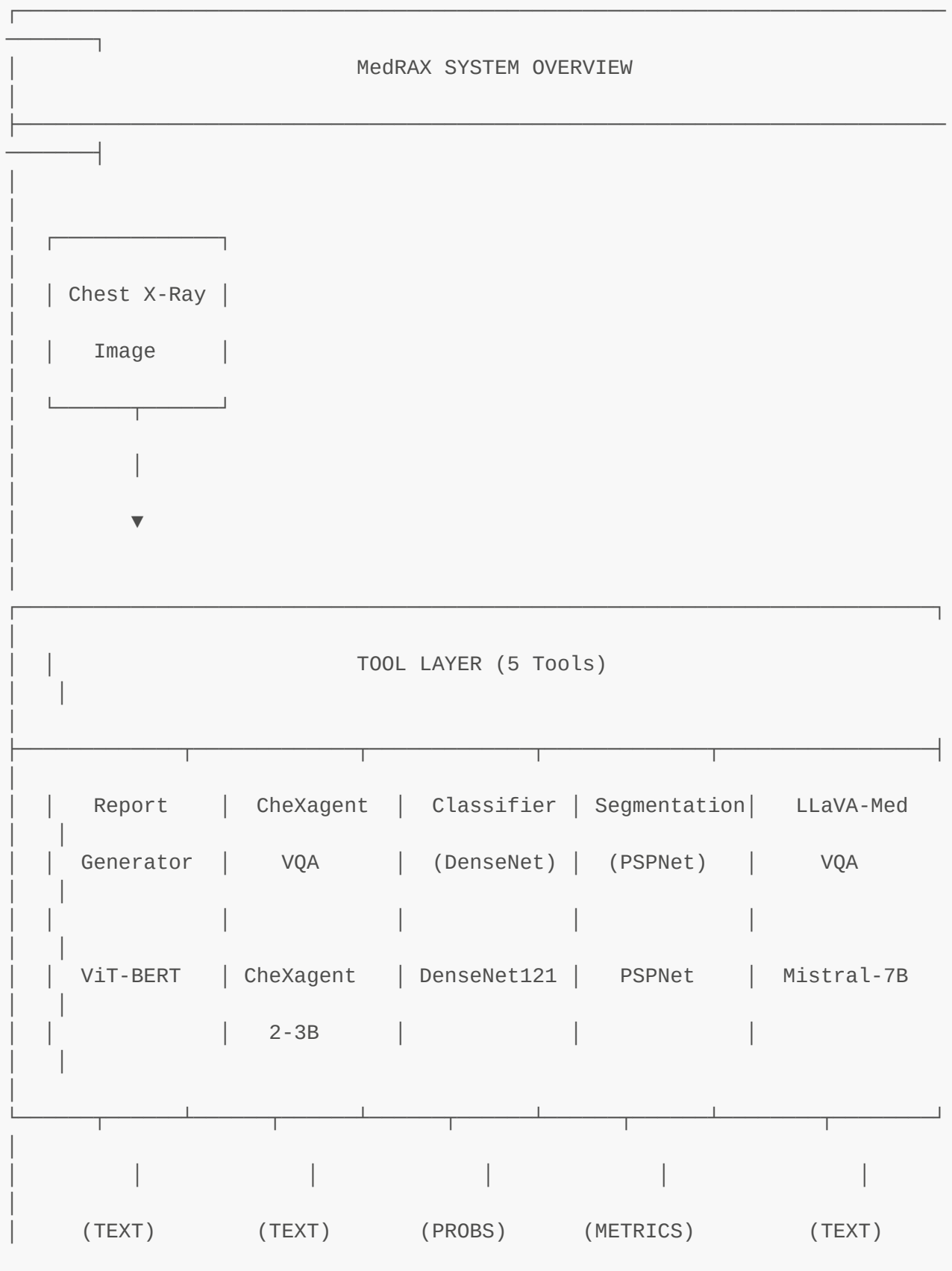
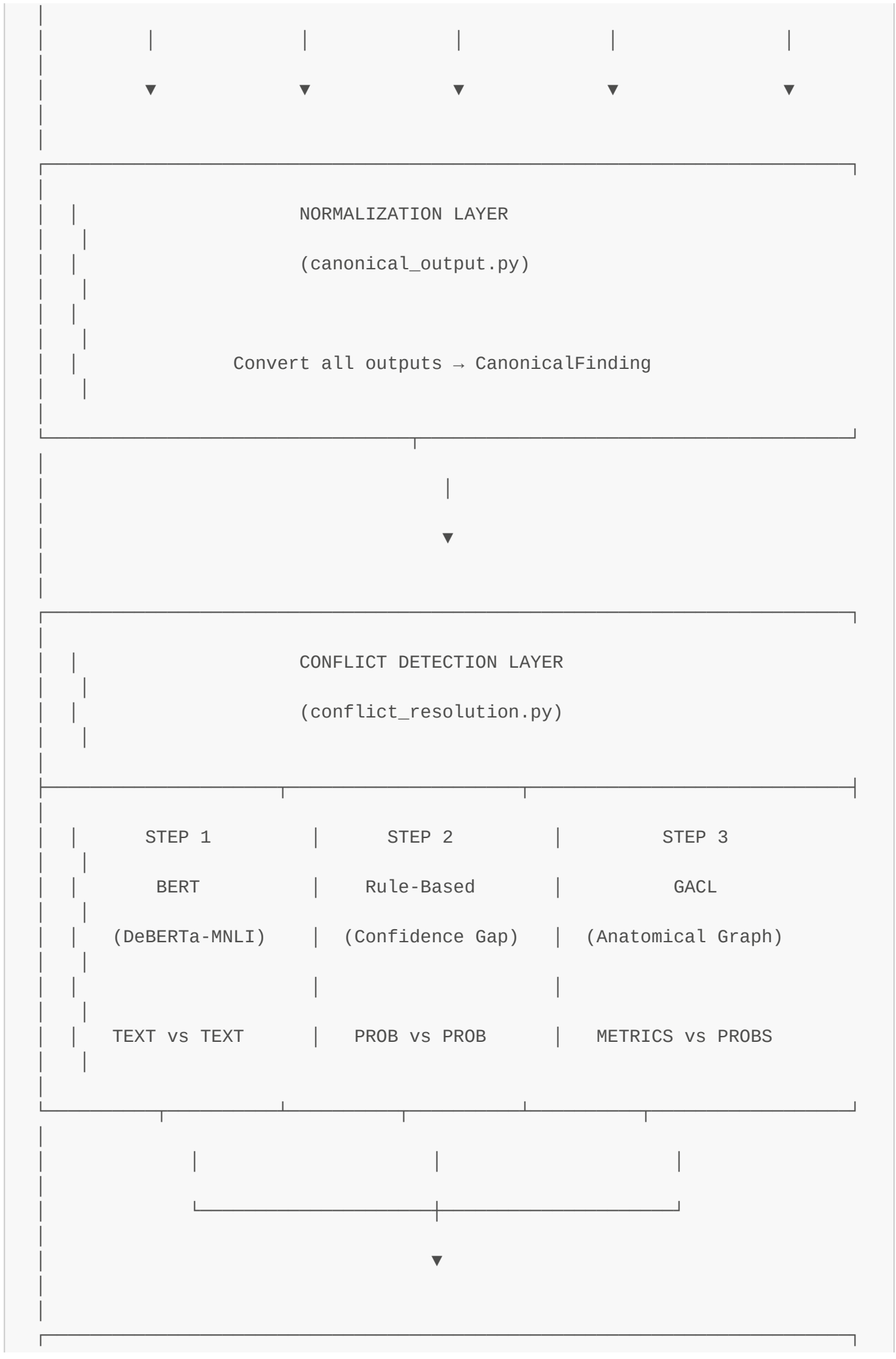
MedRAX Conflict Resolution: Complete Pipeline Analysis

