# iReader: An Intelligent Reader System for the Visually Impaired

Jothi G*, Ahmad Taher Azar†‡, Basit Qureshi†, Nashwa Ahmad Kamal§
*Department of Computer Applications, Sona College of Arts and Science, Salem. Tamil Nadu. India.
Email: jothi.g@sonacas.edu.in, jothiys@gmail.com

†College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia.
Email: aazar@psu.edu.sa, qureshi@psu.edu.sa

‡Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt.
Email: ahmad.azar@fci.bu.edu.eg, ahmad_t_azar@ieee.org
§Faculty of Engineering, Cairo University,Giza, Egypt.
Email: nashwa.ahmad.kamal@gmail.com

2022 7th International Conference on Data Science and Machine Learning Applications (CDMA) | 978-1-6654-1014-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDMA54072.2022.00036

***Abstract*—For visually impaired persons, it is quite difficult to read printed text. Non-visual forms of reading materials, such as Braille, are available as Blind Aiding Technology amoung many others. In recent times, many devices and assistive equipment have been developed and technologies made available to assist visually impaired persons with reading. Most of these research works and products support reading from printed text-based manuscripts only. Due to this limitation, it may not be possible for a visually impaired person to describe and comprehend a printed image. In this paper, we develop iReader, an Intelligent Reader system that not only helps a visually impaired reader to read but also vocally describes an image available in the printed text. The Convolution Neural Network (CNN) is employed to collect features from the printed image and its caption. The Long Short-Term Memory (LSTM) network is used to train the model for describing the image data. The resulting data is sent as a voice message using Text-To-Speech to be read out loud to the user. The efficiency of the LSTM model is examined using the ResNet50 and VGG16. The experimental results show that the LSTM-based training model delivers the best prediction of a picture's description with an accuracy of 83**

**Key words:** Deep Learning; Convolution Neural Network (CNN); Long Short-Term Memory (LSTM); Visually Impaired People; Reader System; Blind Aiding Technology.

## I. Introduction

There are an estimated 253 million people worldwide who are visually impaired. 36 millions of them are blind, and 217 million have low to high vision deficiencies. In this population, 4.8 percent of people are born with a visual deficiency (blindness), while 90 percent of visual deficiency is caused by factors such as age-related macular degeneration, diabetes, glaucoma, accidents, and so on. As the world's aging population grows, the rate of vision loss associated with chronic disease is expected to rise [1]. To gain access to education, visually impaired people may require the assistance of a person or assistive equipment. Students with visual impairments have access to materials in a variety of formats, including Braille, audiotape, and magnified print [2]. These devices are only useful for reading text; it is extremely difficult to read image data. Technology is used in educational initiatives to help blind and special-needs students succeed in school. Recently, researchers have developed technologies and systems using machine learning to assist visually impaired people. Priya et.al. in [3] develop a machine-learning algorithm used in accurate detection of objects during navigation by a blind person and provides an alarm. The wearable device with an ultrasonic sensor measures the distance between the person and an object detected in real-time to actuate a buzzer to alarm the user. Authors in [4] and [5] highlight the importance of machine learning in supporting visually challenged. They develop a convolution neural network, trained on ImageNet dataset that detects objects and narrates the detected objects to the visually impairs person. Durgadevi in [6] use image classification as an object detection approach. In their work they use narration to describe the detected object to the user. Felix et.al in [7] develop an android based mobile application that incorporates a voice assistant, image recognition, currency recognition, e-book reading into one platform. The purpose of this application is to support the visually impared in navigating through daily chores. Denic et.al. in [8] use a convolution neural network to detect objects and communicate them using text-to-speech. Deep learning has grown in popularity as a field of study that attempts to discover new techniques for automating various operations based on input data [9–15]. Deep learning is a subset of artificial intelligence techniques with a wide range of applications including image recognition, virtual assistants, healthcare, natural language processing, fraud detection, and so on. This article describes the creation of a Deep Neural Network-based Intelligent Reader system that allows visually impaired people to read and describe images in a printed text book. In this proposed method, Convolutional Neural Network (CNN) [16] is used to extract features from a dataset of images and Long Short-Term Memory (LSTM) [17] is used to describe the visual information. The LSTM architecture is a deep learning artificial recurrent neural network (RNN) architecture. The intelligent learning system provides a voice message of text and image information from a printed text book using text-to-speech technology. Deep learning-based methods address image-to-speech task performance accuracy and can be used to improve the quality of life for people who

978-1-6654-1014-4/22/$31.00 ©2022 IEEE
DOI 10.1109/CDMA54072.2022.00036

are visually impaired. The framework describing iReader is depicted in Figure 1.

