Hand Gesture Recognition using Deep Learning

Abhinav Thapa
M.Sc Data Analytics
National College of Ireland
Dublin, Ireland
x20259409@student.ncirl.ie

Rupesh Sai Baba Chintakayala
M.Sc. Data Analytics
National College of Ireland
Dublin, Ireland
x21104638 @student.ncirl.ie

Tejas Mahesh Jadhav
M.Sc. Data Analytics
National College of Ireland
Dublin, Ireland
x20206861@student.ncirl.ie

***Abstract*— Hand gestures - a form of non-verbal communication, are now being utilized in developing various Human-Computer Interaction tools such as systems for communicating with deaf-mute people, robot controls, home automation devices and, medical devices. Hand Gesture Recognition systems find their immense application in a wide range of domains which include or may incorporate automated gesture-based utility of machines that are currently being operated manually. Different approaches and combinations of techniques can be used for classifying hand gestures using deep learning. This project aims to explore and develop a system that can recognize and classify hand gestures using deep learning methods and evaluate the extent to which the combination of CNN with LSTM and GRU (Gated Recurrent Unit) model assists in addressing this problem better. Furthermore, the effects of data augmentation on the models were explored. The results of the evaluation metrics suggested that the combination of CNN, GRU, and LSTM teamed with data augmentation, outperformed other methods in terms of both lesser computational expenses, model loss behavior, and better performance accuracy for the chosen Hand Gesture Recognition Kaggle dataset.**

***Keywords*— *Hand Gesture Recognition (HGR), Deep Learning, Classification, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Data augmentation*.**

I. INTRODUCTION

The problem of identifying and classifying hand gestures has found various applications in the field of computer Vision and Human-Computer Interaction [19, 20]. The fundamental goal of a Hand gesture recognition system is to accurately segment the hand gestures [16]. Research and Studies related to the classification of hand gestures has proved useful in the development of tools like sign-language recognition systems, gesture based robotic controls, home automation devices and enhanced gaming consoles.

To bridge the communication gap between the mute-deaf people and others, various researchers and studies came up with different techniques and methods for developing systems capable of operating on both static images and video files to classify hand gestures. [15] developed a system that would accept a live video of the hand gestures as an input and would output a text/audio interpreted output of the gesture. Another study [17] proposed a Gabor-SVM method which pre-processed the images using the Gabor filter and then used the SVM algorithm for classifying the hand gestures.

Along with the technological advancement, the application of hand gesture recognition is being widely explored in developing television without remote. Such television intends to use hand gestures for changing the channels, increase and decrease volume and pause/resume a show [22]. Home automation is another area where gesture recognition is finding its application. Companies are developing devices which turn on and off depending upon the hand gestures [22]. Another example where the application of hand gesture recognition can be explored is for setting an Unmanned Air Vehicle (UAV) for a fight carrier [20]. On flight carriers, the deck officers and crew use gestures to communicate with the pilots. Unlike human pilots, a UAV needs a special tool for understanding gesture command of a deck officer. To solve this problem, UAVs can be equipped with a computer vision system capable of accurately recognizing the set of gestures of a deck officer.

To address the problem of hand gesture recognition, alongside machine learning various deep learning methods and techniques can be employed. To provide a direction on the same front, the aim of this project is to identify and classify hand gestures using deep learning techniques and evaluate how well can the combination of CNN, LSTM and GRU can assist in addressing this problem compared to other deep learning techniques for the chosen dataset. To achieve the project objective, a Hand Gesture Recognition Dataset from the Kaggle data repository [26] has been selected. This dataset encapsulates a set of 20000 (Resolution 640 x240) infrared images containing 10 different hand gestures.

This project follows the given layout: related work is explored in Section II, the proposed methodology and evaluations elaborated in Section III, while results, conclusions and future work in Section IV and V.

II. RELATED WORK

Extensive literature on the gesture recognition problem has been recorded that explored wide arrays of approaches and techniques for both image and video inputs to effectively recognise and classify the desired gestures. Here a study of the past work was done considering a distinction between the inputs as delving in both the approaches would help with investigating a range of methods that were undertaken. This study ranged from investigating the applications of pre-trained models using Transfer Learning [1,15], hyper-parameter optimization of existing models [2,7] for improving performance, employing various feature extraction and segmentation techniques [5,16,17,21], data augmentation [1,3,4], upto RNNs [11] and attention-based deep learning model [6,11,12] architecture for gesture recognition on live videos and surface electromyograph images [3,10].

Two main approaches are typically used for static hand gesture recognition i.e., wearable device-based approach and vision-based approach. Static gesture recognition uses devices such as data glove, electromyograph, electrical impedance tomography and Myo armband with leap motion device [3]. Here a vision-based approach was explored. With its freedom and simplicity of usage and widespread application in multiple domains, vision-based approach comes with its own set of challenges such as hand segmentation, hand skin-tone, hand size, image noise and views. Using a vision-based approach we can classify by extracting important features pertaining to the desired domain-specific use case and then estimating the accuracy. However, this technique may entail designing the networks predominantly for specified conditions and certain applications only, which makes it less generalized in its effectiveness. Thus, meaning that the solution may only work when the gestures are made as per the specified instructions. Deep learning approach solves this conundrum by providing use-case specific solutions with generalized learning and desired estimation accuracy when applied with certain data augmentation and training techniques. In [3], multiple datasets were used for a comparative analysis of an Enhanced Densely connected CNN that improved the feature propagation and gradients flow in the CNN model by employing bottleneck layers for feature reduction and then convolution layer for smoothing unnecessary features. Various data augmentation techniques were also explored to account for data scarcity and limited data diversity while implementing the proposed approach. The experimental results implied on one NUS (National University if Singapore) hand gesture and two ASL (American Sign language) benchmarking datasets established the validity of the superiority of the described technique for greater generalization and accuracy for image classification.

