### **Results Discussion and Reflections**

Among the three models I explored, the **basic deep learning model with Conv1D layers** performed the best overall in terms of accuracy. I initially expected the transfer learning models, especially the ones using GloVe embeddings and BERT, to outperform the simpler Conv1D model, but they didn’t. This result was both surprising and frustrating, as I invested a lot of time tuning hyperparameters and refining the models, only to see the validation accuracy barely improve.

#### **1. Basic Deep Learning Model with Conv1D Layers**

* **Why It Worked Well**: The Conv1D layers excel at picking up local patterns in text, such as n-grams or phrases, which likely gave it an edge over the other models. Since the dataset includes relatively short reviews, Conv1D might have been able to extract the most relevant features without overcomplicating the process.
* **Challenges**: Despite its relatively good performance, the validation accuracy was still lower than I’d hoped. I think this may be because Conv1D struggles with capturing global context or deeper semantic meaning in text. Also, preprocessing inconsistencies (like handling rare words or punctuation) could have impacted its ability to generalize.

#### **2. Transfer Learning Model with GloVe Embeddings**

* **Why It Should Have Done Better**: Pre-trained embeddings like GloVe are supposed to provide a strong foundation by encoding semantic relationships between words. I expected this to significantly boost the model’s ability to classify sentiments accurately.
* **Why It Didn’t**: One big limitation was that I froze the embeddings during training, which meant they couldn’t adapt to the specific dataset. I now realize that tuning the embeddings or even partially unfreezing them might have made a difference. Additionally, GloVe embeddings are static and don’t capture the context-dependent nature of words, which could have limited their effectiveness here. I also experimented with adjusting dropout rates and learning rates, but these only led to minor changes in accuracy.

#### **3. Transfer Learning Model with BERT**

* **Why I Had High Hopes**: BERT is a state-of-the-art model for many NLP tasks, and its ability to capture contextual relationships made me optimistic about its performance. It’s particularly good at understanding the nuances in longer and more complex texts.
* **Why It Fell Short**: In hindsight, I think there were a few issues here. First, BERT requires extensive fine-tuning to work well, and while I tried different configurations for hyperparameters like learning rate and batch size, the improvements were marginal. Second, the dataset’s relatively short and simple reviews didn’t give BERT much room to showcase its strengths in capturing intricate relationships. Lastly, the tokenization and preprocessing steps might not have been optimal, which could have reduced the quality of input to the model.

### **Reflections on Challenges**

One of the most significant challenges was that, no matter how much I adjusted the models, the validation accuracy plateaued early on. This was frustrating because I felt like I was putting in a lot of effort but not seeing the results I wanted. Here are a few reasons I think this happened:

1. **Dataset Limitations**: The dataset may not have been diverse or large enough for the models to fully generalize. If there were noisy or ambiguous samples, this could have confused the models and led to lower accuracy.
2. **Hyperparameter Sensitivity**: I spent a lot of time tuning hyperparameters like the number of layers, dropout rates, and learning rates. While I saw some slight improvements, they weren’t as impactful as I’d hoped, suggesting that the models were hitting their performance ceiling with this dataset.
3. **Mismatch Between Models and Data**: While GloVe and BERT are powerful, they’re designed for tasks that benefit from large datasets and complex context. This dataset, being relatively simple and short, might not have allowed them to shine.

### **Final Thoughts**

Although the results were not as strong as I had hoped, this project taught me a lot about the complexities of working with machine learning models, especially in NLP. I now have a deeper appreciation for how much data quality, preprocessing, and hyperparameter tuning matter. Even though I only saw slight improvements, I feel like the time I spent on this was valuable for understanding where models succeed and fail. Going forward, I think exploring larger datasets or refining preprocessing techniques could help these models achieve better results. This experience also taught me that sometimes, simpler models like Conv1D can outperform more complex ones when the dataset aligns better with their strengths.