Deep Learning Approaches for Age-based Gesture Classification in South Indian Sign Language

Ramesh M. Badiger
Department of Computer Science and Engineering, Tontadarya College of Engineering, India
rameshmbadiger@gmail.com

Rajesh Yakkundimath
Department of Computer Science and Engineering, KLE Institute of Technology, India
rajeshymath@gmail.com

Guruprasad Konnurmath
School of Computer Science and Engineering, KLE Technological University, Hubballi, India
guruprasad.konnurmath@kletech.ac.in

Praveen M. Dhulavvagol
School of Computer Science and Engineering, KLE Technological University, Hubballi, India
praveen.md@kletech.ac.in

Received: 4 January 2024 | Revised: 21 January 2024 | Accepted: 29 January 2024

ABSTRACT

This study focuses on recognizing and categorizing South Indian Sign Language gestures based on different age groups through transfer learning models. Sign language serves as a natural and expressive communication method for individuals with hearing impairments. The intention of this study is to develop deep transfer learning models, namely Inception-V3, VGG-16, and ResNet-50, to accurately identify and classify double-handed gestures in South Indian languages, like Kannada, Tamil, and Telugu. A dataset comprising 30,000 images of double-handed gestures, with 10,000 images for each considered age group (1-7, 8-25, and 25 and above), is utilized to enhance and modify the models for improved classification performance. Amongst the tested models, Inception-V3 achieves best performance with test precision of 95.20% and validation accuracy of 92.45%, demonstrating its effectiveness in accurately categorizing images of double-handed gestures into ten different classes.

Keywords-sign language; age group; gesture identification; transfer learning; Inception-V3; VGG-16; ResNet-50

I. INTRODUCTION

The rapid evolution of AI technologies has opened up significant opportunities to develop the life quality for individuals with disabilities. According to the World Health Organization, approximately 1.2 billion people, or 15% of the worldwide population, live with some type of disability. Recognizing the need for inclusivity, AI solutions to assist the disabled community in their daily activities have been leveraged. Consequently, AI, accessibility, and disabilities have become closely intertwined, with extensive research in AI and machine learning directed towards accommodating the daily lives of disabled individuals. AI has made significant strides in facilitating non-verbal communication for persons with impairments, including the recognition and understanding of sign language and gestures. Sign language, which relies on hand gestures and body motions, serves as a vital means of communication with impaired people. Advancements in computer vision and pattern recognition have been instrumental in enabling AI systems to interpret and respond to sign language effectively. But, despite the recent progress, the fast and accurate recognition of hand gestures remains a challenging task [1]. Researchers are continuously working to get better performance of AI models in recognizing and categorizing these gestures. Gestures may vary in the speed, orientation, and hand alignment. Age category divisions for hand gesture performance are not rigidly defined, as there is considerable overlap and individual variations in hand gesture usage. However, we can broadly categorize the latter into the following age groups:

- Early Childhood (1-7 years), Early Childhood refers to the early stages of development in which individuals have limited skills and communication abilities. They use simple gestures like reaching out for objects, waving, and pointing to objects.
- Adolescence (8-25 years). As children advance through middle childhood to adolescence, their gesture vocabulary expands further, and they become more adept at using gestures for social interaction.
- Adulthood (25+ years). Adults typically possess a fully developed gesture repertoire. They use gestures to enhance verbal communication, convey complex ideas, express emotions, and adhere to cultural norms and social cues.

Gesture accuracy gradually changes as individuals grow up. Table I provides a comprehensive list of double-handed gestures considered for the research work, alongside with their respective letters in various South Indian languages based on different age groups.

TABLE I. LABEL ENCODING/NUMERICAL CODE ASSIGNMENTS FOR THE CONSIDERED DOUBLE-HAND GESTURES

S. No.	**Word in Kannada/Telugu/Tamil/English**	**Class label (alphanumerical code) in years**		
		1-7	**8-25**	**≥ 25**
1	ಮೀನು /చేప /மீன் /Fish	1D	2D	3D
2	ಯಾವುದು /ఏది /எப்படி /Which	1E	2E	3E
3	ಅಡಿ /అడుగు /அடி /Foot	1F	2F	3F
4	ಚರ್ಮ /చర్మం /தோல் /Skin	1J	2J	3J
5	ನವಿಲು /నెమలి /மயில் /Peacock	1M	2M	3M
6	ಸಮಯ /సమయం /நேரம் /Time	1N	2N	3N
7	ಉದ್ದ /పొడవు /நீளம் /Length	1P	2P	3P
8	ಚಿಟ್ಟೆ /తీతీతీబ్బుటి /புறா /Butterfly	1T	2T	3T
9	ಮನೆ /ఇల్లు /வீடு /House	1W	2W	3W
10	ಏನು / ఏమి /என்ன /What	1X	2X	3X

