Gesture Recognition Project:

Contributors: Ravikiran Krishnaprasad and Amit Kumar

Throughout the gesture recognition project, various model architectures and training strategies were experimented with to optimize the accuracy and performance of the model. Below is a detailed explanation of each experiment conducted, including the results and decisions made based on those results

Experiment Number	Model	Result	Decision + Explanation
1	Baseline Conv3D Model	• **Training Accuracy**: 80.80% • **Validation Accuracy**: 27.34% • **Validation Loss**: 10.7707	The baseline Conv3D model showed a significant gap between training and validation accuracy, indicating overfitting. The high validation loss further confirms that the model struggled to generalize to unseen data. To address this, subsequent experiments focused on tuning hyperparameters and exploring different architectures.
2	**Conv3D:** Increasing Number of Epochs to 20	• **Training Accuracy**: 93.90% • **Validation Accuracy**: 71.09% • **Validation Loss**: 1.6455	Increasing the number of epochs from 10 to 20 significantly improved both training and validation accuracy. This suggested that the model benefited from more extended training, allowing it to learn more complex features. However, the validation loss, while reduced, indicated that there was still room for improvement in generalization.
3	**Conv3D:** Decreasing Batch Size to 16	• **Training Accuracy**: 97.92% • **Validation Accuracy**: 80.36%	Decreasing the batch size to 16 resulted in further improvements in both training and validation accuracy. The smaller batch size allowed for more frequent updates to the model's weights, which likely

		• **Validation Loss**: 1.3362	helped the model generalize better to the validation set. This experiment demonstrated that tuning the batch size could positively impact model performance.
4	**ConvLSTM**	• **Training Accuracy**: 49.78% • **Validation Accuracy**: 60.27% • **Validation Loss**: 3.5580	The ConvLSTM model was introduced to capture both spatial and temporal dependencies. However, it underperformed compared to the Conv3D model, with lower training and validation accuracy and higher validation loss. This result suggested that the ConvLSTM architecture, as implemented, might require further tuning or modifications to be effective in this context.
5	**ConvLSTM: Increasing no of epochs**	Experiment could not be completed due to disk errors.	While the experiment was intended to explore whether increasing the number of epochs would improve the ConvLSTM model's performance, technical issues prevented its completion.
6	CNN + RNN Model **(Final Model)**	• **Training Accuracy**: 100.0% • **Validation Accuracy**: 81.70% • **Validation Loss**: 0.7183	The CNN + RNN model achieved the highest validation accuracy among all experiments, with a perfect training accuracy. This indicates that the model was able to effectively capture both spatial and temporal features. The relatively low validation loss further suggests that this architecture generalizes well to unseen data. Given these results, the CNN + RNN model is the most promising architecture and could be considered the final model for this task.