Optically processed Kannada script realization with Siamese neural network model

Ambili Parathra Sreedharanpillai[1], Biku Abraham[2], Arun Kotapuzakal Varghese[1]
[1]School of Computer Science and Applications, REVA University, Bangalore, India
[2]Department of Computer Applications, SAINTGITS College of Engineering, Kerala, India

Article Info

Article history:

Received Jul 11, 2023
Revised Oct 15, 2023
Accepted Nov 9, 2023

Keywords:

Deep neural network
dissimilarity score
Optical character recognition
deep learning
Siamese neural networks

ABSTRACT

Optical character recognition (OCR) is a technology that allows computers to recognize and extract text from images or scanned documents. It is commonly used to convert printed or handwritten text into machine-readable format. This Study presents an OCR system on Kannada Characters based on siamese neural network (SNN). Here the SNN, a Deep neural network which comprises of two identical convolutional neural network (CNN) compare the script and ranks based on the dissimilarity. When lesser dissimilarity score is identified, prediction is done as character match. In this work the authors use 5 classes of Kannada characters which were initially preprocessed using grey scaling and convert it to pgm format. This is directly input into the Deep convolutional network which is learnt from matching and non-matching image between the CNN with contrastive loss function in Siamese architecture. The Proposed OCR system uses very less time and gives more accurate results as compared to the regular CNN. The model can become a powerful tool for identification, particularly in situations where there is a high degree of variation in writing styles or limited training data is available.

Corresponding Author:

Biku Abraham
Department of Computer Applications, SAINTGITS College of Engineering
Kottayam, India
Email: biku.abraham@saintgits.org

1. INTRODUCTION

Optical character recognition (OCR) is a technology that enables the conversion of handwritten or printed text into digital format that can be processed and edited by computers [1]. It has been widely used in various applications, such as document digitization, language translation, and information retrieval. However, developing an OCR system for non-Latin scripts, such as Kannada, is challenging due to the complexity and diversity of the script. Siamese neural networks (SNNs) have been used for various applications, including face recognition, signature verification, text similarity, image retrieval, document matching, speech recognition, recommendation systems, and bioinformatics applications [2]-[5]. This Research shows the development of Kannada OCR system that can recognize the Kannada Swar with high accuracy and precision characters which were trained. With the advancement of deep learning techniques and the availability of large datasets, SNNs are becoming increasingly popular in various industries, including finance, healthcare, and entertainment.

Research Gap Analysis: many research and experiment of OCR is carried forward with convolutional network and recurrent neural network (RNN) which has an upper hand towards finding patterns in images but it takes lot of data and computation resource for training the model [6]-[8]. Siamese model uses far less data than that of traditional convolutional neural network (CNN) and RNN and very few research has been done with OCR with Siamese network. Since Kannada as a language has curves in its writing style Siamese network excels in identifying handwritten characters [9], [10]. The key considerations in the research are: i) Collect and pre-process the data, including image cropping, resizing, and normalization. In the case of OCR, this may involve converting text images to grayscale or binary format, ii) Create a dataset of image pairs, where each pair contains two images of the same or different characters. Label the pairs as positive (1) or negative (0), depending on if it is the same character then that pair will be labelled 1 and if they are having different characters then that pair will be labelled 0, iii) Define the siamese neural network architecture with two identical networks that share the same set of weights. Both the network architecture works are based on CNN [11]-[14]. Two images are passed as input at the same time to both branches of the siamese networks and then are made to train on it and at the end the output of both the networks are compared using distance or similarity metrics, iv) Use of contrastive loss function to reduce the distance between positive data pairs and increase the distance between negative data pairs.

2. LITERATURE REVIEW

Siamese neural networks have been extensively studied for natural language processing tasks, such as sentence similarity, paraphrase identification, and text classification. Koch *et al*. [15] studied the one-shot image classification and they first learnt about a neural network that can discriminate between the class-identity of image pairs as that forms the primary step for image recognition Imposing local affine transformations on the strokes and overlaying them into a composite image the learning produced better results. OCR is an important application of Siamese networks, but limited researches were carried out on applying them for Kannada character recognition.