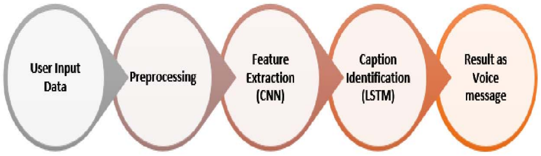


Figure 1 – The Intelligent Reading System Architecture

The rest of the paper is structured as follows: Section II describes the many learning systems available to visually impaired persons. The framework architecture using Deep learning approach is described in Section III. Section IV details the evaluation of the system using the experimental outcomes carried out in the study. Finally, section V presents the conclusion and future work.

## II. Related Work

Advanced technologies such as computer vision, machine learning, and deep learning have recently been used to develop an autonomous learning system for visually impaired people. The author in [18] describe a one-of-a-kind low-cost solar-powered wearable assistive technology (AT) device that provides continuous, real-time object identification to help visually impaired (VI) persons identify objects in their everyday lives. The system's three main components are a small low-cost camera, a system on module (SoM) processing unit, and an ultrasonic sensor. The first is worn on the user's glasses and records real-time video of the surroundings. The second is worn as a belt and detects and recognizes things by processing footage from the camera. Third, it assists in the detection of objects in the surrounding region. Lin et.al.[19] in propose a deep learning-based assistance system to improve visually impaired people's perception of their environment. The system consists of a wearable terminal equipped with an RGBD camera and an earphone. A powerful CPU for deep learning conclusions, as well as a smartphone for touch-based interaction. It expects safe and trustworthy walking instructions using RGBD data and the pre-existing semantic map. Zamir et al. [20] suggested a smart reader system based on the Raspberry Pi that translates text into speech signals. A camera is used to recognize printed text using an optical character recognition (OCR) technique. The proposed method is for an image to audio converter built on the Raspberry Pi single board computer. In [21], a deep convolutional neural network-based architecture is used to build an indoor object detector. The system is based on the "RetinaNet" deep convolutional neural network. In order to increase detection performance, various backbones such as ResNet, DenseNet, and VGGNet are used in the evaluation. Intriguing results with a detection precision of 84.61 To aid visually impaired people, Tasnim et al. [22] propose an automated solution for detecting Bangladeshi banknotes using a convolutional neural network. The system has an average accuracy of 92 percent in identifying the eight banknotes used in Bangladesh and can produce both written and aural output. A fresh collection of over 70,000 photographs of contemporary Bangladeshi banknotes has been created. People who are blind or visually impaired will find it easy to use in routine interactions. In table 1, we summarizes the learning approach used in recent works designed for those who are visually impaired.

Table I – Summarization of the Related works

| Author(s) | Year | Technique used | Developed System |
|---|---|---|---|
| Denic et.al. [8] | 2019 | Convolutional Neural Network | Object Detection System |
| Felix et.al. [7] | 2019 | Android Mobile Application | Blind Assistive technology |
| Durgadevi et. al. [6] | 2020 | Image classification | Indoor object detection |
| Lin et. al. [19] | 2019 | Deep Learning | Assistive system to find the people's perception of their surroundings |
| Zamir et. al. [20] | 2019 | Raspberry Pi | OCR Text Detection system |
| Calabrese et. al. [18] | 2020 | Deep Learning | object detection system |
| Afif et. al. [21] | 2020 | Deep CNN | indoor object detector system |
| Cheng et. al. [23] | 2021 | NetVLAD | image description System |
| Tasnim et. al. [22] | 2021 | CNN | Bangladeshi banknotes detection system |

## III. Techniques for Deep Learning

Deep learning is a machine learning and artificial intelligence area that mimics how people learn certain types of information. Deep learning is extremely valuable for academics who must collect, analyze, and interpret huge amounts of data since it speeds up and simplifies the process. Deep learning makes use of a variety of neural networks, such as recurrent neural networks, convolutional neural networks, artificial neural networks, and feedforward neural networks, each of which has benefits in certain applications [24].