The literature survey on pattern recognition applications in the identification of sign language gestures reveals the following key findings. Authors in [2] employed a time-distributed Convolution Neural Network (CNN) and a Gated Recurrent Unit (GRU) to extract skin features from hand images. The models achieved a performance accuracy of 96.5% in categorizing hand images into 17 age groups ranging from 18 to 75 years old. Authors in [3] utilized MRELBP to extract features from right and left hand images. The images were classified by person, age, and gender, reporting accuracies of 91.4%, 85.9%, and 92.6%, respectively. Authors in [4] presented age detection on an image dataset using combinations of deep learning and image processing techniques, achieving an accuracy of 91%. The literature survey shows that noteworthy research has been explored in the area of sign language and hand gesture identification, with a particular emphasis given on deep learning techniques for recognizing human poses and scenes from images [5-9]. However, there is a noticeable gap in the identification of South Indian Sign Language (SISL) gesture images regarding different age groups [10]. This research article aims to address this issue and provide a novel method for recognizing SISL gestures across various age groups.

II. PROPOSED METHODOLOGY

The proposed method consists of two primary stages: the first involves organizing the image dataset into various age groups, while the second entails employing deep learning methods to classify the gestures; Figure 1 illustrates a block diagram showcasing the stages involved.

A. *Data Preparation*

The dataset preparation process involves selecting 10 categories of double-hand gestures based on different age groups. High-resolution images were captured using a Nikon D3300 camera with a resolution of 24.2 megapixels. The images were taken against black or green backgrounds under natural lighting conditions. Each image has an original size of 1080 × 2400 pixels. For the initial age group (age ranging from 1 to 7), the dataset consists of 5000 images, with 500 images per gesture category. Image extension methods were applied to expand the dataset to 10,000 images, enhancing its diversity. To ensure efficient processing and storage, all images were resized to 300 × 300 pixels. The dataset was divided into three subsets: training (70%), validation (15%), and testing (15%).

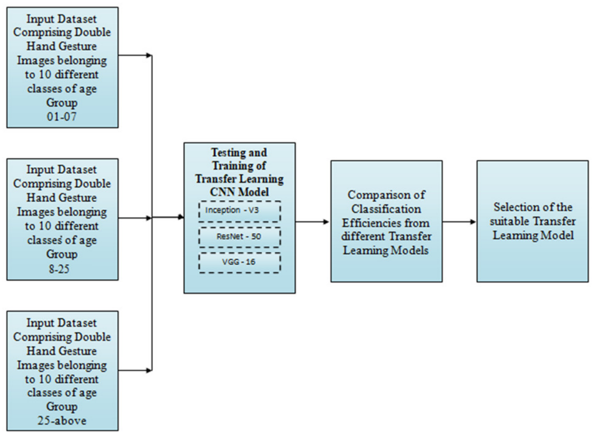

Fig. 1. Block diagram of the proposed methodology.

The training subset contains 7000 images, while the validation and testing subsets contain 1500 images each. This division ensures the acquisition of enough data for the training and reliable estimation of the models. The same methodology was applied to the other age groups. By following this systematic approach, a comprehensive dataset of double-hand gestures is prepared, covering various age groups, facilitating the development, and estimating the accuracy of recognition models. Figure 2 displays a sample of the utilized dataset.