CNN has been the go-to model for OCR for any lingual as they learn in a hierarchal form which is helpful in recognizing the image or the characters. Other methods such as RNN and Long short-term memory (LSTM) also have been experimented with OCR where LSTM used to handle long term dependency which also fairs well with cursive characteristics of a character [13], [16], [17]. CNNs are highly effective for image recognition tasks, but they have some disadvantages. These include complexity, over fitting, lack of interpretability, difficulty with variable-sized inputs, limited context awareness, limited rotation and translation invariance, and limited object orientation and position. CNNs are designed to be invariant to small translations in the input image, but may not be as effective at handling large translations or rotations.

Kannada as a language while making a OCR have a potential disadvantages with CNN, RNN as per lack of training data, difficulty with complex scripts, dependence on pre-processing, difficulty with handwriting recognition, and computationally expensive can make it difficult to deploy an OCR system in real-time or on resource-constrained devices [11], [18], [19]. Kannada is a complex script with multiple components or diacritics, making it difficult to design a CNN that can effectively recognize all of these variations. Handwriting is highly variable and can be difficult to model effectively with a CNN.

Despite the variety of models and techniques used in previous research, there is more sophisticated approaches which was explained in this study where SNNs which is used where the network consists of two identical sub-networks, often called "twin" networks that share the same weights and architecture [20]. Each sub-network takes in one input image and processes it through a series of convolutional and pooling layers to extract relevant features. The extracted feature maps are then flattened and fed through a series of fully connected layers that output a feature vector for each input. Overall this methodology require fewer training samples than CNNs due to their similarity or distance learning. They can be adapted to new tasks by fine-tuning a pre-trained network on the new data [21], [22].

3. METHODOLOGY

Siamese networks can be used to compare two different images of the same character or word and determine whether they represent the same or different characters/words [23]-[25]. This can be useful in scenarios where the input data is noisy or has variations in style or orientation. One of the main advantages of Siamese networks is that they can be trained with relatively small amounts of labelled data, making them useful in applications where large datasets are not available. Additionally, Siamese networks can be trained end-to-end, allowing them to learn features directly from the input data without the need for hand-engineered features. Figure 1 represents the proposed system architecture.

The proposed approach for Kannada script identification will be complete in mainly five steps. The procedure is given below. The proposed procedure for character identification:

Step 1: Data collection and pre-processing:

For each image in the dataset:

a. Load the image.
b. Crop and resize the image to a specified size (e.g., (224, 224)).

c. Convert the image to grayscale or binary format (for OCR).
d. Normalize pixel values to a specific range (e.g., [0, 1] or [-1, 1]).
Step 2: Create image pairs and labels:
Initialize empty arrays for image pairs and labels.
For each image in the dataset:
a. Select another image randomly from the dataset.
b. Create a pair containing the current image and the selected image.
c. Assign a label of 1 if both images are of the same character, otherwise assign a label of 0.
d. Append the image pair and label to their respective arrays.
Step 3: Siamese neural network architecture:
Define a CNN architecture with convolutional layers, pooling layers, and fully connected layers.
Duplicate the architecture to create two identical network branches (Siamese networks).
Share the same set of weights between both branches.
Each branch takes an input image and processes it through the CNN layers.
The output of each branch is a feature vector.
Step 4: Training:
Initialize the siamese network with shared weights.
Define a contrastive loss function:
a. For each image pair and its label:
- Forward pass both images through their respective Siamese branches.
- Calculate the Euclidean distance or another similarity metric between the output feature vectors.
- Compute the contrastive loss based on the predicted similarity and the actual label.
- Backpropagate the loss through both branches to update the shared weights.
Repeat the training process for a specified number of epochs.
Step 5: Inference:
Given a new pair of images:
a. Forward pass both images through their respective Siamese branches.
b. Calculate the Euclidean distance or another similarity metric between the output feature vectors.
c. If the distance is below a certain threshold, classify the pair as a positive match (same character), otherwise as a negative match (different characters).