The proposed CNN model in [4] used a uniform He (Weight) initialization algorithm for all ReLU (Rectified Linear) layers and a uniform Xavier initialization for the Softmax layer (for generalization). The dropout regularization and L2 regularization were used to minimize the loss on the training datasets. Data augmentation such as scale, rotation, translation, illumination, colour-conversion and noise were carried to create variation in the dataset. In addition, the model was able to achieve an accuracy of 99.73% with lesser pre-processing. In [5], an alternate method of image segmentation was used instead of optical flow from RBG frames, by using a light-weight semantic segmentation (FASSD-Net) with TSN – Temporal Segment Networks and TSM – Temporal Shift Modules techniques for higher real-time performance accuracy on IPIN video dataset without the high limitation of computational expense.

Another study [9] evaluated gesture recognition using a depth extraction approach and skeleton-based approach, and applying score-level fusion for higher accuracy of upto 89% for skeleton-based and 78.17% for the depth-based networks. CNN was used for feature extraction while RNN for temporal sequence learning in the multiple frames. [13] developed a hand gesture recognition system using LSTM network and a short-range radar - Soli. The system would detect the signal of the gestures using the Constant False Rate Alarm (CFAR) algorithm which was then fed to the LSTM network for making the classification. The LSTM model operated on the motion profile sequences to make the predictions. The proposed system was tested on data collected from 10 different individuals. The evaluation results showcased the system was able to make correct classification with an accuracy of 99.10%. Furthermore, [14] explored the problem of human activity and hand gesture recognition using deep learning and a dataset consisting of 3D sequences of full-body and hand skeleton. To address the problem, a combination of Convolutional Neural Network and Long Short-Term Memory (LSTM) network was proposed. CNN was used for extracting the features from the data, while LSTM was used for carrying out the classification task. The experiment results showcased that the proposed model was a better performing candidate compared to other identified methods. In addition to this, the experiment also suggested that the model was able to perform better in case of small datasets, data augmentation has a significant impact on the model and that controlling data augmentation could also reduce computational expense for the model.

[17] presented a method for identifying hand gestures using Gabor filters and Support Vector Machine (SVM) algorithm. Gabor filter - a well-known application for image processing was used by the authors for extracting features from an Image. The dimensionality of the extracted features was then further reduced using Principal Component Analysis (PCA) upon which the SVM (Support Vector Machine) algorithm was applied for identifying the hand gesture. The proposed method was tested on an augmented dataset and showcased an accuracy of 95%. [19] came with a new approach for identifying hand gestures using the ORAB method. Oriented FAST and rotated BRIEF also known as ORAB is a famous feature detector tool and is used for image processing. After performing segmentation on the images, the ORAB method was used for extracting features of the hand gesture. Post extracting the features, different machine learning algorithms were applied on them for identifying the hand gesture. To assess the robustness of the proposed approach, the algorithm was tested against different pre-processing techniques like PCA, LBP (Local Binary Pattern) and Histogram of Gradients. The evaluation results suggested that the proposed ORAB method outperformed the other mentioned techniques for KNN (K Nearest Neighbor), Naive Bayes and Logistic Regression algorithm, while the PCA outperformed all the other techniques for Random Forest, MLP (Multilayer Perceptron) and SVM classifier.

Since optimization of pre-existing model solutions could also contribute to novel models with faster execution and higher performance, investigating Harris Hawks and Crow Search algorithms on CNN models resulted in satisfying outcomes. In [2] hyperparameter tuning of CNN models for hand gesture recognition was done using a metaheuristic algorithm, Harris Hawks Optimization (HHO) algorithm. This optimized CNN image classification model experimentally proved to attain 100% accuracy with reduced training time in this study when compared to other metaheuristic algorithms with CNN model for evaluation. While in [7] an efficient CNN model was developed that was optimized by crow search method for hyperparameter tuning parameters such as convolution layers, number of epochs, batch size, loss function, activation function, number of dense layers and pooling layers, in order to provide 100 % accuracy. Even though optimizers like grid search, random search and bayesian approach are the go-to methods for hyperparameter tuning, methods inspired by nature, nevertheless lacking the explainability, perform better and faster by identifying optimal values for the associated parameters during model building. Crow search algorithm implementation on the proposed CNN model delivers faster convergence rate, higher efficiency, and with fewer control parameters for ease of complexity.

Several studies on surface electromyograph images (sEMG) for gesture recognition have also illuminated another approach. This paper [12] provides a comparative analysis of sEMG-based hand gesture recognition deep learning solution derived from a CNN-only NN (Neural Network), a combination of transfer learning a CNN-LSTM NN and Non-Transfer Learning techniques. A comprehensive input from 30 gestures with corresponding finger joint state, wrist state and elbow state were used to compare the model performances and computational expense. Transfer learning proved its competence in adaptation to new gestures, new users and lower data quality and quantity, and could still work with small amounts of the training data with faster execution time for generalized pattern recognition. On the other hand, an attention-based [11] hybrid CNN_RNN for EMG signal recognition performed better than the state of the art when compared for 200ms window length with NinaProDB1 and NinaProDB2 and 150ms for BioPatRec database. The attention-based learning mechanism [6] guides the model to learn by paying attention to different parts of the signal image sequentially to improve the recognition accuracy of the hybrid CNN-RNN framework. In contrast, [20] proposed a hand recognition method using the Electromyography (EMG) features of the muscles. Electromyography also commonly known as EMG is a method for extracting and evaluating the electrical activities of the skeletal muscles. The authors proposed that better results can be obtained if the EMG features are normalized to AUC-RMS values. To evaluate the proposed method, five different algorithms were developed on three different versions of the EMG feature. First - the original EMG feature set, second - EMG features normalized to maximum peak values and last - EMG feature set normalized to AUC-RMS value. The evaluation results showcased that the machine learning models in case of the proposed approach performed much better compared to other two versions of the feature set. Maximum accuracy on the proposed method was obtained by RF algorithm (96.38%).