B. *CNN Classifiers*

In this study, three prominent transfer learning CNN [11] models are utilized for the recognition and classification of double-handed gesture images. These models include Inception-V3 [12], VGG-16 [13], and ResNet-50 [14], acknowledged for their effectiveness in image recognition tasks. The training process of the CNN models on the prepared dataset follows a specific set of steps, as demonstrated in the following Algorithm.

```
Algorithm: Training Steps for CNN Models
Inputs: Prepared dataset of double-handed
gesture images, Chosen CNN models
(Inception-V3, VGG-16, or ResNet-50), and
hyper parameters: learning rate, batch
size, number of epochs.
Outputs: Trained CNN model
Start
Step 1. Load the prepared dataset
Step 2. Initialize the chosen CNN model
Step 3. Set the hyper parameters for
training
Step 4. Split the dataset into training,
validation, and testing subsets
Step 5. Training Phase:
for each epoch in the specific number of
epochs
do:
for each batch in the training subset
do:
Load a batch of images and their
corresponding labels
Perform forward propagation through the
CNN model
Compute the loss using the loss function
Perform backward propagation to calculate
the gradients
Update the model's parameters using the
optimization algorithm
Step 6. Validation Phase:
for each batch in the validation subset
do:
Load a batch of images and their
corresponding labels
Perform forward propagation through the
CNN model
Compute the validation loss and accuracy
Step 7. Evaluation:
After training, evaluate the trained model
on the testing subset
do:
for each batch in the testing subset
do:
Load a batch of images and their
corresponding labels
Perform forward propagation through the
trained CNN model
Compute the testing accuracy
Step 8. Fine-tuning and Hyper parameter
Tuning:
Adjust hyper parameters and fine-tune the
model based on the validation results
Step 9. Repeat steps 5-8 until
satisfactory performance is achieved
Step 10. Save the trained model for future
use in gesture recognition and
classification tasks
Stop.
```

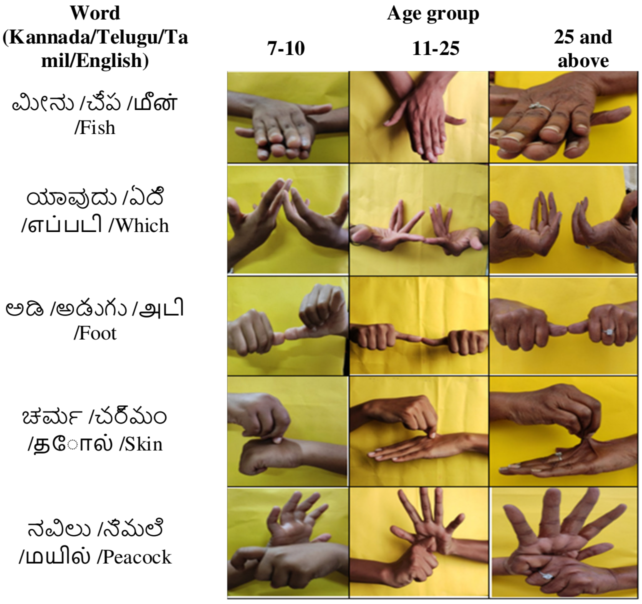

Fig. 2. Dataset samples.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The experiments on double-handed gesture classification for different age categories were conducted using the Deep Learning Toolbox given by the MATLAB R2022b platform. The pre-trained CNN models employed in the experiments were imported and prepared for transfer learning by modifying the properties of suitable layers utilizing the Deep Network Designer application. Specifically, the last learnable layer and the output or classification layer were replaced to align with the classes in the newly constructed double-handed gesture image dataset. To control the training process, specific options were set for the CNN models. The initial learning rate, validation frequency, number of epochs, and mini-batch size were initialized to 0.0001, 10, 30, and 35, respectively. These values were chosen based on experimentation and empirical knowledge to achieve optimal training performance. In terms of activation functions, all hidden layers in the CNN models were activated implementing the Rectified Linear Unit (ReLU) function, which has been usually used in deep learning due to its ability to introduce non-linearity and handle vanishing gradients. The output layer, responsible for classification, was activated utilizing the softmax function, which produces a probability distribution over the different classes. To fine-tune the network and optimize its performance, the Stochastic Gradient Descent (SGD) algorithm was employed as the optimization algorithm. The customized CNN models were trained and validated applying the prepared augmented image dataset. The training progress of each CNN model, along with the corresponding validation accuracy and loss, is monitored and visualized in Figures 3 to 5. These figures provide a graphical representation of the training process, allowing for a better knowledge of the model's performance and convergence.