### A. *Convolutional Neural Network (CNN)*

Convolutional Neural Networks [25] are a type of deep learning neural network. It is a significant advancement in picture classification and recognition. CNN is made up of layers, and in ConvNet, each layer is combined with activation functions. The input layer, the Convolutional Layer, the Pooling Layer, and the Fully-Connected Layer are the four primary types of layers in CNN architecture. The convolution process in the convolutional layer retrieves low-level to high-level features from the input layer using a sequence of convolutional layers. The Pooling layer is in charge of shrinking the spatial size of the convolved features. Pooling is classified into two types: maximum pooling and average pooling. Average Pooling returns the average of all the values from the portion of the picture covered by the kernel, whereas Max Pooling returns the maximum value. The Softmax Classification approach is used by the Fully Connected layer to compute the class label scores and classify them. When we train the network on a huge dataset, we train all of the neural network's parameters, and so the model is learned.

### B. *Long Short Term Memory (LSTM)*

Long Short-Term Memory (LSTM)[26] networks are Recurrent Neural Networks (RNN) that may learn order dependency in sequence prediction tasks. A Recurrent Neural Network (RNN) is a feedforward neural network with internal memory. RNN is a form of Neural Network in which the previous step's output is used as input for the current phase.

After the output is generated, it is replicated and returned to the recurrent network. It evaluates the current input and the output that it has learned from the prior input while making a decision. The Recurrent Neural Network (RNN) assists in identifying the sequence on the images. Time series prediction, voice recognition, rhythm learning, grammar learning, handwriting identification, human action detection, activity recognition, picture description, video description, and so on are all uses of LSTMs. The LSTM network architecture is depicted Figure 2. A conventional LSTM network is made up of several memory blocks known as cells, which are represented by the rectangles in the picture. Memory blocks are in charge of remembering things, and modifications to this memory are performed via one of four gate methods. LSTMs deal with both Long Term Memory (LTM) and Short Term Memory (STM), and the notion of gates is used to make the computations simple and effective. The LTM in this design travels to the forget gate and forgets information that is no longer helpful. Learn gate is used to learn vital information that we have recently acquired from STM and remember gate is utilised to update the LTM information. Finally, the use gate predicts the current event's output.

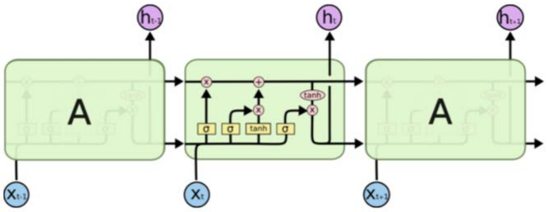


Figure 2 – LSTM Network Architecture [27]

### C. The proposed CNN-LSTM Design

In the proposed approach, the input file is fed into the intelligent reader system. The system reads the contents of the file and extracts the characters from it using an optical character recognition (OCR) tool. Text-to-Speech (Google TTS) technology is used to convert a user's written input into a voice response. If the file contains images, the trained CNN-LSTM model predicts the caption for each image and delivers the prediction to the intelligent reader system. All information will be delivered through voice message by the reader system. CNN-LSTM is a deep learning architecture that incorporates two deep learning algorithms: Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) (LSTM). The CNN model is utilized to extract features from the input data for sequence prediction, and the LSTM model is used to predict the image caption. This approach is intended for sequence prediction problems including spatial inputs such as photos or videos. The entire architecture of the proposed CNN-LSTM model is depicted in Figure 3.

## IV. RESULTS AND DISCUSSION

### A. Dataset Collection

The model in this article is trained using the Flickr8k dataset [28,29]. The dataset is divided into two repositories: Flickr8k Dataset (8092 images) and Flickr8k Text (image names and captions). The dataset is divided into three parts: training, validation, and testing, with 80

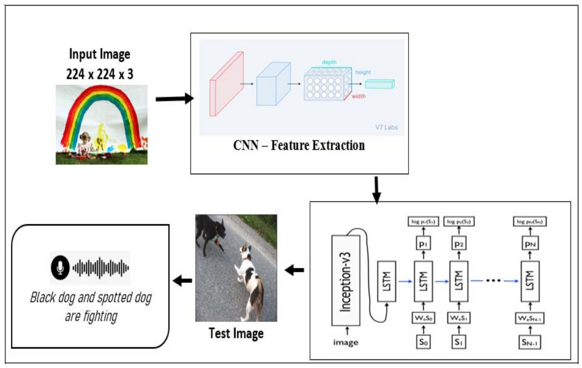


Figure 3 – CNN-LSTM Design

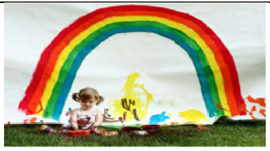
| Sample Image | Description/Caption |
|---|---|
|  | A small girl in the grass plays with finger paints in front of a white canvas with a rainbow on it. |
| | A little boy has fallen asleep in his food , which is sitting on a blue and yellow tablecloth |
| | A man attempts to start a flying model of a bi-plane with British air force rondel markings |