Furthermore, evaluating transfer learning approaches in [1, 15] revealed their significance and limitations. [1] applied an end-to-end fine-tuned method on a pre-trained CNN model with score-level fusion technique to recognize hand gestures in the American sign language dataset with a low number of gesture images and the results were evaluated using Leave one out-CV and CV test. Due to the limitation of available data and training capacity for CNN models from scratch, authors used pre-trained models such as AlexNet, VGG, GoogleNet and ResNet and performed fine-tuning on the target datasets. This research due to its relatively smaller input size for fine-tuning made use of the Transfer Learning from the pre-trained CNN models, AlexNet and VGG-16 models and calculated the score-level fusion for the recognition of static hand gestures. The technique implemented the depth thresholding technique for hand region segmentation for noise reduction and also eliminated the requirement for illumination variation, rotation and hand region segmentation in some cases. The technique proved to be superior to the existing methods and was able to distinguish even closely related gestures postures with greater accuracy. In [15], Transfer learning from the Inception V3 pre-trained CNN model was used for web-cam image classification on MNIST Sign Language dataset with 3000 images for gesture of each letter. Data augmentation was also performed by random rotation about 20 degrees, height and width shift, zoom, brightness and horizontal flip. The dataset was subset as two and experimented for comparison post input video-frames breakdown and pre-processing for gesture recognition and then converted into text/ audio output. Inception V3 is better than base models but plateaus after a certain level (around 70%) and doesn’t change with optimization.

[21] presented a comparative study of the performance of random forest algorithm on two different approaches for identifying and classifying gestures. The first method performs the segmentation and classification task simultaneously while the second method performs these tasks sequentially. The first approach utilizes one single random forest algorithm for recognizing the gestures from a video of 3D body joints. The second approach uses two random forest models for identifying the gestures: one binary random forest for differentiating the gestures and a multi-class random forest model for segmenting the gestures. After evaluating the results, it found that the first approach of simultaneous segmentation and classification yielded better results.

In the paper, [16] studied the vision-based hand gesture recognition from a real-time video using the Viola-jones method in which the hand is located using a blue rectangle by the system and then processed for gesture pattern recognition using SVM classifier based on Hu invariant moment vectors.

Another vision-based gesture recognition based on CNN-LSTM model [8] for Asian culture gestures worked with both augmented and un-augmented datasets to produce recognition of both single and double hand gestures with an accuracy of upto 99% better than Transferred CNN models. Although performance was good, diversity in features such as skin tone, light intensity and background noise in the training set could improve the reliability of the results. Whereas [10] focused on gesture recognition using temporal pattern extraction using the LSTM and then feature reduction through CNN for the sEMG signal images. The proposed model LCNN constitutes 2 LSTM layers with 52 cells (64 hidden layers each) each, 1 1-D convolution layers and 1 output layer with 0.3 dropout and Batch normalization for regularization due to smaller data volume of the DB5 dataset and self-collected Myo Dataset. Feature design complexity is reduced intensely due to the sequence of models applied i.e., first LSTM then CNN.

Finally, [24] employed a CNN-LSTM neural network to diagnose Covid-19 from CT-scan X-ray images as an alternative to existing yet scarce and slower RT-PCR tests. The model achieved an accuracy of 99.4% on a 4575 X-ray image dataset including 1525 Covid-19 infected patient chest X-ray images.

This use case was also handled in the past using transfer learning for covid-19 classification along with ML techniques such as SVN, CNN and random forest. The proposed system was faster (18372 s and 113 s for training and testing) and also performed better in accuracy – 99.4 and 99.2 for COVID class. Even though the higher performance, proposed model was not generalized enough as the dataset was limited in size, existing disease could still affect the model accuracy. The image diversity considering all views of the X-rays and validation from radiologists could improve the model.

A comprehensive analysis of the related work revealed that Convolutional Neural Networks although add to model design complexity, they help immensely and effectively in feature reduction ultimately contributing to greater model performance. Moreover, when employed with LSTM or GRU network it can lead to promising results and simpler learning curve while maintaining lower execution time on an enriched dataset with large enough sample size for each gesture class.

III. Methodology

It is essential for a Data Mining And Machine Learning project to follow a certain structured framework to achieve the project objectives. Methodologies make sure that no crucial steps are overlooked during the project development phase.