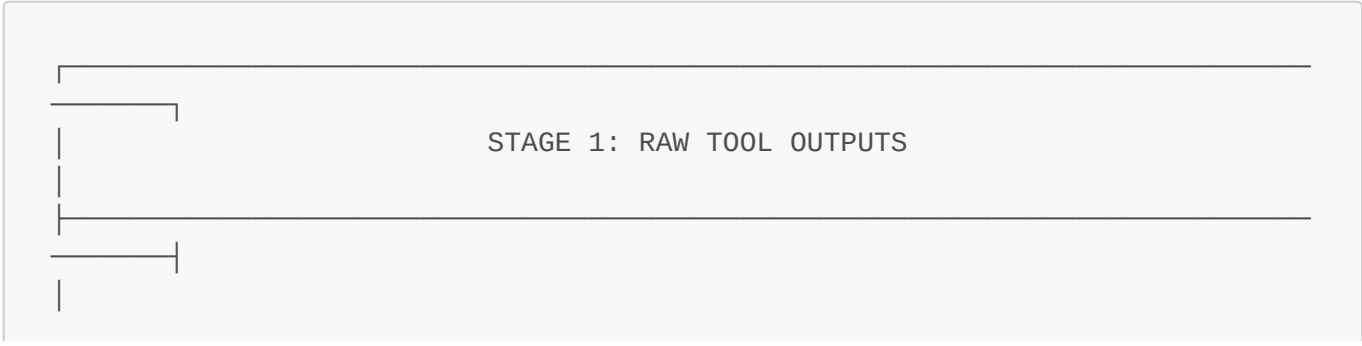
High-Level Architecture




NORMALIZATION LAYER
(canonical_output.py)
Convert all outputs → CanonicalFinding
CONFLICT DETECTION LAYER
(conflict_resolution.py)
STEP 1
BERT
(DeBERTa-MNLI)
TEXT vs TEXT
STEP 2
Rule-Based
(Confidence Gap)
PROB vs PROB
STEP 3
GACL
(Anatomical Graph)
METRICS vs PROBS

Complete Example Scenario: Patient with Suspected Cardiomegaly

Stage 1: Tool Outputs (Raw)




REPORT GENERATOR (ViT-BERT)
Output: "CHEST X-RAY REPORT
FINDINGS:
The cardiac silhouette is mildly enlarged. Lungs are clear.
IMPRESSION:
Mild cardiomegaly. No acute cardiopulmonary process."
Metadata: {overall_confidence: 0.82}
CHEXAGENT VQA (CheXagent-2-3B)
Question: "Is there cardiomegaly in this image?"
Output: "The heart size appears within normal limits. No evidence
of cardiomegaly is seen on this frontal chest radiograph."
Metadata: {self_consistency_score: 0.68}

```
    │
    │
    │
    │
    ┌──────────────────────────────────────────────────────────────────────────────┐
    │
    │   CLASSIFIER (DenseNet-121)
    │    │
    │
    ├──────────────────────────────────────────────────────────────────────────────┤
    │
    │   Output: {
    │    │
    │   │   "Atelectasis": 0.12,
    │    │
    │   │   "Cardiomegaly": 0.22,        ← 22% probability
    │    │
    │   │   "Consolidation": 0.08,
    │    │
    │   │   "Edema": 0.05,
    │    │
    │   │   "Effusion": 0.15,
    │    │
    │   │   "Pneumonia": 0.11,
    │    │
    │   │   "Pneumothorax": 0.03,
    │    │
    │   │   ... (18 pathologies total)
    │    │
    │   }
    │    │
    │
    └──────────────────────────────────────────────────────────────────────────────┘
    │
    │
    │
    │
    ┌──────────────────────────────────────────────────────────────────────────────┐
    │
    │   SEGMENTATION (PSPNet)
    │    │
    │
    ├──────────────────────────────────────────────────────────────────────────────┤
    │
    │   Output: {
    │    │
    │   │   "segmentation_image_path": "temp/segmentation_8022ced0.png",
    │    │
    │   │   "metrics": {
    │    │
    │   │     "Heart": {
    │    │
    │   │       "area_pixels": 58000,
    │    │
```

```
  │   │       "area_cm2": 23.2,
  │   │
  │   │       "width": 185,            ← Heart width in pixels
  │   │
  │   │       "height": 168,
  │   │
  │   │       "confidence_score": 0.94
  │   │
  │   │     },
  │   │
  │   │     "Left Lung": {"width": 162, ...},
  │   │
  │   │     "Right Lung": {"width": 178, ...}
  │   │
  │   │   }
  │   │
  │   │ }
  │   │
  │
  └──────────────────────────────────────────────────────────────────────────────┘
  │
  │
  │
  └──────────────────────────────────────────────────────────────────────────────
  ──────┘
```

