'National Conference on Advancements in Computer Science and Engineering'
International Journal of Scientific Research in Computer Science,
Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Prescription Recognition for the Visually Impaired using Convolutional Neural Networks (CNN)

Jessica Vareed, Keerthana Salome Joshua
Department of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India

ABSTRACT

In order to help visually impaired people, manage their medications, this paper suggests a novel approach for prescription recognition using the Convolutional Neural Network (CNN) algorithm. The proposed system is made to recognize prescription labels and translate them into speech, making it simple for people who are visually impaired to manage and identify their medication. OCR algorithms are typically designed to recognize specific fonts or handwriting styles and may struggle with variations in handwriting, different languages, or unusual characters. The CNN algorithm is used in this suggested architecture to recognize and categorize images. The algorithm can identify different prescription label formats and styles after being trained on a dataset of prescription labels. The design of CNN models has been extensively automated using reinforcement learning (RL), leading to notable improvements and intriguing results in this field.

Keywords: Convolutional Neural Network (CNN), Prescription recognition, visually impaired, Image recognition

I. INTRODUCTION

According to a survey conducted by World Health Organization (WHO) in 2019, 1181.4 million is the total population count of India and out of which 152.23 million are people who suffer from blindness, low vision, and visual impairment. Visually impaired people are struggling in their daily life activities such as recognizing objects such as medicines, and need a third person's help to read.

The society we live in today is increasingly digital, and technology is developing quickly. Artificial intelligence, computer vision, and other technologies are being used to create applications that can extract text and then transform it into audio utilizing text-to-speech or image- to-speech technology. This can be applied to pharmaceuticals to extract appropriate text and then provide audio. Using this advanced technology and device researchers are making economical and portable systems to assist visually impaired people.

Prescription recognition has attracted a lot of attention in recent years thanks to the use of machine learning methods like OCR (Optical Character Recognition) and CNN (Convolutional Neural Network). OCR algorithms have been widely used for text recognition tasks, including prescription recognition. However, OCR algorithms may not be suitable for recognizing complex patterns or objects in images, which can lead to errors in prescription recognition.

In contrast, CNN algorithms have demonstrated promising performance in image recognition tasks, such as prescription recognition. CNN algorithms can learn to identify objects from incomplete or partial images and can be trained on various image recognition tasks. In addition, CNN algorithms may be more resistant to changes in image quality, lighting, and other elements that may have an impact on prescription recognition.
In image recognition and computer vision tasks, convolutional neural networks (CNNs) are a common type of artificial neural network. CNNs are built to identify patterns and features in images and are inspired by the structure and operation of the visual cortex in the human brain. Convolutional, pooling, and fully connected layers are among the layers that together make up a typical CNN. The network applies a set of filters to the input image in the convolutional layers, searching the image for patterns and features like edges, corners, and shapes. The final step is to pass the generated feature maps through pooling layers, which down-sample the feature maps by using the maximum or average values from each pooling region. By reducing the feature maps' dimensionality, overfitting is prevented and processing becomes simpler.