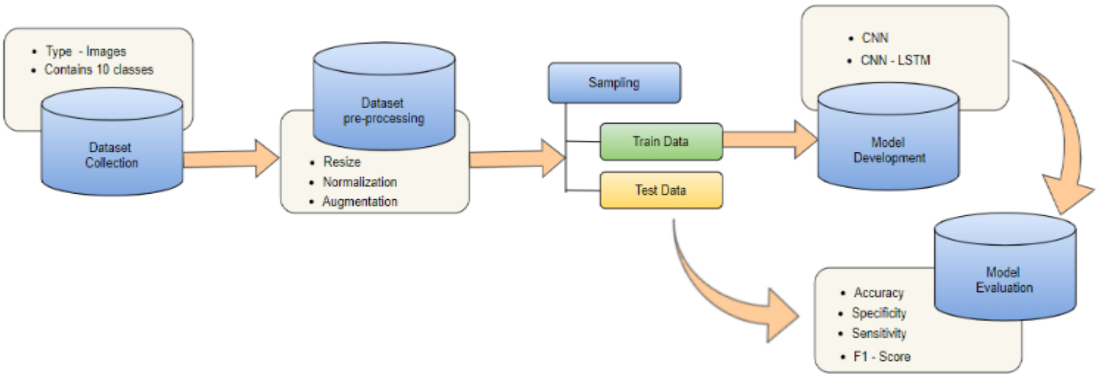

Figure 1: Project Methodology

Knowledge Discovery in Databases (KDD) is a methodology which is widely used for discovering knowledge and insights from a Data Mining project. Apart from being cost effective and simple, KDD provides a set of iterative steps which can assist in enhancing and refining the project execution and provide more appropriate insights. Thus, for achieving the project objectives and finding an answer to the question, a simple methodology was tailored using the KDD framework. Fig. 1 shows the components and flow of the project methodology.

A. *Data Selection*

To achieve the project objectives, a Hand Gesture Recognition Dataset from the Kaggle Data repository was selected. This dataset encapsulates a set of near infrared images containing 10 different hand gestures namely - palm (suggesting to stop), L shape, fist, fist moved, thumb, index, okay, palm moved, C shape and down. Overall the dataset contains 20000 images, each with a resolution of 640 x240.

B. *Data Preparation*

Resizing and normalizing images is an essential pre-processing measure in Image classification [23]. Deep learning models perform faster on images with lower resolution. As the resolution of an input image increases, the learning time of the neural network architecture also increases [23]. Thus for reducing the training time and saving computational expenses, all the images from the dataset which initially had a resolution of 640x240 were then resized and normalized to 150x150.

In addition to this, Image Augmentation was carried out to improve variation in the datasets. Image augmentation only enriches the dataset, but also assists a model in achieving generalization and avoid overfitting. For this project, augmentation methods like - random rotation, height and width shift, increasing the brightness and horizontal flip were conducted on the dataset to carry out Image Augmentation.

These datasets were then used for developing the deep learning models and were divided in the ratio of 75:25 for training and validation purposes.

C. *Data Modelling*

After preparing the datasets, six different deep learning models were developed using CNN and RNN algorithms.

The primary reason for choosing CNN algorithm is because of its ability to extract features from an image. For any given image, the algorithm has the ability to learn distinctive features on its own all the while reducing the feature size. Furthermore, CNN models are very powerful and efficient in terms of computation.

The RNN (Recurrent Neural Network) algorithm, on the other hand, has its own set of benefits. RNN models are very memory efficient, even for the inputs with larger size, the size of the RNN models does not increase. For this particular reason, two specialized types of RNN algorithms have been explored - LSTM and GRU. Another advantage of RNN is the ability to use previous knowledge from earlier predictions for making further new predictions. Also, RNN can be connected with CNNs to extend the pixel neighbourhood.

Hence, CNN and special RNN algorithms were used for developing the models. For this project, CNN was used for extracting features from the images and RNN was used for making the predictions. Six models were developed using the CNN and RNN algorithm are mentioned in the Table No. 1.

Sr. No.	**Model**	**Data Augmentation**	**Total Parameters**
1.	CNN2D	No	4,144,714
2.	CNN2D	Yes	4,144,714
3.	CNN2D-LSTM	No	1,002,442
4.	CNN2D-LSTM	Yes	1,002,442
5.	CNN2D-GRU	No	608,202
6.	CNN2D-GRU	Yes	608,202

Table No. 1: Project Methodology

The execution time of the model can be observed in the Fig. 2

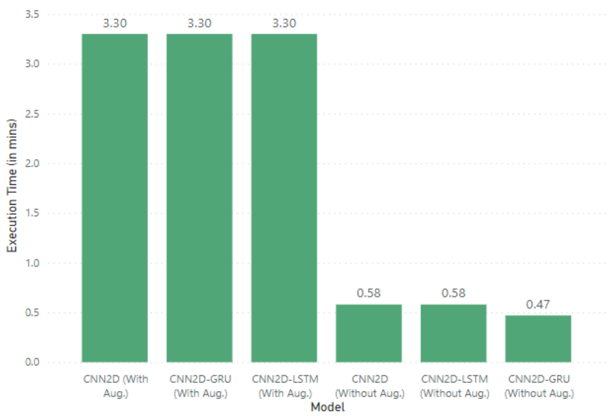

Figure 2: Execution Time (in minutes)

D. Model Evaluation

To quantify the quality of the models, each model was evaluated using Confusion Matrix and then Accuracy, Precision, Recall and F1 Score of each of the models was calculated. Fig. 2 showcases the confusion matrix for each of the models.

From the Confusion Matrix, Accuracy, Precision, Recall and F1 Score of the models was calculated. Table No. 2 encapsulates the figures for each of the mentioned measure.

Performance of the models during each epoch can be seen in Fig. 3 and Fig. 4.