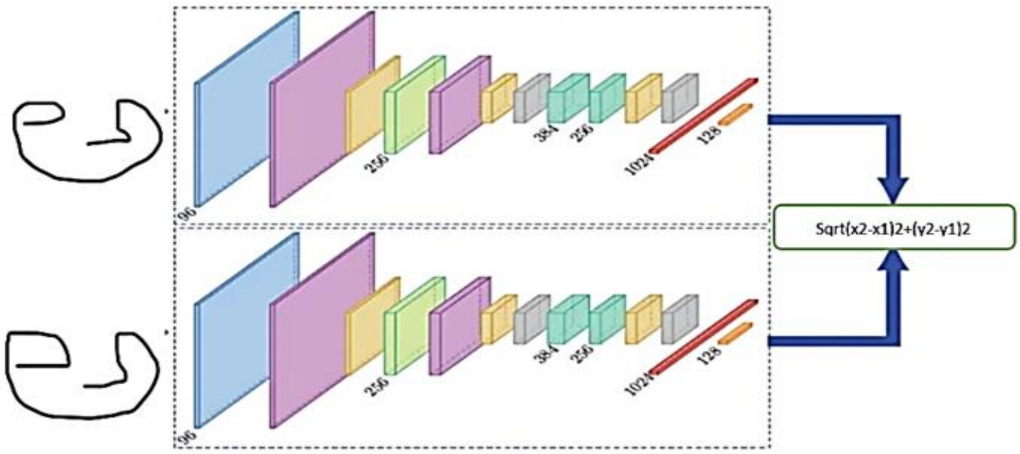

Figure 1. The proposed siamese neural network architecture for Kannada script

Once the feature vectors have been generated, a distance metric is computed between them to determine their similarity. This distance metric can be computed using various techniques such as Euclidean distance, cosine distance, or contrastive loss function. This network takes in two inputs and computes a similarity score between them, which can be used to determine if they belong to the same class or not. The Siamese network is trained using a contrastive loss function, which helps to optimize the network's parameters for better accuracy.

Adam optimizer is a gradient-based optimization algorithm used in deep learning models to update the weights of the neural network during training [8]. Adam optimizer uses the first and second moments of the gradients to estimate the adaptive learning rates for each parameter in the neural network. It works by computing the current gradient of the loss function with respect to the parameters, then updating the parameters in the opposite direction of the gradient, with a learning rate proportional to the first moment divided by the square root of the second moment. It can handle sparse gradients and noisy data, has low memory requirements, and is computationally efficient. It also has automatic learning rate tuning, which helps to converge faster and more accurately to the global minimum.

The contrastive loss function (clf) is a loss function used in the training of Siamese neural networks to measure the similarity between two inputs. It takes two inputs and a label as input and penalizes the model if the distance between positive pairs is greater than a margin value and the distance between negative pairs is smaller than the margin value. The positive term encourages the network to make similar data points close together and the negative term encourages the network to push dissimilar data points apart. The overall loss for a pair of data points is calculated as a combination of positive term and the negative term.

$$clf = yd^2 + (1-y)\max(margin - d, 0)^2 \tag{1}$$

Where ‘y’ is the label and ‘d’ is the distance between the two inputs.
‘d’ is calculated as

$$d = \sqrt{(x2-x1)^2 + (y2-y1)^2} \tag{2}$$

The label y is 1 if positive pair and 0 if negative pair. If the inputs are a positive pair (y=1) and if the labels are of the same class (positive pair) or different classes (negative pair), the loss function penalizes them if the margin value is greater than 0.

4. RESULTS AND DISCUSSION

Unlike other classification and regression models which gives out classes and continuous values as output, Siamese neural network provides the distance between the given input by considering the distance between the input class and the class which is been passed. Siamese network is trained to ensure that the embeddings of similar pairs are close together in this feature space, while embeddings of dissimilar pairs are far apart. The output is based on the similarity and dissimilarity values which is achieved. The lesser the dissimilarity score it is likely that the input value is same as the class. Since the subnetworks produce embeddings for each input data point and these subnetworks share the same architecture and weight, more accuracy in the result is achieved. Additionally, hyperparameters like learning rate, batch size, and threshold values are tuned for optimal performance.