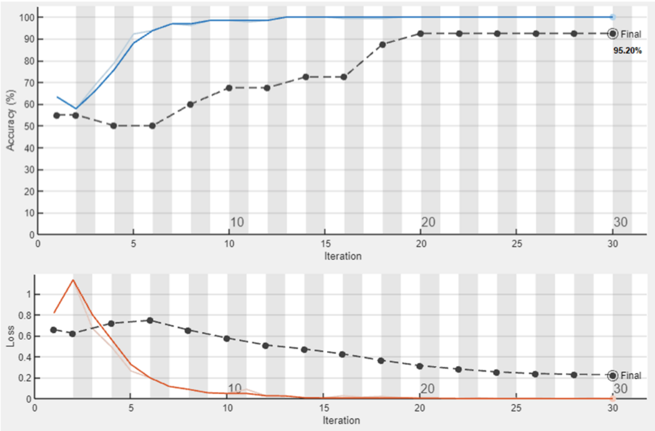

Fig. 3. Validation accuracy and loss graph while training the Inception V3 CNN model.

Figure 6 displays the confusion matrix of the trained Inception V3 CNN model on the test dataset.

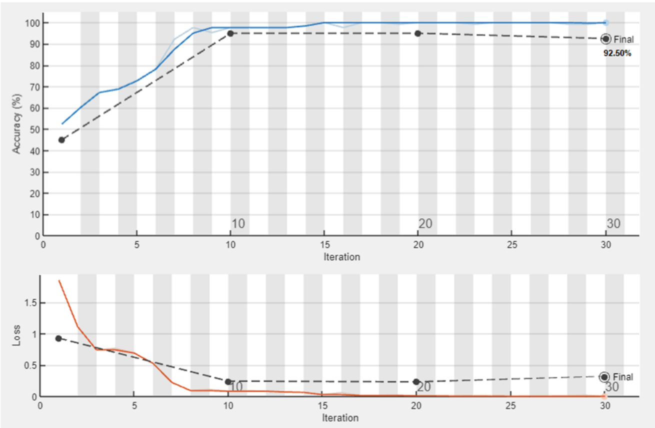

Fig. 4. Validation accuracy and loss graph while training the ResNet 50 CNN model.

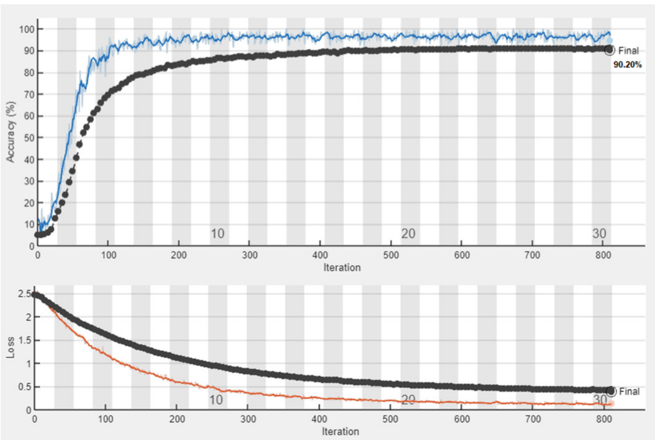

Fig. 5. Validation accuracy and loss graph while training the VGG-16 CNN model.

The confusion matrices provide insights into the performance of the models by illustrating the circulation of the predicted and actual class labels. They permit for the identification of any misclassifications or patterns in the model's predictions, aiding in the measurement of its efficacy in recognizing double-handed gestures. By analyzing the training progress and the confusion matrices, valuable information can be obtained regarding the performance and reliability of the CNN models in classifying double-handed gestures for different age categories. Table II presents the average scores of these evaluation metrics for all the considered CNN models across the classes of double-handed gesture images, which include different age categories. According to the results in Table II, it is apparent that the Inception-V3 model achieves the highest performance among the three models, followed by the ResNet-50 and VGG-16. The Inception-V3 model demonstrates superior evaluation metric scores and achieves the highest validation and testing accuracies of 95.20%, 92.50%, and 90.20% for age groups 1 to 7, 8 to 25, and 25 and above, respectively. The performance comparison analysis provides valuable insights into the effectiveness of the CNN models in classifying double-handed gesture images for different age categories. To visually depict the performance comparison results of all the considered pre-trained CNN models, Figure 7 is provided.

