| **Model** | **Accuracy** | **Macro Precision** | **Macro Recall** | **Macro F1** |
| --- | --- | --- | --- | --- |
| **RNN** | 0.67 | 0.68 | 0.67 | 0.67 |
| **LSTM** | **0.70** | **0.71** | **0.70** | **0.71** |
| **BiLSTM** | 0.68 | 0.68 | **0.69** | 0.68 |
| **GRU** | 0.66 | **0.77** | 0.63 | 0.65 |
| **BiGRU** | 0.64 | 0.67 | 0.67 | 0.65 |

Metric Summary Table

Accuracy Comparison

| RNN | ██████████████████████░░░ 67%

| LSTM | █████████████████████████ 70%

| BiLSTM | ███████████████████████░░ 68%

| GRU | █████████████████████░░░░ 66%

| BiGRU | ███████████████████░░░░░░ 64%

F1-Score Breakdown by Class

| **Model** | **Negative F1** | **Neutral F1** | **Positive F1** |
| --- | --- | --- | --- |
| RNN | 0.65 | 0.64 | 0.73 |
| LSTM | 0.68 | 0.68 | 0.76 |
| BiLSTM | 0.68 | 0.63 | 0.74 |
| GRU | 0.57 | 0.69 | 0.68 |
| BiGRU | 0.65 | 0.55 | 0.74 |

Model Behaviour Analysis

* **RNN (Baseline)**
* Strengths: Simpler model; faster training.
* Weaknesses: Lower accuracy and macro F1; struggles with long-term dependencies.
* Conclusion: Performs decently but not recommended for production.
* **LSTM**
* Strengths: Best accuracy (70%), well-balanced precision, recall, and F1 across all classes.
* Weaknesses: Slightly heavier than RNN.
* Conclusion: Strong candidate for deployment.
* **BiLSTM**
* Strengths: Leverages context from both directions; high recall.
* Weaknesses: Marginal gain over LSTM; more computational overhead.
* Conclusion: Good alternative to LSTM but not significantly better.
* **GRU**
* Strengths: Lightweight alternative to LSTM; higher precision for some classes.
* Weaknesses: Lower accuracy and inconsistent recall (biased towards neutral class).
* Conclusion: Not stable across all classes; not ideal for deployment.
* **BiGRU**
* Strengths: Captures bidirectional dependencies.
* Weaknesses: Lower accuracy, suffers in neutral classification.
* Conclusion: Underperforms; not suitable for production.

Validation Against Expected Outcomes

| **Expected vs Observed** | **✅ Match** |
| --- | --- |
| **BERT > LSTM/GRU > RNN** (Expected) | ✅ Partially validated (LSTM > RNN) |
| LSTM/GRU better than RNN | ✅ Yes |
| RNN weakest | ✅ Yes |