The study used Kannada character set of 85,697 characters belonging to approximately 97 classes with 205-490 sample images in each class. The study actually classified five classes of Kannada characters using Siamese Neural Network with almost 350 sample images in each class. Figure 2 is a representation of sample characters from the dataset.

Figure 2. Sample charaters from dataset

Figure 3 and Figure 4 depicts the results obtained in terms of dissimilarity index. As evident from Figure 3, similar swaras are showing low dissimilarity index. The dissimilar swaras are identified correctly shown in Figure 4. The high dissimilarity index values depicted identified them correctly. The study also identified similar characters with angular displacement looking as dissimilar.

4.1. Evaluation and deployment of the model

The trained model is evaluated by separate test set which checks the accuracy and performance by the metrics of precision, F1-Score and Recall, Dissimilarity. The deployment of the model was done on Kiviy tool used for user interface creation and it works well on python language [26]. The training and validation accuracy are presented in Figure 5 and training and validation loss graphs is Figure 6.

Major advantages of proposed system are: i) Effective identification of handwritten and Optically processed Kannada scripts. ii) One of the most precise models in deep learning by using two CNNs to identify accurate dissimilarity between the input characters. iii)The model can become a powerful tool for identification, particularly in situations where there is a high degree of variation in writing styles or limited training data is available.

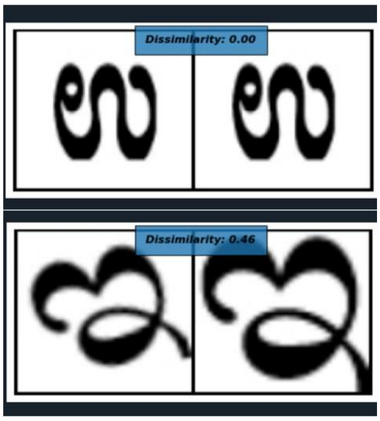

Figure 3. Similar swaras showing low dissimilarity

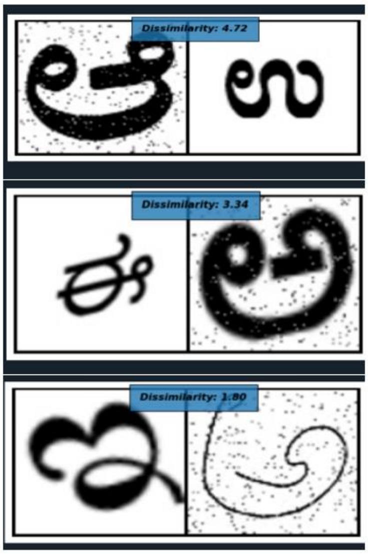

Figure 4. Dissimilar swaras showing high dissimilarity

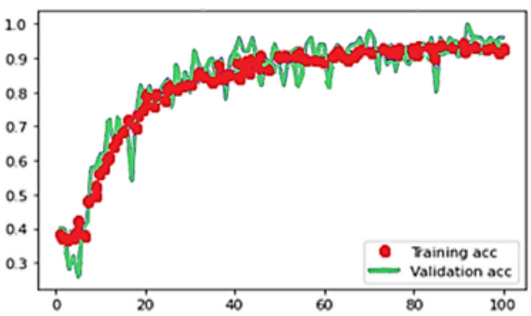

Figure 5. Accuracy of training and validation

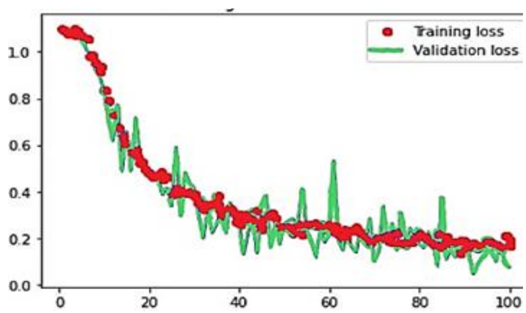

Figure 6. Training and validation loss

5. CONCLUSION

This Research illustrates a SNN with custom training methodology on a PyTorch framework. It successfully classifies all the trained classes of Kannada characters and gives the dissimilarity of each class with respect to the newly passed image. The developed SNNs have been shown to be effective in handwritten character recognition, where there is a high degree of variation in writing styles. This methodology also supports when data to be trained are very diverse which are true for handwritten texts. Also fairs well compared to the RNN and CNN which uses more computation, training data and memory as compared to the Siamese network. In short Siamese network can be fitting to made an OCR model and also gives promising results.