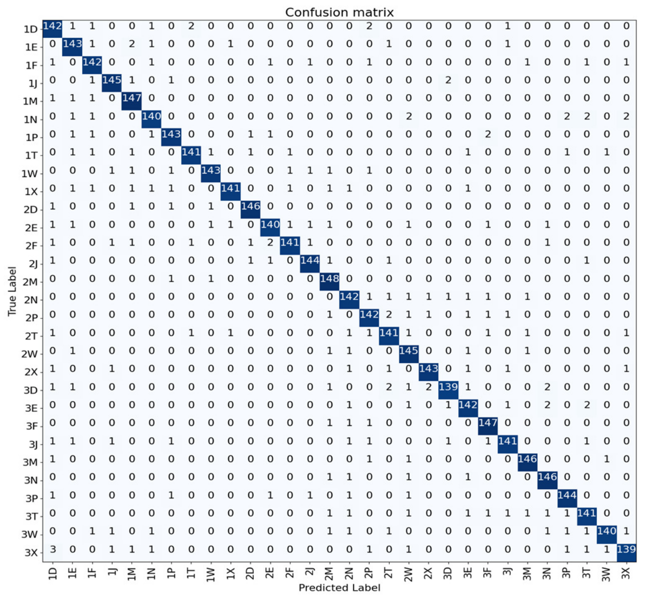

Fig. 6. Confusion matrix of the test dataset for the trained Inception V3 CNN model.

According to the experimental results, the class labeled "2M" representing the word" ನವಿಲು /నెమలి /മയിൽ /Peacock" in the age group from 8 to 25 achieved the maximum validation accuracy among all the double-hand gestures, while the class labeled "2F" representing the word " ಅಡಿ /అడుగు /அடி /Foot" in the age group from 8 to 25 obtained the lowest validation accuracy. These results indicate the varying performance of the CNN models in correctly classifying different double-hand gestures based on their age groups and letters.

TABLE II. EVALUATION METRICS DERIVED FROM THE CONFUSION MATRICES OF THE CONSIDERED CNN MODELS

CNN model	Average performance metrics across all the single-hand gesture classes			Average validation accuracy (%)	Average test accuracy (%)
	Precision	Recall	F1 Score		
Inception-V3	0.9520	0.9520	0.9500	95.20%	96.10%
ResNet-50	0.9247	0.9253	0.923	92.50%	93.20%
VGG-16	0.9027	0.903	0.901	90.20%	91.70%

It can be seen that the Inception-V3 model demonstrated the best performance among all the considered pre-trained CNN models. It achieved the highest evaluation metrics, as well as the maximum validation and testing accuracies for different age groups. The high accuracy values of the Inception-V3 model indicate its efficacy in precisely classifying double-hand gestures based on age groups. The superior performance of the Inception-V3 model suggests its appropriateness for recognizing and categorizing South Indian Sign Language gestures across different age groups. The performance comparison results of all the considered pre-trained CNN models can be seen in Figure 7.

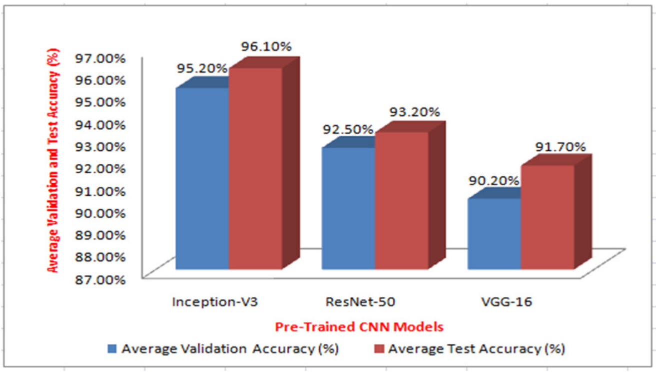

Fig. 7. Performance evaluation results of all the pre-trained CNN models.

IV. CONCLUSION

The current study utilized popular pre-trained CNN models, namely Inception-V3, ResNet-50, and VGG-16, to classify double-hand gesture images across 30 different classes, focusing on 10 signs for different age categories (1–7, 8–25, and 25 and above). These models were customized and fine-tuned to accommodate the specific image classes and improve classification performance. The study results demonstrated impressive performance across all three models. However, Inception-V3 emerged as the top-performing model, achieving an average classification accuracy of 95.20%, indicating its efficiency in accurately classifying double-hand gestures based on age categories. The outcomes of this work have potential applications in constructing automated systems that can identify South Indian sign language gestures from both still images and streaming videos [15-17]. By leveraging these advanced CNN models, communication barriers can be reduced, allowing for easier and more effective communication with the outside world. Future research endeavors could explore the use of the latest CNN models and incorporate publicly available image datasets to further enhance the image dataset employed in this study.