II. METHODS AND MATERIAL

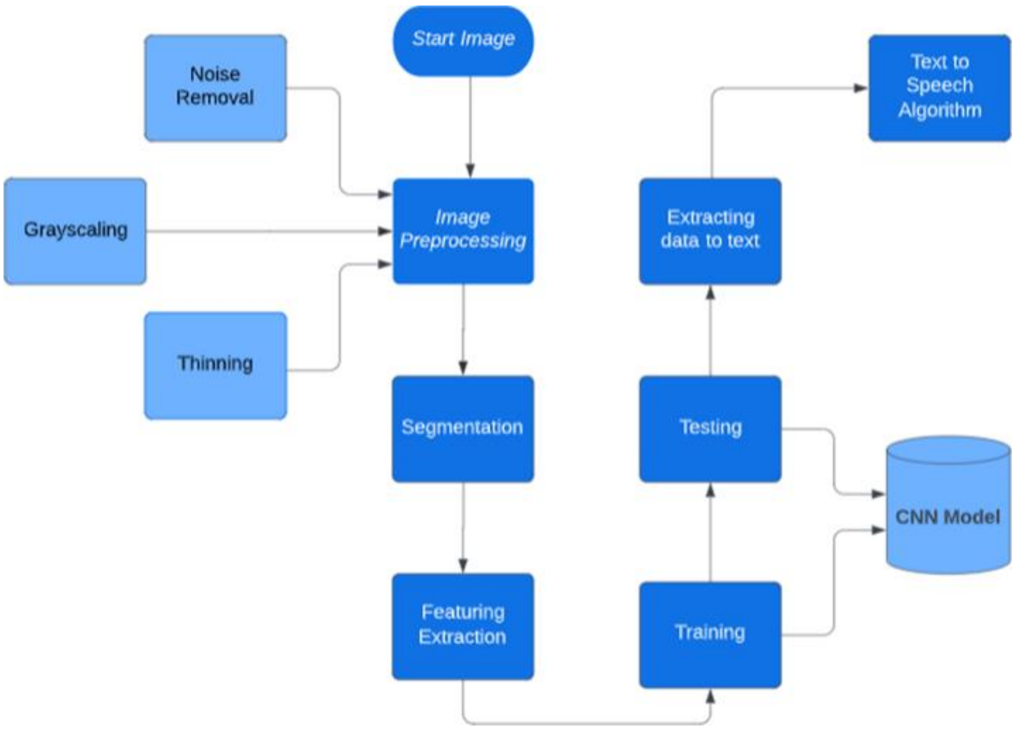

Fig 1: Methodology

To automatically understand the scanned prescriptions uploaded, a portable system was created. All of the prescriptions are scanned using a portable scanner and stored on the Raspberry Pi, where the scanned images are converted from pixmap to jpg format. A laptop or mobile phone receives the jpg-formatted image from the Raspberry Pi. The CNN models are on the laptop that is used to upload the scanned prescriptions. First, the scanned images are segmented so that the machine learning algorithm only receives the handwritten alphabets and numbers. Thus, following segmentation, training data is provided to CNN models so that they can be properly trained to recognize handwritten answers, and testing data is provided so that the trained CNN model can be loaded and the recognized text is then converted to an audio form of output.

A. Data collection

The first step is to collect a dataset of prescription images with corresponding labels. The dataset should include a range of font sizes and styles in addition to different handwritten prescription formats. Scanners are used to acquire these images of prescriptions. The Raspberry Pi can record and store the original image.

B. Pre-processing

Pre-processing is a critical step in the character recognition of prescription images. The primary goal of pre-processing is to enhance the quality of the prescription images and reduce noise. In order to achieve this, various operations may be used, such as resizing the images to a standard size, normalizing the pixel values to lessen variations in lighting and contrast, and image binarization to turn the prescription image into a binary black-and-white image. Other pre- processing methods might include segmenting the individual characters in the prescription image and removing any non-text regions in the prescription image, like logos or graphics.

Once the writing is segmented into smaller units, these units are sent to a module that extracts features in the data, essential to the employed shape classification algorithm. It involves transforming raw data into a representation that captures relevant information and discriminative characteristics for the given task. It involves identifying distinctive visual elements, such as edges, corners, or texture patterns, which can be used to differentiate between different objects or classes. These pre-processing methods are crucial for increasing the character recognition algorithm's accuracy because they make the characters' clarity and unnecessary information more comprehensible.

C. Training and Testing

The CNN model is trained on a large collection of prescription images with corresponding labels during the training phase. To minimize the discrepancy between the predicted and actual labels, the model is optimized using backpropagation and gradient descent algorithms. Train the model using the training set, where the prescription images are fed as input, and the corresponding ground truth labels (e.g., drug names, dosages) are provided for supervision. The model learns to recognize patterns and features from the prescription images to make accurate predictions.

The CNN model is tested using a different dataset of prescription images that weren't used during training. Measuring metrics such as accuracy, precision, recall, and F1 score are used to assess the model's performance. These metrics aid in assessing how well the CNN model recognizes characters in prescription images.

D. Convolutional Neural Network (CNN)

The CNN is a particular kind of neural network that can draw out important details from both time series and image data. It is very useful for pattern recognition, object classification, and image recognition. A CNN employs concepts from linear algebra, such as matrix multiplication, to find patterns within an image. They are made up of several layers, including pooling, convolutional, and fully connected layers.

The foundational component of the CNN is the convolution layer. By applying filters to the input image and identifying spatial relationships between pixels, it carries the majority of the computational load. This layer creates a dot product between two matrices, one of which is the kernel—a set of learnable parameters—and the other of which is the constrained area of the receptive field. Although the kernel is more detailed than an image, it is spatially smaller. As a result, a two-dimensional representation of the image called an activation map is created, revealing the kernel's response at each location in the image.

The input parameter count is decreased and some information is lost as a result of the pooling layer. On the plus side, this layer streamlines operations and boosts CNN effectiveness. By calculating an aggregate statistic from the nearby outputs, the pooling layer substitutes for the network's output at specific locations.