Table II – Sample Image and Description

### B. Results and Discussion

The training dataset was fed into the model as an input for training. In this study, two deep learning algorithms, Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), were used to determine the caption of an image. The CNN method is used to extract the features, and an LSTM trained model is used to determine the caption for the image. After training, the model accepts an image as input and generates a short-written summary explaining the image's content. The trained model assists it in capturing all of the detail within the image, as shown in Figure 4.

The efficiency of the LSTM model is examined using the ResNet50 and VGG16 architectures. The parameter is depicted in Figure 5.

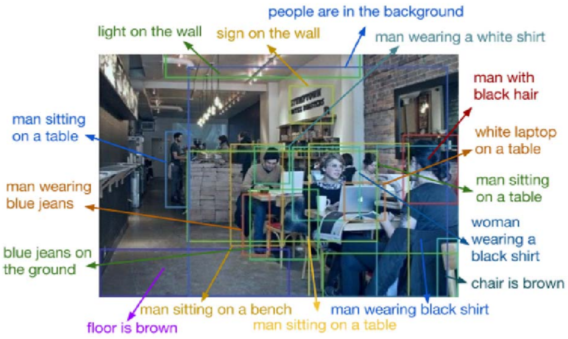
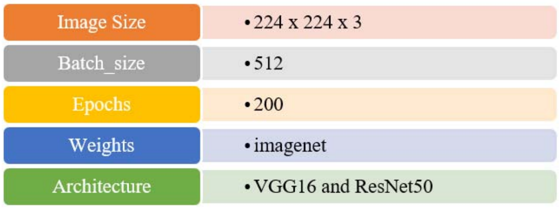


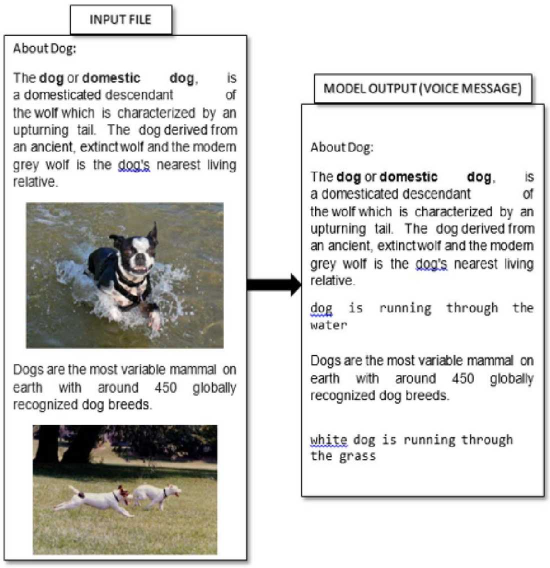
Figure 4 – Image Description using Trained Model [29]



| | |
|---|---|
| Image Size | • 224 x 224 x 3 |
| Batch_size | • 512 |
| Epochs | • 200 |
| Weights | • imagenet |
| Architecture | • VGG16 and ResNet50 |

Figure 5 – Parameter Values of Proposed Method

The proposed method receives an input file containing both text and images. The trained LSTM model is used to predict the image's caption, and the system provides all of the information through voice message. Figure 6-7 depict an example's output of the proposed intelligent reader system.



INPUT FILE

About Dog:

The **dog** or **domestic dog**, is a domesticated descendant of the wolf which is characterized by an upturning tail. The dog derived from an ancient, extinct wolf and the modern grey wolf is the dog's nearest living relative.

Dogs are the most variable mammal on earth with around 450 globally recognized dog breeds.

MODEL OUTPUT (VOICE MESSAGE)

About Dog:

The **dog** or **domestic dog**, is a domesticated descendant of the wolf which is characterized by an upturning tail. The dog derived from an ancient, extinct wolf and the modern grey wolf is the dog's nearest living relative.