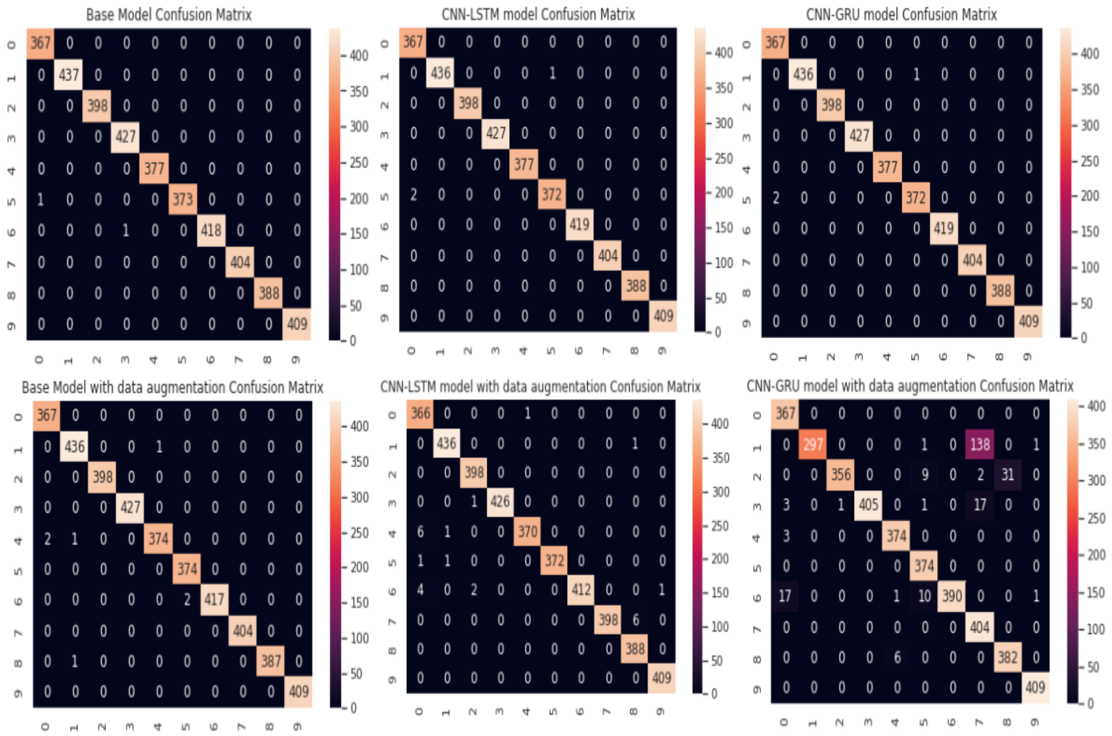

Figure 2: Confusion Matrix

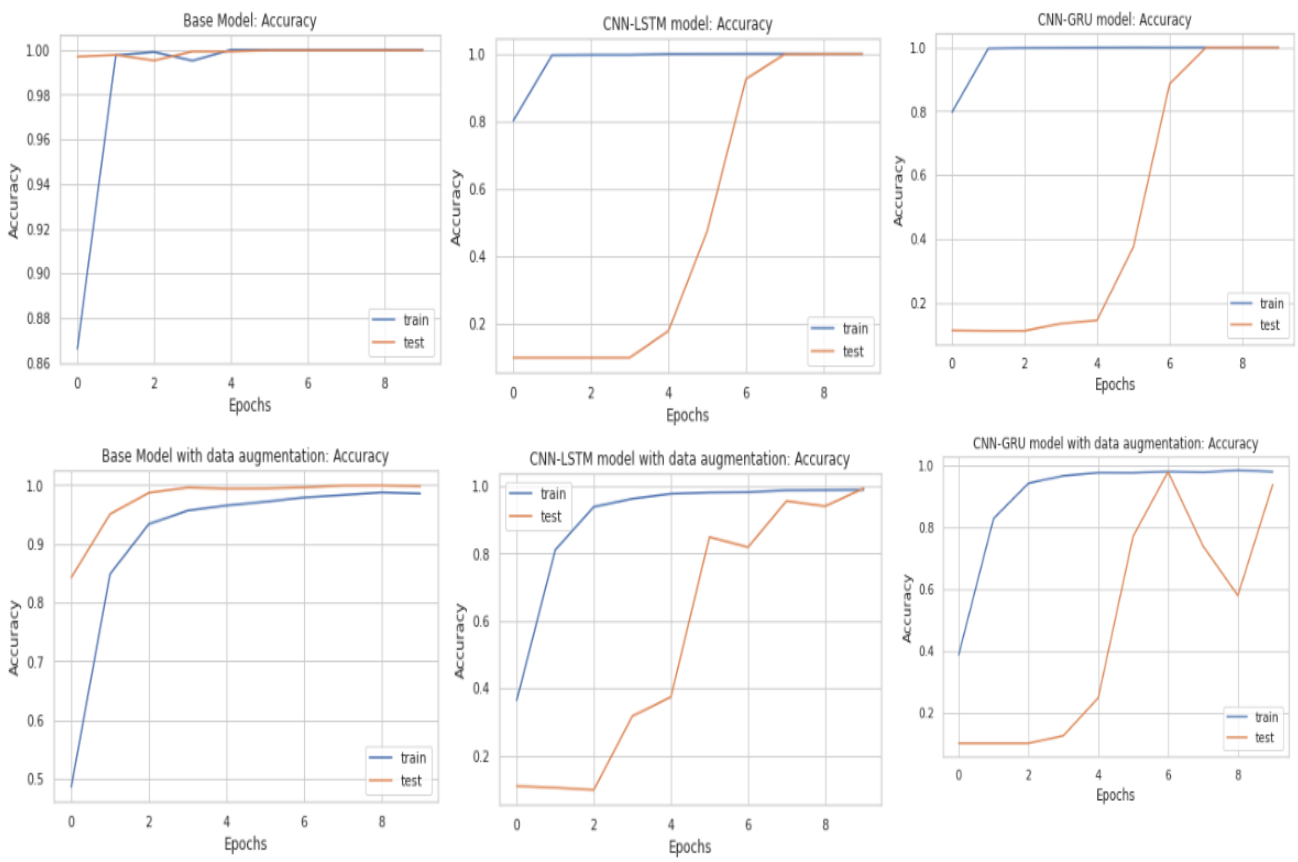

Figure 3: Accuracy of the models per epochs

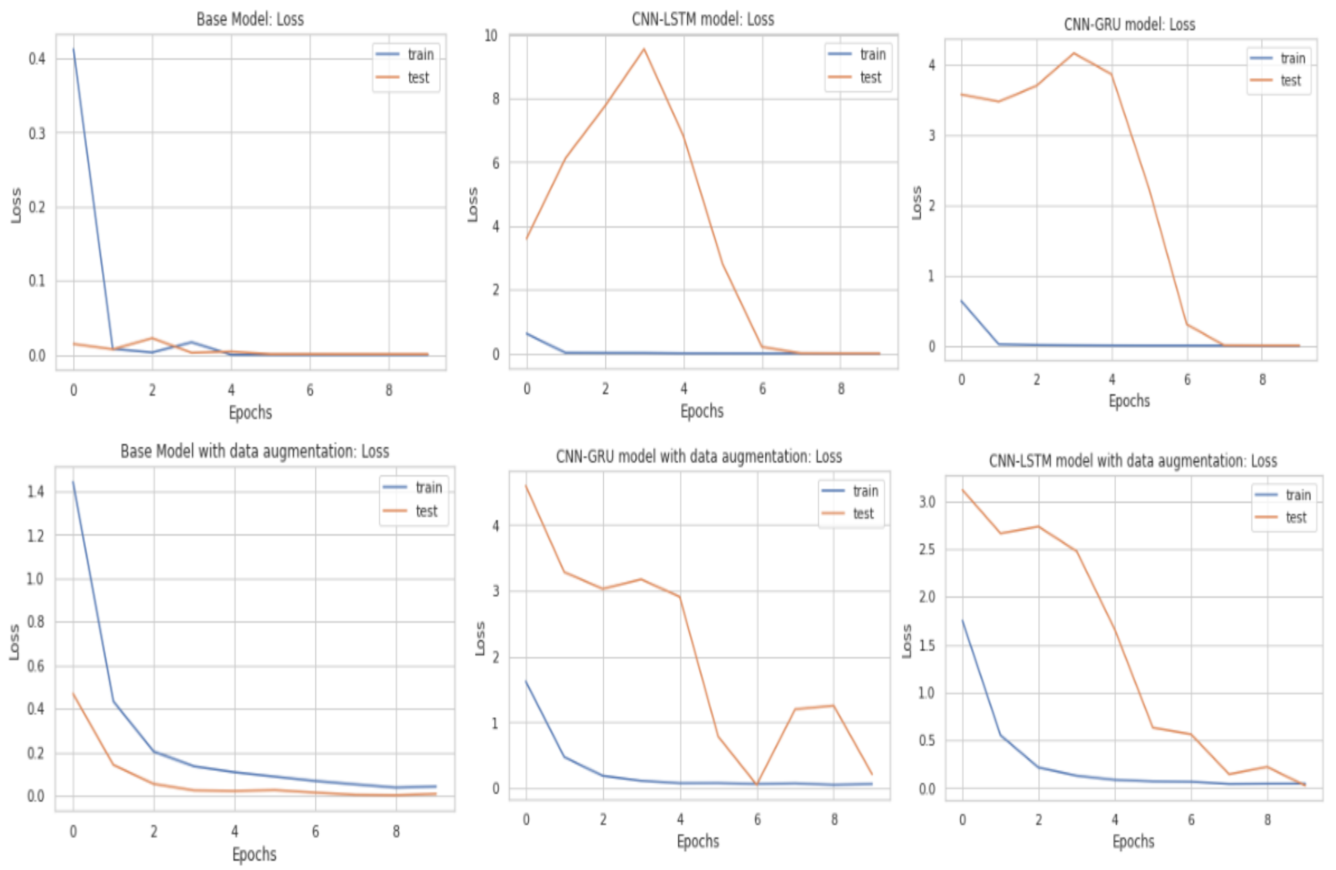

Figure 4: Loss per epochs

Fig. 3 showcases the accuracy of the models per epochs. From the Fig. 3 it can be seen that in case of CNN-LSTM and CNN-GRU model, the test accuracy showcased an irregular increase compared to the CNN2D (base) model test accuracy. Also, these two models achieved their top accuracies over around 6 epochs, while the CNN2D (base) model achieved it as faster as in one epoch.

Conversely, the loss graphs of the models follow a similar pattern in reverse direction. Fig. 4 showcases the loss graphs of the models.

Model	Data Augmentation	Test Accuracy	Precision	Recall	F1 Score
CNN2D	No	0.9994	0.9997	0.9995	0.9996
CNN2D	Yes	0.99825	0.9984	0.9975	0.9979
CNN2D-LSTM	No	0.9992	0.99925	0.99925	0.99925
CNN2D-LSTM	Yes	0.9937	0.99375	0.99375	0.99375
CNN2D-GRU	No	0.9992	0.9994	0.99925	0.9993
CNN2D-GRU	Yes	0.9394	0.9448	0.93375	0.9392

Table No. 2: Evaluation Results of the models

. From the Fig.4, it can be observed that even though the CNN2D (base) model achieved the top accuracy in less epochs, it may lead to over-fitting. Hence, the other models may give an advantage to avoid the problem of over-fitting.

To assess the accuracy of the models further, the accuracy of each model was drilled down into three components: test accuracy and training accuracy.

