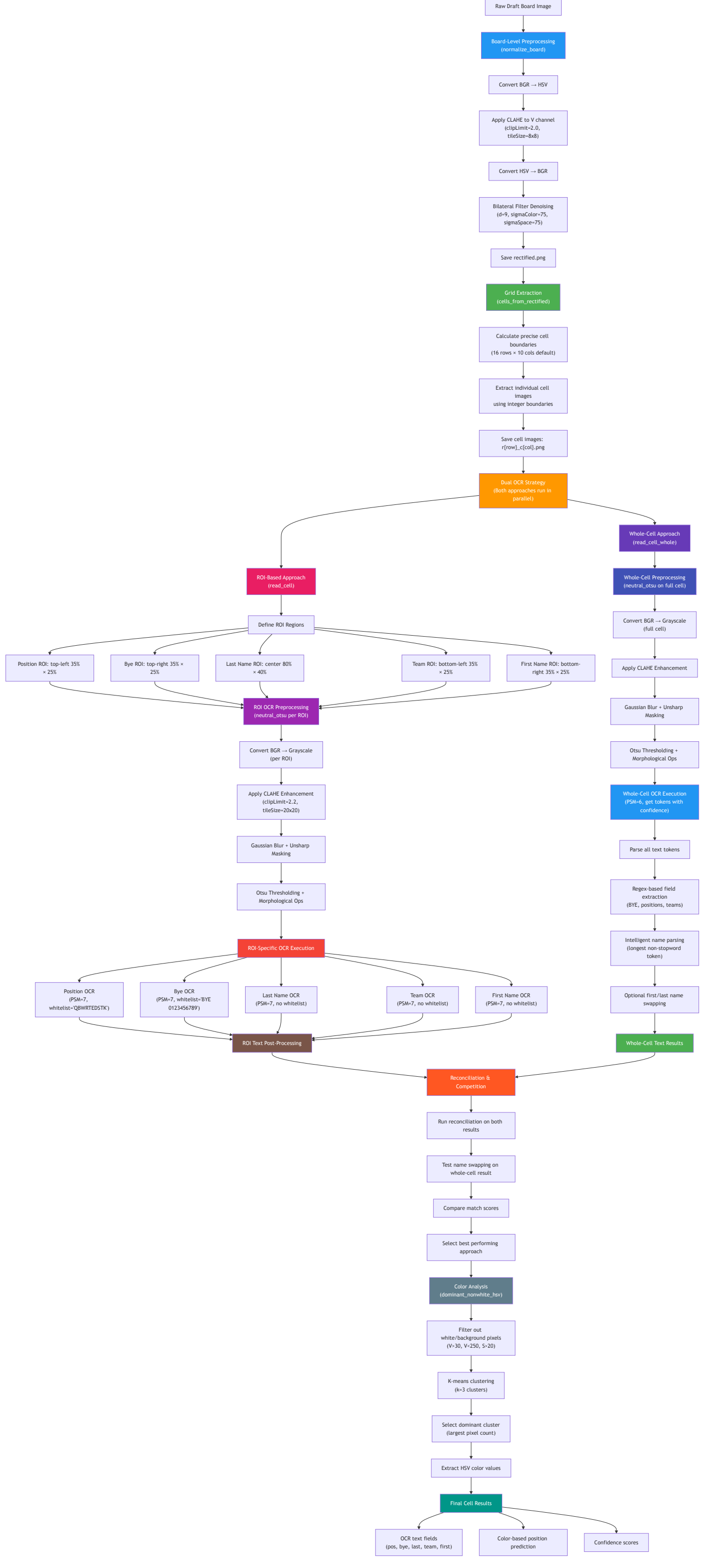


# Image Preprocessing and OCR Pipeline



## Pipeline Stages Legend

- Board-Level Preprocessing:** CLAHE enhancement and bilateral filtering on full board
- Grid Extraction:** Precise cell boundary calculation and individual cell extraction
- Dual OCR Strategy:** Both ROI and whole-cell approaches run in parallel
- ROI-Based Approach:** Targeted OCR on 5 specific regions per cell
- Whole-Cell Approach:** Single OCR pass on entire cell with intelligent parsing
- Reconciliation & Competition:** Both approaches compete, best result wins
- Color Analysis:** K-means clustering for position prediction
- Final Results:** Combined OCR text, color prediction, and confidence scores

## Key Features:

- Dual Strategy:** ROI-based and whole-cell OCR run simultaneously
- Competitive Selection:** System automatically chooses the best performing result
- Robust Preprocessing:** Multi-step neutral\_otsu enhancement pipeline
- Intelligent Parsing:** Regex-based field extraction with stopwords filtering
- Name Swapping:** Tests first/last name combinations for better matches
- Color Validation:** K-means clustering provides secondary position verification