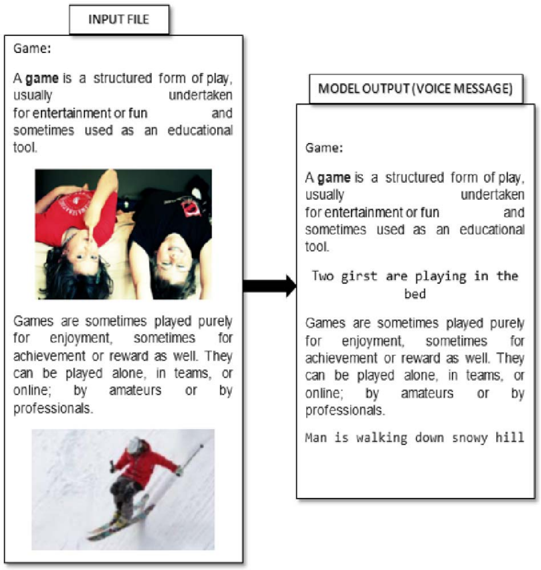
dog is running through the water

Dogs are the most variable mammal on earth with around 450 globally recognized dog breeds.

white dog is running through the grass

Figure 6 – The Proposed System Output 1

The image captioning system's performance is evaluated using BLEU [30]. It is used to investigate the n-gram correlation between the translation statement in question and the reference translation statement. The greater the BLEU score, the better the performance. The 1-gram and 2-gram BLEU score values for VGG16 and ResNet50 are shown in Table III. The VGG16 architecture is said to outperform the ResNet50 design. To evaluate the effectiveness of the LSTM algorithm, evaluation measures such as Precision, Recall, and F-score were used, which are defined as:



INPUT FILE

Game:

A **game** is a structured form of play, usually undertaken for entertainment or fun and sometimes used as an educational tool.

Games are sometimes played purely for enjoyment, sometimes for achievement or reward as well. They can be played alone, in teams, or online; by amateurs or by professionals.

MODEL OUTPUT (VOICE MESSAGE)

Game:

A **game** is a structured form of play, usually undertaken for entertainment or fun and sometimes used as an educational tool.

Two girst are playing in the bed

Games are sometimes played purely for enjoyment, sometimes for achievement or reward as well. They can be played alone, in teams, or online; by amateurs or by professionals.

Man is walking down snowy hill

Figure 7 – The Proposed System Output 2

Table III – BLEU Score Values

| Architecture | BLEU-1 | BLEU-2 |
|---|---|---|
| VGG 16 | 0.7824 | 0.7303 |
| ResNet50 | 0.7416 | 0.7086 |

$$Precision = \frac{TruePositive}{(TruePositives + FalsePositive)} \tag{1}$$

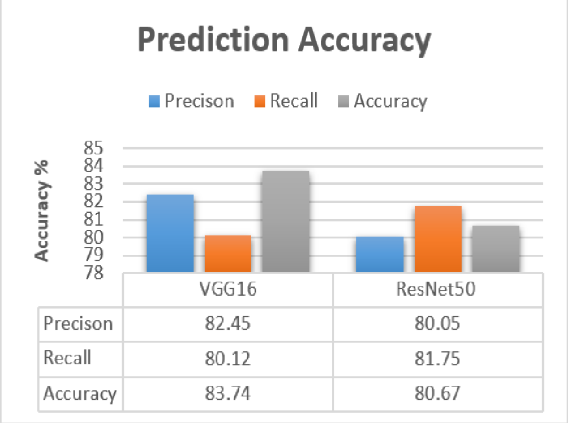
$$Recall = \frac{TruePositive}{(TruePositive + FalseNegative)} \tag{2}$$

$$Accuracy = \frac{No.ofCorrectedPredictions}{TotalNo.ofPredictions} \tag{3}$$

The predicted and tagged image captions are compared, and the algorithm's performance is evaluated. The VGG16 design outperforms the ResNet50 architecture in terms of image captioning. The produced data is then converted into a voice message using text-to-speech recognition. Figure 8 illustrates the evaluation metric values for two different architectures, VGG16 and ResNet50. VGG16 appears to have the highest accuracy value.

The loss chart is used in the deep learning model assessment process to determine how well the model fits the training data (training loss) and new data (validation loss). The training and validation loss of VGG16 and ResNet50 architectures are depicted in Figure 9 and Figure 10, respectively.