Stage 2: Normalization to CanonicalFinding

```
  ┌──────────────────────────────────────────────────────────────────────────────
  ──────┐
  │                        STAGE 2: NORMALIZED FINDINGS
  │
  ├──────────────────────────────────────────────────────────────────────────────
  ──────┤
  │
  │
  │  All tool outputs converted to uniform CanonicalFinding format:
  │
  │
  │
  │
  ┌──────────────────────────────────────────────────────────────────────────────┐
  │
  │   │ Finding 1: Report Generator
  │   │
  │   │   pathology: "Cardiomegaly"
  │   │
  │   │   source_tool: "chest_xray_report_generator"
  │   │
  │   │   confidence: 0.82
```

```
│    │
│  │   evidence_type: "report"
│    │
│  │   raw_value: "Mild cardiomegaly..."
│    │
│  │   metadata: {text: "cardiac silhouette is mildly enlarged..."}
│    │
│
└──────────────────────────────────────────────────────────────────────────────┘
│
│
│
│
┌──────────────────────────────────────────────────────────────────────────────┐
│
│  │ Finding 2: CheXagent VQA
│    │
│  │   pathology: "Cardiomegaly"
│    │
│  │   source_tool: "chest_xray_expert"
│    │
│  │   confidence: 0.25    ← Low because says "normal"
│    │
│  │   evidence_type: "vqa"
│    │
│  │   raw_value: "No evidence of cardiomegaly..."
│    │
│  │   metadata: {text: "heart size appears within normal limits..."}
│    │
│
└──────────────────────────────────────────────────────────────────────────────┘
│
│
│
│
┌──────────────────────────────────────────────────────────────────────────────┐
│
│  │ Finding 3: Classifier
│    │
│  │   pathology: "Cardiomegaly"
│    │
│  │   source_tool: "chest_xray_classifier"
│    │
│  │   confidence: 0.22
│    │
│  │   evidence_type: "classification"
│    │
│  │   raw_value: {"Cardiomegaly": 0.22, ...}
│    │
│  │   metadata: {text: "Cardiomegaly probability: 22%"}
│    │
│
└──────────────────────────────────────────────────────────────────────────────┘
│
```

```
  │
  │
  │
┌─────────────────────────────────────────────────────────────────────────────┐
│
│   | Finding 4: Segmentation
│     │
│   |   pathology: "Cardiomegaly"
│     │
│   |   source_tool: "chest_xray_segmentation"
│     │
│   |   confidence: 0.89    ← High because CTR > 0.5
│     │
│   |   evidence_type: "segmentation"
│     │
│   |   raw_value: {"Heart": {"width": 185}, "Left Lung": {"width": 162}...}
│     │
│   |   metadata: {cardiothoracic_ratio: 0.544}
│     │
│
└─────────────────────────────────────────────────────────────────────────────┘
  │
  │
  │
└──────────────────────────────────────────────────────────────────────────────
──────┘
```