REFERENCES

[1] H. G. Doan and N. T. Nguyen, "Fusion Machine Learning Strategies for Multi-modal Sensor-based Hand Gesture Recognition," *Engineering, Technology & Applied Science Research*, vol. 12, no. 3, pp. 8628–8633, Jun. 2022, https://doi.org/10.48084/etasr.4913.

[2] M. A. Abderrahmane, I. Guelzim, and A. A. Abdelouahad, "Human Age Prediction Based on Hand Image using Multiclass Classification," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, Sakheer, Bahrain, Jul. 2020, https://doi.org/10.1109/ICDABI51230.2020.9325591.

[3] E. Aydemir and R. Alalawi, "Classification Of Hand Images by Person, Age and Gender with The Median Robust Extended Local Binary Model," *Balkan Journal of Electrical and Computer Engineering*, vol. 11, no. 1, pp. 78–87, Jan. 2023, https://doi.org/10.17694/bajece.1171905.

[4] G. Bakshi, A. Aggarwal, D. Sahu, R. R. Baranwal, G. Dhall, and M. Kapoor, "Age, Gender, and Gesture Classification Using Open-Source Computer Vision," in *Emerging Technologies in Data Mining and Information Security*, Singapore, 2023, pp. 63–73, https://doi.org/10.1007/978-981-19-4052-1_8.

[5] S. K. Gupta and N. Nain, "Review: Single attribute and multi attribute facial gender and age estimation," *Multimedia Tools and Applications*, vol. 82, no. 1, pp. 1289–1311, Jan. 2023, https://doi.org/10.1007/s11042-022-12678-6.

[6] W.-B. Horng, C.-P. Lee, and C.-W. Chen, "Classification of Age Groups Based on Facial Features," *Tamkang Journal of Science and Engineering*, vol. 4, no. 3, pp. 183–192, Sep. 2001.

[7] K. R, "Deep Learning for Age Group Classification System," *International Journal of Advances in Signal and Image Sciences*, vol. 4, no. 2, pp. 16–22, Dec. 2018, https://doi.org/10.29284/ijasis.4.2.2018.16-22.

[8] M. Madhiarasan and P. P. Roy, "A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets." arXiv, Apr. 7, 2022, https://doi.org/10.48550/arXiv.2204.03328.

[9] N. S. Russel and A. Selvaraj, "Gender discrimination, age group classification and carried object recognition from gait energy image using fusion of parallel convolutional neural network," *IET Image Processing*, vol. 15, no. 1, pp. 239–251, 2021, https://doi.org/10.1049/ipr2.12024.

[10] R. M. Badiger and D. Lamani, "Deep Learning Based South Indian Sign Language Recognition by Stacked Autoencoder Model and Ensemble Classifier on Still Images and Videos," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 21, pp. 6587–6597, 2022.

[11] N. C. Kundur, B. C. Anil, P. M. Dhulavvagol, R. Ganiger, and B. Ramadoss, "Pneumonia Detection in Chest X-Rays using Transfer Learning and TPUs," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11878–11883, Oct. 2023, https://doi.org/10.48084/etasr.6335.

[12] N. Kumar, A. Hashmi, M. Gupta, and A. Kundu, "Automatic Diagnosis of Covid-19 Related Pneumonia from CXR and CT-Scan Images," *Engineering, Technology & Applied Science Research*, vol. 12, no. 1, pp. 7993–7997, Feb. 2022, https://doi.org/10.48084/etasr.4613.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.

[14] S. Khan, I. Ali, F. Ghaffar, and Q. Mazhar-ul-Haq, "Classification of Macromolecules Based on Amino Acid Sequences Using Deep Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9491–9495, Dec. 2022, https://doi.org/10.48084/etasr.5230.

[15] G. Konnurmath and S. Chickerur, "An Investigation into Power Aware Aspects of Rendering 3D Models on Multi-Core Processors," *Procedia*

Computer Science, vol. 218, pp. 887–898, Jan. 2023, https://doi.org/10.1016/j.procs.2023.01.069.

[16] G. Konnurmath and S. Chickerur, "GPU Shader Analysis and Power Optimization Model," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12925–12930, 2024, https://doi.org/10.48084/etasr.6695.

[17] G. Konnurmath and S. Chickerur, "Power-Aware Characteristics of Matrix Operations on Multicores," *Applied Artificial Intelligence*, vol. 35, no. 15, pp. 2102–2123, Dec. 2021, https://doi.org/10.1080/08839514.2021.1999013.