***Classification of Events***

True Positives (TP) - Correct Predictions by LLM

These are events that the LLM correctly predicted and also happened during Operation Cast Lead.

| **Event** | **Definition** |
| --- | --- |
| The operation was initiated by Israel | The Israeli Defense Forces (IDF) launched the operation in response to Hamas rocket attacks on Israeli cities. |
| The conflict involved heavy aerial bombardments | The IDF conducted extensive airstrikes against Hamas military targets, including infrastructure and leadership. |
| A ground invasion followed the air campaign | After initial airstrikes, the IDF launched a ground assault into Gaza to dismantle Hamas' capabilities. |
| Urban warfare was a major feature of the conflict | Combat occurred in densely populated areas, increasing the difficulty of military operations. |
| Hamas used underground tunnel networks | Hamas utilized an extensive tunnel system for smuggling, ambushes, and movement. |
| Rocket attacks from Gaza targeted Israeli cities | Hamas fired thousands of rockets at Israeli population centers, prompting the IDF response. |
| Israel enforced a blockade on Gaza | Israel restricted the movement of goods and people to limit Hamas’ access to weapons and supplies. |
| The conflict resulted in significant civilian casualties | Both Israeli and Palestinian civilian casualties were reported, drawing international scrutiny. |
| International criticism and calls for ceasefire increased as the conflict continued | The operation drew widespread condemnation from human rights groups and governments. |
| The conflict ended with a unilateral Israeli ceasefire | The IDF declared a ceasefire after achieving its military objectives, though Hamas claimed victory. |

Total True Positives (TP): **10**

**False Positives (FP) - Incorrect Predictions by LLM**

These are events that the LLM predicted but did NOT actually happen in Operation Cast Lead.

| **Event** | **Why It Did NOT Happen?** |
| --- | --- |
| The conflict resulted in a negotiated peace settlement | No formal peace agreement was reached; Israel unilaterally ceased hostilities, and Hamas continued governing Gaza. |
| A major prisoner exchange occurred as part of the resolution | While prisoner exchanges occurred in later conflicts, this battle did not include one as part of its conclusion. |
| The IDF suffered heavy casualties due to Hamas counterattacks | Israeli losses were relatively low compared to Hamas casualties. |
| Hamas gained territorial control through the conflict | Hamas remained in control of Gaza, but its military infrastructure was severely damaged. |
| Israel aimed for permanent occupation of Gaza | Israel did not seek long-term occupation; it withdrew after the operation. |
| Large-scale naval battles were a key part of the operation | While Israel maintained a naval blockade, no major maritime battles occurred. |
| The war led to immediate political changes in either Israel or Gaza | Neither the Israeli government nor Hamas experienced leadership changes directly due to this battle. |

Total False Positives (FP): **7**

**False Negatives (FN) - Missed Predictions by LLM**

These are real events that happened during Operation Cast Lead but were NOT predicted by the LLM.

| **Event** | **Definition** |
| --- | --- |
| The operation was triggered by the breakdown of a ceasefire between Israel and Hamas | A previous truce between Hamas and Israel collapsed, leading to the start of the conflict. |
| Media and information warfare played a significant role | Both sides engaged in propaganda and used media to shape international perceptions. |
| Israel was accused of using white phosphorus in urban areas | The use of white phosphorus munitions by Israel became a major point of controversy. |
| Hamas attempted to smuggle weapons through tunnels during the battle | Tunnels were not only for movement but also for arms smuggling. |
| Cyber warfare and intelligence played a role in the conflict | Israel conducted cyber operations to disrupt Hamas communications. |
| Hamas' rocket capabilities had improved significantly | Hamas used longer-range rockets capable of reaching deeper into Israel. |
| Hamas executed suspected collaborators during the conflict | Reports indicated that Hamas targeted and killed individuals suspected of working with Israel. |
| The conflict had long-term effects on the Israel-Gaza dynamic | The war did not resolve tensions but shaped future conflicts and military strategies. |
| International legal investigations and accusations followed the war | The UN's Goldstone Report accused Israel of war crimes, adding to diplomatic pressure. |

Total False Negatives (FN): **9**

***Final Results***

| **Metric** | **Value** |
| --- | --- |
| Total True Positives (TP) | **10** |
| Total False Positives (FP) | **7** |
| Total False Negatives (FN) | **9** |
| Precision | **58.82%** |
| Recall | **52.63%** |
| F1-Score | **55.48%** |

**Critical Analysis of LLM’s Prediction**

1. Major Issue: The LLM Predicted the Wrong Winner

* The LLM incorrectly assumed Hamas won, when in reality, Israel achieved its operational objectives and declared victory.
* Hamas suffered major losses, but continued governing Gaza and claimed symbolic victory.

2. Strengths in LLM’s Predictions

Correctly identified major aspects of the battle:

* Israeli airstrikes and ground invasion were central to the conflict.
* Hamas’ use of tunnels and rocket attacks played a significant role.
* International legal scrutiny and calls for ceasefire shaped the war’s political dimension.
* The war ended with a unilateral Israeli ceasefire rather than a negotiated settlement.

3. Weaknesses in LLM’s Predictions

Missed key factors that influenced the war’s impact:

* Did not mention the ceasefire breakdown that triggered the war.
* Ignored the role of cyber warfare and intelligence in Israel’s strategy.
* Overlooked how propaganda and information control shaped international perceptions.

Invented historical inaccuracies:

* There was no negotiated peace settlement.
* Naval battles did not play a major role.
* Israel did not suffer major military losses.

***Final Verdict***

* The LLM’s prediction was weak (~55.48% F1-score) due to incorrect outcome prediction and missing key historical elements.
* Improving recall (identifying more real-world details) and correcting historical inaccuracies would significantly improve future assessments.