Stage 3: Conflict Detection (Three-Step Process)

```
┌──────────────────────────────────────────────────────────────────────────────
──────┐
│                        STAGE 3: CONFLICT DETECTION
│
├──────────────────────────────────────────────────────────────────────────────
──────┤
│
│
│
═══════════════════════════════════════════════════════════════════════════════
═══════
│
│  STEP 1: BERT (Text vs Text)
│
│
═══════════════════════════════════════════════════════════════════════════════
═══════
│
│
│
│  Comparing: Report Generator ↔ CheXagent VQA
│
│
```

```
│
│  Text 1: "cardiac silhouette is mildly enlarged...Mild cardiomegaly"
│
│  Text 2: "heart size appears within normal limits. No evidence of
cardiomegaly"│
│
│
│  BERT NLI Output:
│
│    ┌──────────────────────────────────────┐
│
│    │ contradiction_prob: 0.91  ✓ HIGH │
│
│    │ entailment_prob:    0.03         │
│
│    │ neutral_prob:       0.06         │
│
│    └──────────────────────────────────────┘
│
│
│
│  Result: ✅ CONFLICT #1 DETECTED
│
│    Type: semantic
│
│    Severity: critical (0.91 > 0.85)
│
│    Tools: [report_generator, chest_xray_expert]
│
│
│
│
═══════════════════════════════════════════════════════════════════════
│
│  STEP 2: Rule-Based (Confidence Gap)
│
│
═══════════════════════════════════════════════════════════════════════
│
│
│
│  Confidences: [0.82, 0.25, 0.22, 0.89]
│
│
│
│  Max: 0.89 (Segmentation)
│
│  Min: 0.22 (Classifier)
│
│  Gap: 0.89 - 0.22 = 0.67 > 0.4 threshold ✓
│
│
│
│  Check: Max (0.89) > 0.7? YES
```

```
│
│  Check: Min (0.22) < 0.3? YES
│
│
│  Result: ✓ CONFLICT #2 DETECTED (but deduplicated - BERT caught similar)
│
│
│
═══════════════════════════════════════════════════════════════════════════
│
│  STEP 3: GACL (Measurements vs Probabilities)
│
═══════════════════════════════════════════════════════════════════════════
│
│
│  Segmentation Measurements:
│
│    Heart width: 185 pixels
│
│    Thorax width: 162 + 178 = 340 pixels
│
│    CTR: 185 / 340 = 0.544
│
│
│  Medical Threshold: CTR > 0.50 = Cardiomegaly
│
│
│  GACL Analysis:
│
│    ┌──────────────────────────────────────────────────────────┐
│    │ Measurements say: CARDIOMEGALY (CTR = 54.4% > 50%)      │
│    │ Classifier says:  NO CARDIOMEGALY (22% probability)     │
│    │                                                          │
│    │ INCONSISTENCY DETECTED!                                  │
│    └──────────────────────────────────────────────────────────┘
│
│
│  Result: ✓ CONFLICT #3 DETECTED
│
│    Type: semantic (anatomical)
│
│    Tools: [segmentation_tool, classifier]
```

```
  │
  │    Explanation: "CTR of 0.544 suggests cardiomegaly but classifier
reports 22%"│
  │
  │
  └──────────────────────────────────────────────────────────────────────────────
──────┘
```

Stage 4: Conflict Resolution

```
  ┌──────────────────────────────────────────────────────────────────────────────
──────┐
  │                      STAGE 4: CONFLICT RESOLUTION
  │
  ├──────────────────────────────────────────────────────────────────────────────
──────┤
  │
  │
  │
  ═══════════════════════════════════════════════════════════════════════════════
  │
  │  RESOLVING CONFLICT #1: Report Generator vs CheXagent
  │
  ═══════════════════════════════════════════════════════════════════════════════
  │
  │
  │
  │  Step 4.1: Analyze BERT Scores
  │
  │    ┌──────────────────────────────────────────────────────────┐
  │
  │    │ contradiction_prob: 0.91 (HIGH)                          │
  │
  │    │ entailment_prob: 0.03 (LOW)                              │
  │
  │    │ is_false_positive: NO (entailment < 0.7)                 │
  │
  │    │ severity_adjustment: 1.0 (no discount needed)            │
  │
  │    └──────────────────────────────────────────────────────────┘
  │
  │
  │
  │  Step 4.2: BERT-Guided Resolution
  │
  │    - BERT contradiction > 0.85? YES (0.91)
  │
  │    - Confidence gap: 0.82 - 0.25 = 0.57 > 0.3? YES
  │
```

```
│    - Clear winner: Report Generator (0.82)
│
│
│  Resolution #1:
│
│    ┌──────────────────────────────────────────────────────────┐
│    | decision: "bert_high_confidence_leader"                  |
│    | selected_tool: "chest_xray_report_generator"             |
│    | value: TRUE (Cardiomegaly present)                       |
│    | confidence: 0.82                                         |
│    | should_defer: FALSE                                      |
│    └──────────────────────────────────────────────────────────┘
│
│
│
═══════════════════════════════════════════════════════════════════
│
│  RESOLVING CONFLICT #3: Segmentation vs Classifier (GACL)
│
═══════════════════════════════════════════════════════════════════
│
│
│  Step 4.3: Task-Aware Arbitration
│
│    - Conflict type: semantic (anatomical)
│
│    - For anatomical measurements: Trust segmentation
│
│    - Segmentation confidence: 0.89
│
│    - Classifier confidence: 0.22
│
│
│  Resolution #3:
│
│    ┌──────────────────────────────────────────────────────────┐
│    | decision: "trust_segmentation_measurements"              |
│    | selected_tool: "chest_xray_segmentation"                 |
│    | value: TRUE (Cardiomegaly present)                       |
│
```

```
│    | confidence: 0.89                                          │
│
│    | reasoning: "CTR=0.544 objectively exceeds 0.50"           │
│
│    | should_defer: FALSE                                       │
│
│    └──────────────────────────────────────────────────────────┘
│
│
│
└──────────────────────────────────────────────────────────────────────────────
──────┘
```