ACKNOWLEDGEMENTS

Authors acknowledge the support from REVA University, Bangalore for the facilities provided to carry out the research.

REFERENCES

[1] K. Sheshadri, P. K. T. Ambekar, D. P. Prasad, and R. P. Kumar, "An OCR system for printed Kannada using k-means clustering," in *2010 IEEE International Conference on Industrial Technology*, 2010, pp. 183–187. doi: 10.1109/icit.2010.5472676.

[2] N. S. Rani, A. C. Subramani, A. Kumar P., and B. R. Pushpa, "Deep Learning Network Architecture based Kannada Handwritten Character Recognition," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Jul. 2020, pp. 213–220. doi: 10.1109/icirca48905.2020.9183160.

[3] S. Tejas and K. K. Dutta, "Siamese Neural Networks for Kannada Handwritten Dataset," in *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)*, Oct. 2022, pp. 1–6. doi: 10.1109/gcat55367.2022.9971971.

[4] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 815–823. doi: 10.1109/cvpr.2015.7298682.

[5] S. Pingali, P. Mondal, D. Chakder, S. Saha, and A. Ghosh, "Towards Developing a Multi-Modal Video Recommendation System," in *2022 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2022, pp. 1–8. doi: 10.1109/ijcnn55064.2022.9892382.

[6] S. Siddiqua, N. C, and S. S. Manvi, "Recognition of Kannada Characters in Scene Images using Neural Networks," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, Nov. 2019, pp. 146–150. doi: 10.1109/iciip47207.2019.8985672.

[7] K. Asha and H. K. Krishnappa, "Kannada Handwritten Document Recognition using Convolutional Neural Network," in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, Dec. 2018, pp. 299–301. doi: 10.1109/csitss.2018.8768745.

[8] D. Chicco, "Siamese Neural Networks: An Overview," in *Artificial Neural Networks*, United States: Springer US, 2020, pp. 73–94. doi: 10.1007/978-1-0716-0826-5_3.

[9] J. Wang *et al.*, "Learning Fine-Grained Image Similarity with Deep Ranking," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1386–1393. doi: 10.1109/cvpr.2014.180.

[10] J. Wang, S. Kumar, and S.-F. Chang, "Semi-supervised hashing for scalable image retrieval," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 3424–3431. doi: 10.1109/cvpr.2010.5539994.

[11] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 1335–1344. doi: 10.1109/cvpr.2016.149.

[12] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1701–1708. doi: 10.1109/cvpr.2014.220.

[13] L. Gomez, A. Nicolaou, and D. Karatzas, "Improving patch-based scene text script identification with ensembles of conjoined networks," *Pattern Recognition*, vol. 67, pp. 85–96, Jul. 2017, doi: 10.1016/j.patcog.2017.01.032.

[14] S. P. T. Parashivamurthy and S. V. Rajashekararadhya, "Recognition of Kannada Character Scripts Using Hybrid Feature Extraction and Ensemble Learning Approaches," *Cybernetics and Systems*, pp. 1–36, Mar. 2023, doi: 10.1080/01969722.2023.2175145.

[15] G. Koch, "Siamese Neural Networks for One-shot Image Recognition," University of Toronto, 2011.

[16] A. P S and B. Abraham, "A Predictive Model for Student Employability Using Deep Learning Techniques," *ECS Transactions*, vol. 107, no. 1, pp. 10149–10158, Apr. 2022, doi: 10.1149/10701.10149ecst.

[17] R. Li, M. Moh, and T.-S. Moh, "Siamese Neural Networks for Content-based Visual Art Recommendation," in *2023 17th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, Jan. 2023, pp. 1–8. doi: 10.1109/imcom56909.2023.10035645.