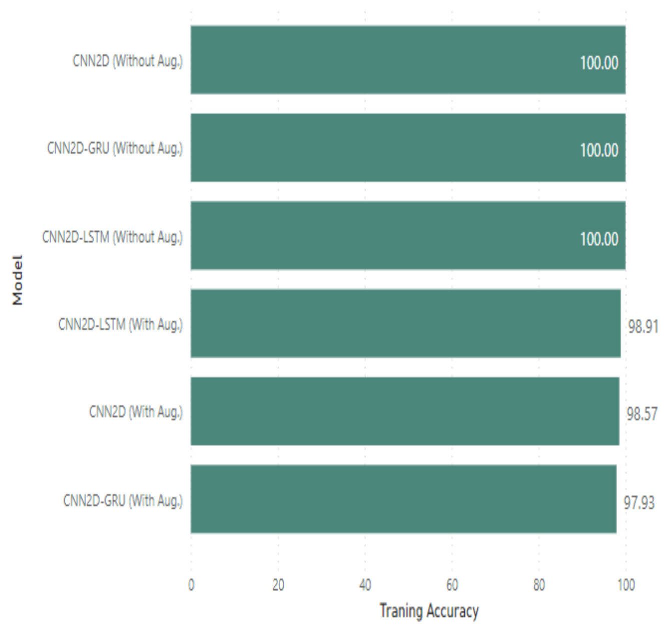

Figure 5: Training Accuracy of the models

Fig. 5 shows the training accuracy of each model. 100 % training accuracy results can be observed in case of the models which were trained on a plain dataset without the augmentation. Such a high accuracy suggests the problem of overfitting in these models.

Fig 6 and Fig 7 showcase the test and validation accuracies of the models respectively. All the models show high test and validation accuracies.

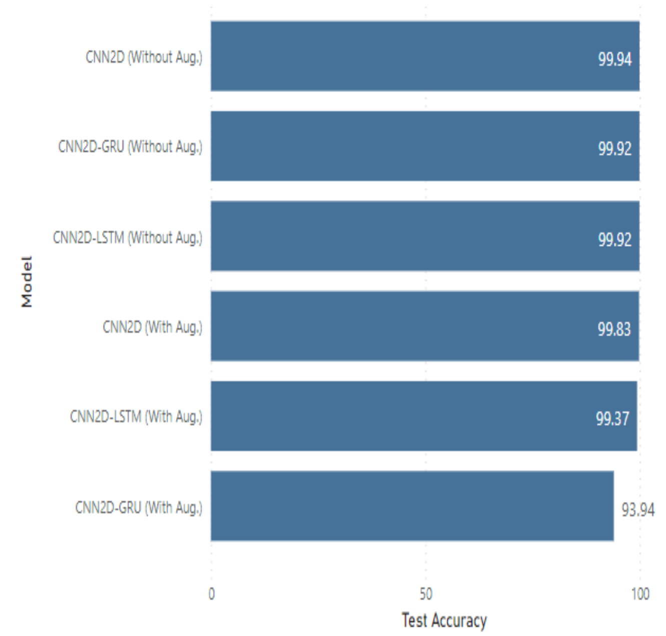

Figure 6: Test Accuracy of the models

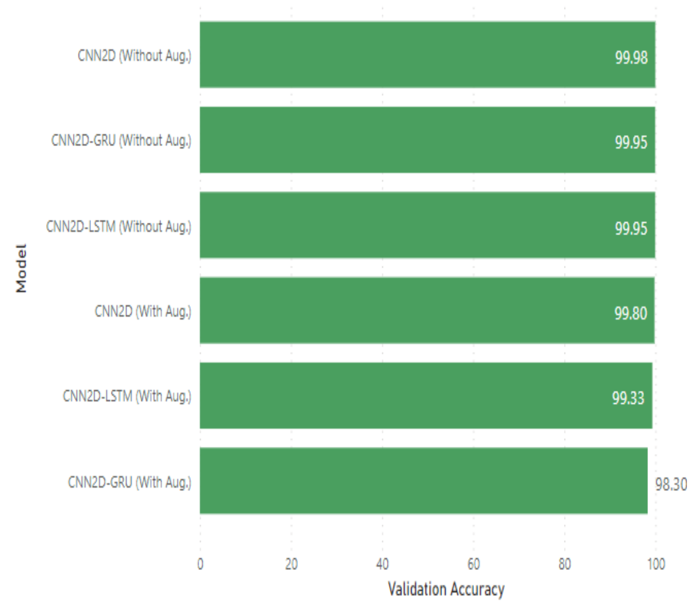

Figure 7: Validation Accuracy of the models

IV. Results and Conclusions

Evaluating multiple cutting-edge deep learning image recognition models such as CNN, CNN-LSTM and CNN-GRU, with and without augmented data provided an in-depth insight into their workings on a public dataset available on Kaggle. We observed the highest accuracy and a comparable speed for execution for the CNN-LSTM model which suits our requirement due to the availability of an enriched dataset. When dealing with a lower quality dataset, we can use the CNN-GRU for similar results with higher speed.

Even though both the models, CNN-LSTM and CNN-GRU produced similar results when it came to accuracy, partly due to the enriched nature of the dataset in which each gesture class is represented with almost 200 samples without data augmentation, CNN-LSTM can be selected for this particular use case over CNN-GRU because of its inherent nature of higher accuracy when a diverse dataset is available. However, for a dataset with lesser diversity and sample size, a CNN-GRU approach can perform better.