This aids in shrinking the representation's spatial size, which lowers the amount of computation and weights needed. Each slice of the representation is subjected to the pooling operation separately. Although there are various pooling functions, such as the L2 norm of the rectangular neighbourhood, the most well-known one is max pooling, which reports the maximum output of neighbourhood. Based on the features extracted in the preceding layers, image classification in the CNN takes place in the FC layer. The representation between the input and the output is mapped using the FC layer.

Non-linearity layers are frequently added right after the convolutional layer to add non-linearity to the activation map because convolution is a linear operation and images are anything but linear. ReLU speeds up convergence by six times and is more dependable than sigmoid and tanh. They are in charge of ensuring that the model isn't both over- and under-fitted for a particular dataset. One can get the results one wants for a particular problem by altering the parameters. The various CNN parameters that were used to build the CNN models are covered in this section.

III. RESULTS AND DISCUSSION

The results and output of a CNN algorithm for prescription recognition have improved the ability of visually impaired individuals to comprehend prescription images and convert them into machine- readable text. The CNN algorithm may recognize the characters in the prescription image and translate them into text that can be read aloud on a screen or by text-to- speech software.

The dataset contained a wide range of prescription styles, typefaces, and layouts that were carefully selected to reflect heterogeneity in the real world. We split the dataset into a training set (80%) and a testing set (20%) in order to ensure a thorough analysis. The high accuracy demonstrates how well the suggested algorithm understands information from prescriptions, including names, doses, and instructions for use. The model's robustness in handling variations in font styles and layouts further illustrated its ability to be generalized. Our CNN algorithm's capacity to recognize intricate features and patterns in prescription images is what gives it such high accuracy.

The model can extract hierarchical representations using the CNN architecture's numerous convolutional layers, gathering both low-level characteristics (like edges and corners) and high-level semantic information (like text regions and structure).

The accuracy of the CNN algorithm can be evaluated by comparing the output with the actual prescription data. The accuracy of the algorithm can vary depending on the size of the training dataset for the CNN model, the complexity of the prescription format, and the calibre of the prescription image. Using text-to-speech software, the output can be read aloud or displayed on a screen, providing visually impaired people with a user-friendly interface to access essential prescription information.

For tasks involving character recognition, including prescription recognition for people with visual impairments, CNN algorithms have generally shown to be very effective.

IV. CONCLUSION

In conclusion, prescription recognition algorithms developed by CNN have the potential to greatly improve the accessibility of prescription data for blind people. Using deep learning techniques, CNN algorithms can recognise and convert prescription images into machine- readable text with accuracy. Despite the fact that OCR algorithms have long been a popular option for character recognition tasks, CNN algorithms have shown to be more effective at identifying characters in difficult images, like prescription images. The accuracy of the CNN model can be improved by increasing the training dataset, adjusting the hyperparameters, and applying sophisticated pre-processing techniques.

To improve CNN algorithms for prescription recognition and to confirm their efficacy in real-world settings, more research is needed. Further research must be done on the CNN algorithm's integration with assistive technologies like text-to-speech software in order to improve the system's usability for people with visual impairments.

Future work can be done by increasing the speed and accuracy of recognition in a much better way. It can still be enhanced to obtain accurate results in noisy environments. Future work will involve creating a mobile application that will scan a doctor's prescription and remind the user to take their medication.

V. REFERENCES

[1]. Pritam S. Dhande,Reena Kharat (2017) "Character Recognition for Cursive English Handwriting to Recognize Medicine Name from Doctor's Prescription" International Conference on Computing, Communication, Control and Automation (ICCUBEA).

[2]. Eman Shaikh, Iman Mohiuddin, Ayisha Manzoor, Ghazanfar Latif, Nazeeruddin Mohammad (2019) "Automated Grading for Handwritten Answer Sheets using Convolutional Neural Networks". 2nd International Conference on New Trends in Computing Sciences (ICTCS).

[3]. Lovely Joy Fajardo, Niño Joshua Sorillo, Jaycel Garlit, Cia Dennise Tomines , Mideth B. Abisado, Joseph Marvin R. Imperial , Ramon L. Rodriguez,Bernie S. Fabito (2019) "Doctor's Cursive Handwriting Recognition System Using Deep Learning". IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)

[4]. Snigdha Kesh, Ananthanagu U (2017) "Text Recognition and Medicine Identification by Visually Impaired People" INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) ICPCN – 2017 (Volume 5 – Issue 19)