Stage 5: Final Output

```
┌──────────────────────────────────────────────────────────────────────────────
──────┐
│                          STAGE 5: FINAL OUTPUT
│
├──────────────────────────────────────────────────────────────────────────────
──────┤
│
│
│  ⚠  CONFLICT DETECTION REPORT
│
│  ═══════════════════════════════════════════════════════════════════════
│
│  Detected 2 conflict(s)
│
│  Timestamp: 2026-02-05 14:30:00
│
│
│
│  ──────────────────────────────────────────────────────────────────────
│
│  Conflict #1 - CRITICAL SEVERITY
│
│  ──────────────────────────────────────────────────────────────────────
│
│  Type: semantic
│
│  Finding: Cardiomegaly
│
│  Tools: chest_xray_report_generator, chest_xray_expert
│
│    • report_generator: "Mild cardiomegaly" (confidence: 82.0%)
│
│    • chest_xray_expert: "No evidence of cardiomegaly" (confidence: 25.0%)
│
│
```

```
  │
  │  Resolution:
  │
  │    Decision: bert_high_confidence_leader
  │
  │    Selected: chest_xray_report_generator
  │
  │    Confidence: 82.0%
  │
  │    Reasoning: BERT detected high contradiction (91%). Trusting report
  │
  │                generator with significantly higher confidence (82% vs
25%).     │
  │
  │  ──────────────────────────────────────────────────────────────────────
  │
  │  Conflict #2 - CRITICAL SEVERITY
  │
  │  ──────────────────────────────────────────────────────────────────────
  │
  │  Type: semantic (anatomical)
  │
  │  Finding: Anatomical pattern consistency
  │
  │  Tools: segmentation_tool, chest_xray_classifier
  │
  │    • segmentation: CTR=0.544 (confidence: 89.0%)
  │
  │    • classifier: Cardiomegaly=22% (confidence: 22.0%)
  │
  │
  │  Resolution:
  │
  │    Decision: trust_segmentation_measurements
  │
  │    Selected: chest_xray_segmentation
  │
  │    Confidence: 89.0%
  │
  │    Reasoning: Graph-based anatomical analysis shows CTR=0.544 > 0.50
  │
  │                threshold. Objective measurements override classifier.
  │
  │
  │  ══════════════════════════════════════════════════════════════════════
  │
  │  FINAL DIAGNOSIS: CARDIOMEGALY PRESENT
  │
  │  ══════════════════════════════════════════════════════════════════════
  │
  │  Confidence: 85.5% (average of 82% + 89%)
```

```
  │
  │  Supporting Evidence:
  │
  │    ✓ Report Generator: "Mild cardiomegaly" (82%)
  │
  │    ✓ Segmentation: CTR = 0.544 > 0.50 (89%)
  │
  │  Contradicting Evidence:
  │
  │    ✗ CheXagent VQA: "No cardiomegaly" (25%)
  │
  │    ✗ Classifier: 22% probability
  │
  │
  │
  │  Status: ✅ RESOLVED (No human review needed)
  │
  │
  │
  └──────────────────────────────────────────────────────────────────────────────
──────┘
```