V. Future Work

Although the analysis and experiments were thorough, further comparisons with customization of Multilayer Perceptron, Convolution Neural Network, Graph Neural Networks could provide more insights into the performance and generalization of the CNN-LSTM and CNN-GRU models. This research could also be extended by upgrading the model to incorporate video inputs along with coloured (RGB) input features.

Reference

[1] J. P. Sahoo, A. J. Prakash, P. Pławiak, and S. Samantray, “Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network,” Sensors, vol. 22, no. 3, p. 706, 2022.

[2] T. R. Gadekallu et al., “Hand gesture recognition based on a Harris hawks optimized convolution neural network,” Computers and Electrical Engineering, vol. 100, p. 107836, 2022.

[3] Y. S. Tan, K. M. Lim, and C. P. Lee, “Hand gesture recognition via enhanced densely connected convolutional neural network,” Expert Systems with Applications, vol. 175, p. 114797, 2021.

[4] A. A. Alani, G. Cosma, A. Taherkhani, and T. McGinnity, “Hand gesture recognition using an adapted convolutional neural network with data augmentation,” in 2018 4th International conference on information management (ICIM), 2018, pp. 5–12.

[5] G. Benitez-Garcia et al., “Improving real-time hand gesture recognition with semantic segmentation,” Sensors, vol. 21, no. 2, p. 356, 2021.

[6] C. Li, S. Li, Y. Gao, X. Zhang, and W. Li, “A two-stream neural network for pose-based hand gesture recognition,” IEEE Transactions on Cognitive and Developmental Systems, 2021.

[7] T. R. Gadekallu et al., “Hand gesture classification using a novel CNN-crow search algorithm,” Complex & Intelligent Systems, vol. 7, no. 4, pp. 1855–1868, 2021.

[8] N. Basnin, L. Nahar, and M. S. Hossain, “An integrated CNN-LSTM model for micro hand gesture recognition,” in International Conference on Intelligent Computing & Optimization, 2020, pp. 379–392.

[9] K. Lai and S. N. Yanushkevich, “CNN+ RNN depth and skeleton based dynamic hand gesture recognition,” in 2018 24th international conference on pattern recognition (ICPR), 2018, pp. 3451–3456.

[10] Y. Wu, B. Zheng, and Y. Zhao, “Dynamic gesture recognition based on LSTM-CNN,” in 2018 Chinese Automation Congress (CAC), 2018, pp. 2446–2450.

[11] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, “A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition,” PloS one, vol. 13, no. 10, p. e0206049, 2018.

[12] X. Chen, Y. Li, R. Hu, X. Zhang, and X. Chen, “Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method,” IEEE Journal of

Biomedical and Health Informatics, vol. 25, no. 4, pp. 1292–1304, 2020.

[13] J.-W. Choi, S.-J. Ryu, and J.-H. Kim, “Short-range radar based real-time hand gesture recognition using LSTM encoder,” IEEE Access, vol. 7, pp. 33610–33618, 2019.

[14] J. C. Nunez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Velez, “Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition,” Pattern Recognition, vol. 76, pp. 80–94, 2018.

[15] M. Agrawal, R. Ainapure, S. Agrawal, S. Bhosale, and S. Desai, “Models for hand gesture recognition using deep learning,” in 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 2020, pp. 589–594.

[16] L. Yun and Z. Peng, “An automatic hand gesture recognition system based on Viola-Jones method and SVMs,” in 2009 Second international workshop on computer science and engineering, 2009, vol. 2, pp. 72–76.

[17] D.-Y. Huang, W.-C. Hu, and S.-H. Chang, “Vision-based hand gesture recognition using PCA+ Gabor filters and SVM,” in 2009 fifth international conference on intelligent information hiding and multimedia signal processing, 2009, pp. 1–4.

[18] S. Kajan, J. Goga, and O. Zsíros, “Comparison of algorithms for dynamic hand gesture recognition,” in 2020 Cybernetics & Informatics (K&I), 2020, pp. 1–5.

[19] A. Sharma, A. Mittal, S. Singh, and V. Awatramani, “Hand gesture recognition using image processing and feature extraction techniques,” Procedia Computer Science, vol. 173, pp. 181–190, 2020.

[20] M. F. Wahid, R. Tafreshi, M. Al-Sowaidi, and R. Langari, “Subject-independent hand gesture recognition using normalization and machine learning algorithms,” Journal of computational science, vol. 27, pp. 69–76, 2018.

[21] A. Joshi, C. Monnier, M. Betke, and S. Sclaroff, “Comparing random forest approaches to segmenting and classifying gestures,” Image and Vision Computing, vol. 58, pp. 86–95, 2017.

[22] L. Kane, B. K. Dewangan, and T. Choudhury, “Challenges and Applications for Hand Gesture Recognition,” 2022.

[23] S. Saponara and A. Elhanashi, “Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance,” in International Conference on Applications in Electronics Pervading Industry, Environment and Society, 2022, pp. 10–17.

[24] M. Z. Islam, M. M. Islam, και A. Asraf, ‘A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images’, Informatics in Medicine Unlocked, τ. 20, σ. 100412, 2020.

[25] M. Panwar and P. S. Mehra, “Hand gesture recognition for human computer interaction,” in 2011 International Conference on Image Information Processing, 2011, pp. 1–7.

[26] T. Mantecón, C.R. del Blanco, F. Jaureguizar, N. García, “Hand Gesture Recognition using Infrared Imagery Provided by Leap Motion Controller”, Int. Conf. on Advanced Concepts for Intelligent Vision Systems, ACIVS 2016, Lecce, Italy, pp. 47-57, 24-27 Oct. 2016. (doi: 10.1007/978-3-319-48680-2_5)