When compared to the ResNet50 design, the VGG16 architecture is believed to fit the training data and predict the new data more effectively.



| | VGG16 | ResNet50 |
|---|---|---|
| Precison | 82.45 | 80.05 |
| Recall | 80.12 | 81.75 |
| Accuracy | 83.74 | 80.67 |

Figure 8 – Prediction Accuracy

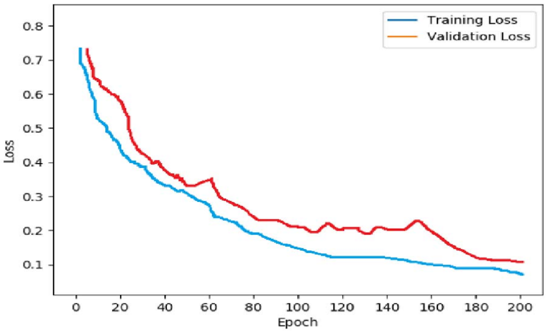


Figure 9 – Training and Validation loss (VGG16)

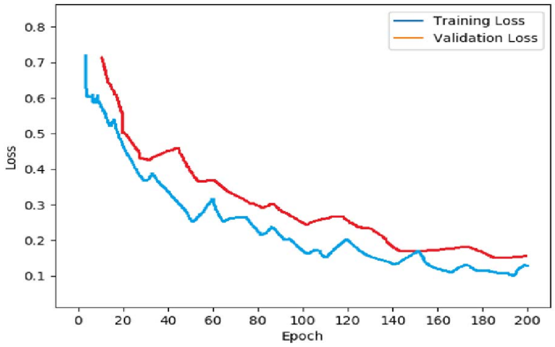


Figure 10 – Training and Validation loss (ResNet50)

## V. CONCULUTION

This study constructs a deep learning-based intelligent system for visually impaired persons. In this system, enter the text and image data from the textbook. The CNN method is used to extract features, whereas the LSTM method characterizes the visual input. All of this information is delivered to users as a as a voice message using the text-to-speech module. The Resnet50 architecture is utilized to train the model in the LSTM. The experimental results show that the LSTM-based training model delivers the best picture prediction and description. The intelligent reading technology allows visually impaired people to read text and graphic information with ease. The system's limitation is that it describes picture information using the Flicker8k dataset. Transfer learning will be utilized in the future to adjust image descriptions depending on real-time photographs and their descriptions.

## ACKNOWLEDGMENT

The authors would like to thank Prince Sultan University, Riyadh, Saudi Arabia for supporting this work. Special acknowledgement to Automated Systems & Soft Computing Lab (ASSCL), Prince Sultan University, Riyadh, Saudi Arabia.

## REFERENCES

[1] B. H. Lee and Y. J. Lee, "Evaluation of medication use and pharmacy services for visually impaired persons: Perspectives from both visually impaired and community pharmacists," *Disability and health journal*, vol. 12, no. 1, pp. 79–86, 2019.

[2] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: a deep learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2014.

[3] T. Priya, K. S. Sravya, and S. Umamaheswari, "Machine-learning-based device for visually impaired person," in *Artificial Intelligence and Evolutionary Computations in Engineering Systems*. Springer, 2020, pp. 79–88.

[4] P. Chandankhede and A. Kumar, "Deep learning technique for serving visually impaired person," in *2019 9th International Conference on Emerging Trends in Engineering and Technology-Signal and Information Processing (ICETET-SIP-19)*. IEEE, 2019, pp. 1–6.

[5] J. Nasreen, W. Arif, A. A. Shaikh, Y. Muhammad, and M. Abdullah, "Object detection and narrator for visually impaired people," in *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*. IEEE, 2019, pp. 1–4.

[6] S. Durgadevi, K. Thirupurasundari, C. Komathi, and S. M. Balaji, "Smart machine learning system for blind assistance," in *2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*. IEEE, 2020, pp. 1–4.

[7] S. M. Felix, S. Kumar, and A. Veeramuthu, "A smart personal ai assistant for visually impaired people," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2018, pp. 1245–1250.

[8] D. Denić, P. Aleksov, and I. Vučković, "Object recognition with machine learning for people with visual impairment," in *2021 15th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS)*. IEEE, 2021, pp. 389–392.