Summary: What Each Component Does

```
  ┌──────────────────────────────────────────────────────────────────────────────
──────┐
  │                          COMPONENT RESPONSIBILITIES
  │
  ├──────────────────────────────────────────────────────────────────────────────
──────┤
  │
  │
  │  BERT (DeBERTa-MNLI)
  │
  │  ├── Input: Two text strings
  │
  │  ├── Output: contradiction/entailment/neutral probabilities
  │
  │  ├── Detects: Text-based contradictions
  │
  │  └── Example: "enlarged heart" vs "normal heart" → 91% contradiction
  │
  │
  │
  │  GACL (Graph-Based Anatomical Consistency)
  │
  │  ├── Input: Segmentation metrics + Classifier probabilities
  │
  │  ├── Output: Conflict detected or not + explanation
  │
```

```
│  ├── Detects: Measurement vs probability inconsistencies
│
│  └── Example: CTR=0.544 vs Cardiomegaly=22% → CONFLICT
│
│
│
│  Rule-Based
│
│  ├── Input: List of confidence scores
│
│  ├── Output: Presence conflict detected or not
│
│  ├── Detects: Large confidence gaps (one says YES, another says NO)
│
│  └── Example: 89% vs 22% gap = 67% > 40% threshold → CONFLICT
│
│
│
│  ConflictResolver
│
│  ├── Input: Detected conflicts + all findings
│
│  ├── Output: Resolution decision with reasoning
│
│  ├── Methods: BERT scores → Tool expertise → Weighted average → Defer
│
│  └── Example: High BERT contradiction + confidence leader → Trust winner
│
│
│
└──────────────────────────────────────────────────────────────────────────
──────┘
```

Coverage Matrix

```
┌──────────────────────────────────────────────────────────────────────────
──────┐
│                    WHAT EACH DETECTOR CAN COMPARE
│
├──────────────────────┬──────────────────────┬──────────────────────────────
──────┤
│                      │        BERT          │           GACL
│
├──────────────────────┼──────────────────────┼──────────────────────────────
──────┤
│ Report ↔ CheXagent   │          ✅          │           ❌
│
│ Report ↔ LLaVA-Med   │          ✅          │           ❌
│
│ CheXagent ↔ LLaVA    │          ✅          │           ❌
```

```
┌──────────────────────────┬──────────────────────────┬──────────────────────────────────┐
│ Segmentation ↔ Class     │            ✗             │              ✓                   │
├──────────────────────────┼──────────────────────────┼──────────────────────────────────┤
│ Coverage                 │    TEXT vs TEXT          │    NUMBERS vs NUMBERS            │
└──────────────────────────┴──────────────────────────┴──────────────────────────────────┘

Together: Complete coverage of all tool output combinations ✓
```

Models Used

Tool	Model	Output Type
Report Generator	`IAMJB/chexpert-mimic-cxr-*-baseline` (ViT-BERT)	Text
CheXagent VQA	`StanfordAIMI/CheXagent-2-3b`	Text
Classifier	`torchxrayvision.models.DenseNet` (densenet121-res224-all)	18 Probabilities
Segmentation	`torchxrayvision.baseline_models.chestx_det.PSPNet`	14 Organ Metrics
LLaVA-Med	`microsoft/llava-med-v1.5-mistral-7b`	Text
BERT Conflict Detector	`microsoft/deberta-v3-large-mnli`	NLI Scores

Key Files

File	Purpose
`medrax/agent/conflict_resolution.py`	Conflict detection & resolution logic
`medrax/agent/bert_conflict_detector.py`	BERT NLI-based text comparison
`medrax/agent/anatomical_consistency_graph.py`	GACL measurement analysis
`medrax/agent/canonical_output.py`	Output normalization
`medrax/tools/classification.py`	DenseNet classifier
`medrax/tools/segmentation.py`	PSPNet segmentation
`medrax/tools/report_generation.py`	ViT-BERT report generator
`medrax/tools/xray_vqa.py`	CheXagent VQA