leading to the final '*softmax*' layer for multi-class classification. The CNN structure can be visualised as given in the figure, and Figure 3 describes the CNN model summary.

After defining the model as mentioned above, the model is trained on a training dataset with a batch size of 32, and at 15 epochs, the validation split used for the model was 0.2. After training the model, evaluation of the model is done on combinations of activation functions and optimisers using accuracy and the loss information obtained by testing the model against training data and testing data.

3.5. Post-processing

Model evaluation is performed in this phase to know how well the current model is working or how correctly the model can perform the desired task. The above-defined model is first evaluated on the trained data, which used the dictionary containing labels and corresponding medicine names and the predicted output to compare the result and based on that comparison output, the binary cross entropy loss is used to measure the loss parameter and accuracy is used to calculate the percentage of correctly predicted samples out of the total number of samples. Similarly, after evaluation of the training data, optimisation techniques or hyperparameter tuning is performed wherever required to improve the performance of the CNN model. After getting satisfactory training accuracy, the model can be evaluated against the testing data on the same parameters mentioned above. The training, testing, and evalua tion cycle followed by model optimisation was performed multiple times to get significant accuracy and lower the loss.

4. Result and Discussion

4.1. Dataset Description

An experiment is conducted on the large number of real samples collected by the team. The training and testing dataset is prepared for conducting model training and testing. The implementation of the model is done using Python programming language. Before training the data on the collected samples, manual and programmed preprocessing is done of the trained and test data. This preprocessing includes normalisation, resizing, and image binarisation. The output of the pre-processing stage is then used to train the CNN model, which works in two stages, primarily feature extraction and classification for medical prescription.

The dataset is collected from multiple doctors and hospitals across the city. The dataset is collected in a fashion that it contains several different prescriptions with the same medicines with different handwritings, which helps the model achieve generalisation and avoid overfitting.

4.2. Results

In the scanner prescriptions shown in Figure 1, we extracted the medicine names for model training and evaluation. After that, data pre-processing is done, which includes normalisation, image resizing, and image binarisation. The dataset includes 15 classes, i.e., names of 15 different medicines. The CNN model is trained over the epoch of 15 and batch size of 32 using binary cross-entropy loss for backpropagation learning. The maximum training accuracy was 89%, and the maximum test accuracy was 70% for the best-performing model parameter sets. The result obtained by the CNN model or different combinations of Activation function, optimiser and batch size is as follows.

Activation Function	Optimizer	Batch Size	Train Loss	Train Accuracy	Test Loss	Test Accuracy
ReLU	Adam	32	0.40	0.89	0.87	0.70
ReLU	RMSprop	32	0.38	0.90	1.35	0.56
ReLU	Adam	16	0.39	0.89	1.99	0.51
ELU	Adam	32	0.29	0.92	1.95	0.63
ELU	RMSprop	32	0.45	0.89	2.08	0.67
ELU	Adam	16	0.40	0.89	1.81	0.54
tanh	Adam	32	0.23	0.93	1.36	0.66
tanh	Adam	16	2.70	0.07	2.79	0.02
LeakyReLU	Adam	32	0.34	0.90	1.58	0.61
LeakyReLU	Adam	16	0.43	0.89	1.88	0.58
LeakyReLU	RMSprop	32	0.41	0.89	1.69	0.62

Table 1. Model Accuracy for different combinations of activation functions and optimisers

5. Conclusion and Future Work

In this paper, we trained and tested the Convolution Neural Network model for recognising medicine names from paper-based prescriptions. The model is based on Python and uses the *Tensorflow-Keras* library to work with CNN efficiently and effectively. The work in this paper proves its usefulness in the domain of Electronic Healthcare as this could help in the digitalisation of paper-based prescriptions, which can help maintain electronic health records. Digitalised prescription data can also be anonymised and aggregated to analyse trends, drug utilisation patterns, and public health outcomes. Such data can contribute to valuable research, help identify potential medication-related issues, and support evidence-based decision-making in healthcare. Even though the method used is not very innovative, the application towards recognising medical prescriptions is the greatest contribution to this work.

With digital prescriptions, tracking and monitoring patients' medication history becomes easier. This helps identify patterns, monitor adherence, and ensure patients receive the appropriate medications. It also facilitates medication reconciliation during transitions of care, reducing the risk of medication discrepancies.

After digitalising medical prescriptions, several areas of future work and potential advancements can further improve healthcare systems and patient care. This area includes Interoperability and Data Exchange, Integration with Clinical Decision Support Systems, Mobile Applications and Patient Engagement, Artificial Intelligence and Machine Learning, Blockchain Technology, Prescription Analytics and Population Health Management, Telemedicine and telepharmacy, and Continuous Quality Improvement.

Refrences

[1] Cowie, M. R., Blomster, J. I., Curtis, L. H., Duclaux, S., Ford, I., Fritz, F., others. (2017). Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*, 106(1), 1–9.

[2] Häyrinen, K., Saranto, K., Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International Journal of Medical Informatic*s, 77, 291–304.

[3] Waegemann, C. P., Tessier, C., Barbash, A., Blumenfeld, B. H., Borden, J., Brinson Jr, R. M. (2002). Healthcare documentation: A report on information capture and report generation. Newton, MA: Medical Records Institute.

[4] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature,* 521, 436–444.

[5] Mithe, R., Indalkar, S., Divekar, N. (2013). Optical character recognition. *International Journal of Recent Technology and Engineering* (IJRTE), 2, 72–75.

[6] Dhar, D., Garain, A., Singh, P. K., Sarkar, R. (2021). HP_DocPres: A method for classifying printed and handwritten texts in doctor's prescription. Multimedia Tools and Applications, 80, 9779–9812.

[7] Jain, T., Sharma, R., Malhotra, R. (2021). Handwriting recognition for medical prescriptions using a CNN-BiLSTM model. *In:* 2021 6th International Conference for Convergence in Technology (I2CT).

[8] Hassan, E., Tarek, H., Hazem, M., Bahnacy, S., Shaheen, L., Elashmwai, W. H. (2021). Medical prescription recognition using machine learning. *In:* 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC).

[9] Achkar, R., Ghayad, K., Haidar, R., Saleh, S., & Al Hajj, R. (2019). Medical handwritten prescription recognition using CRNN. *In:* 2019 International Conference on Computer, Information and Telecommunication Systems (CITS).

[10] Tabassum, S., Abedin, N., Rahman, M. M., Ahmed, M. T., Islam, R., Ahmed, A. (2022). An online cursive handwritten medical words recognition system for busy doctors in developing countries for ensuring efficient healthcare service delivery. *Scientific Reports*, 12, 3601.

[11] Balci, B., Saadati, D., Shiferaw, D. (2017). Hand

written Text Recognition using Deep Learning. Report of Stanford University. Link to the report.

[12] Bhattacharyya, S. (2011). A brief survey of color image preprocessing and segmentation techniques. *Journal of Pattern Recognition Research*, 1, 120–129.

[13] Pal, K. K., Sudeep, K. S. (2016). Preprocessing for image classification by convolutional neural networks. *In:* 2016 *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology* (RTEICT).

[14] Kohli, R., Tan, S. S.-L. (2016). Electronic Health Records. *MIS Quarterly*, 40, 553–574.

[15] Jha, A. K., DesRoches, C. M., Campbell, E. G., Donelan, K., Rao, S. R., Ferris, T. G., Blumenthal, D. (2009). Use of electronic health records in US hospitals. *New England Journal of Medicine,* 360, 1628–1638.