[9] Z. Li, F. Song, B. C. Clark, D. R. Grooms, and C. Liu, "A wearable device for indoor imminent danger detection and avoidance with region-based ground segmentation," *IEEE Access*, vol. 8, pp. 184 808–184 821, 2020.

[10] H. A. Elkholy, A. T. Azar, A. Magd, H. Marzouk, and H. H. Ammar, "Classifying upper limb activities using deep neural networks." in *AICV*, 2020, pp. 268–282.

[11] N. A. Mohamed, A. T. Azar, N. E. Abbas, M. A. Ezzeldin, and H. H. Ammar, "Experimental kinematic modeling of 6-dof serial manipulator using hybrid deep learning." in *AICV*, 2020, pp. 283–295.

[12] H. A. Ibrahim, A. T. Azar, Z. F. Ibrahim, H. H. Ammar, A. Hassanien, T. Gaber, D. Oliva, and F. Tolba, "A hybrid deep learning based autonomous vehicle navigation and obstacles avoidance." in *AICV*, 2020, pp. 296–307.

[13] A. S. Sayed, A. T. Azar, Z. F. Ibrahim, H. A. Ibrahim, N. A. Mohamed, and H. H. Ammar, "Deep learning based kinematic modeling of 3-rrr parallel manipulator." in *AICV*, 2020, pp. 308–321.

[14] A. T. Azar, A. Koubaa, N. Ali Mohamed, H. A. Ibrahim, Z. F. Ibrahim, M. Kazim, A. Ammar, B. Benjdira, A. M. Khamis, I. A. Hameed *et al.*, "Drone deep reinforcement learning: A review," *Electronics*, vol. 10, no. 9, p. 999, 2021.

[15] A. Koubâa, A. Ammar, M. Alahdab, A. Kanhouch, and A. T. Azar, “Deepbrain: Experimental evaluation of cloud-based computation offloading and edge computing in the internet-of-drones for deep learning applications,” *Sensors*, vol. 20, no. 18, p. 5240, 2020.

[16] T. Guo, J. Dong, H. Li, and Y. Gao, “Simple convolutional neural network on image classification,” in *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*. IEEE, 2017, pp. 721–724.

[17] A. Sherstinsky, “Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network,” *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.

[18] B. Calabrese, R. Velázquez, C. Del-Valle-Soto, R. de Fazio, N. I. Giannoccaro, and P. Visconti, “Solar-powered deep learning-based recognition system of daily used objects and human faces for assistance of the visually impaired,” *Energies*, vol. 13, no. 22, p. 6104, 2020.

[19] Y. Lin, K. Wang, W. Yi, and S. Lian, “Deep learning based wearable assistive system for visually impaired people,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[20] M. F. Zamir, K. B. Khan, S. A. Khan, and E. Rehman, “Smart reader for visually impaired people based on optical character recognition,” in *International Conference on Intelligent Technologies and Applications*. Springer, Singapore, 2019, pp. 79–89.

[21] M. Afif, R. Ayachi, Y. Said, E. Pissaloux, and M. Atri, “An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation,” *Neural Processing Letters*, pp. 1–15, 2020.

[22] R. Tasnim, S. T. Pritha, A. Das, and A. Dey, “Bangladeshi banknote recognition in real-time using convolutional neural network for visually impaired people,” in *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE, 2021, pp. 388–393.

[23] R. Cheng, W. Hu, H. Chen, Y. Fang, K. Wang, Z. Xu, and J. Bai, “Hierarchical visual localization for visually impaired people using multimodal images,” *Expert Systems with Applications*, vol. 165, p. 113743, 2021.

[24] A. Koubaa and A. T. Azar, “Deep learning for unmanned systems,” 2021.

[25] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[26] A. Graves, “Long short-term memory,” in *Supervised sequence labelling with recurrent neural networks*. Springer, 2012, pp. 37–45.

[27] S. Yan, “Understanding lstm networks,” *Online). Accessed on August*, vol. 11, 2015.

[28] M. Hodosh, P. Young, and J. Hockenmaier, “Framing image description as a ranking task: Data, models and evaluation metrics,” *Journal of Artificial Intelligence Research*, vol. 47, pp. 853–899, 2013.

[29] J. Johnson, A. Karpathy, and L. Fei-Fei, “Densecap: Fully convolutional localization networks for dense captioning,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4565–4574.

[30] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a method for automatic evaluation of machine translation,” in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.