[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[19] P. Arora and T. Chaspari, "Exploring Siamese Neural Network Architectures for Preserving Speaker Identity in Speech Emotion Classification," in *Proceedings of the 4th International Workshop on Multimodal Analyses Enabling Artificial Agents in Human-Machine Interaction*, Oct. 2018, pp. 15–18. doi: 10.1145/3279972.3279980.

[20] V. Oriol, B. Charles, L. Tim, kavukcuoglu koray, and W. Daan, "Matching Networks for One Shot Learning," *Advances in Neural Information Processing Systems*, pp. 3630–3638, 2016, [Online]. Available: http://arxiv.org/abs/1606.04080

[21] B. Abraham and P. S. Ambili, "An Enhanced Career Prospect Prediction System for Non-computer Stream Students in Software Companies," in *Computational Intelligence for Engineering and Management Applications*, singa: Springer Nature Singapore, 2023, pp. 811–819. doi: 10.1007/978-981-19-8493-8_60.

[22] A. Hajavi and A. Etemad, "Siamese Capsule Network for End-to-End Speaker Recognition in the Wild," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jun. 2021, pp. 7203–7207. doi: 10.1109/icassp39728.2021.9414722.

[23] S. Ramoji, P. Krishnan, and S. Ganapathy, "PLDA inspired Siamese networks for speaker verification," *Computer Speech & Language*, vol. 76, p. 101383, Nov. 2022, doi: 10.1016/j.csl.2022.101383.

[24] J. Yuan, H. Guo, Z. Jin, H. Jin, X. Zhang, and J. Luo, "One-shot learning for fine-grained relation extraction via convolutional siamese

neural network," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec. 2017, pp. 2194–2199. doi: 10.1109/bigdata.2017.8258168.

[25] G. M. Messa, F. Napolitano, S. H. Elsea, D. di Bernardo, and X. Gao, "A Siamese neural network model for the prioritization of metabolic disorders by integrating real and simulated data," *Bioinformatics*, vol. 36, no. 2, pp. i787–i794, Dec. 2020, doi: 10.1093/bioinformatics/btaa841.

[26] Y. Aufar and I. S. Sitanggang, "Face recognition based on Siamese convolutional neural network using Kivy framework," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 2, pp. 764–772, May 2022, doi: 10.11591/ijeecs.v26.i2.pp764-772.

BIOGRAPHIES OF AUTHORS

Ambili Parathra Sreedharanpillai holds a Ph.D. degree in Computer Science from Anna University, Chennai from 2017. She received her Master of Computer Applications degree from the University of Madras. She is currently an Associate Professor at the School of Computer Science and Applications of REVA University, Bangalore, India. Her research includes Data Mining, Data Analytics, Machine Learning, Deep Learning and Cyber Security. She has published several papers in peer reviewed journals and is a reviewer of several International Journals and Conferences. She has acted as a Resource Person for many faculty development programmerss. She is a research supervisor at REVA University and is currently guiding five scholars. She can be contacted at email: ambili.ps@reva.edu.in or amispillai20@gmail.com.

Biku Abraham received her Ph.D. in Computer Science from Anna University, Chennai and Master of Computer Applications and M.Phil. in Computer Science from Bharathiar University, Coimbatore. Of her 20 years of teaching career in post-graduation, she has been working as Professor since 2016 at Saintgits College of Engineering. Dr. Biku held an administrative position for five years and is currently leading the curriculum design of the department. She published several papers in peer reviewed journals and served as the reviewer of many journals and conferences. Her areas of interest include Big Data, Deep Learning, and Information systems. She can be contacted at email: biku.abraham@saintgits.org.

Arun Kotapuzakal Varghese is an exceptional professional with a Master of Science (MSc) degree in Data Science and Bachelors degree in Computer Applications (BCA). Currently he is working as Metaverse AI Engineer at Wise Works, Bangalore. He has presented the project SIPAS at IIT-Kanpur in 2022- an ingenious smart wheelchair designed to empower disabled individuals with comprehensive health monitoring capabilities. With a strong research focus on AI, Metaverse, Space and Technology Arun is committed to pushing the boundaries of innovation and creating a positive impact on society. He can be contacted at email: kv.arun32@gmail.com.