

Solstice: Technical Architecture

Computer Vision + Multi-LLM Pipeline for Medical Document Verification

1 Technical Implementation

The Solstice pipeline processes medical documents through computer vision-based layout detection followed by multi-stage LLM analysis. Each stage builds on the previous, creating a traceable verification chain from raw PDF to final evidence report.

2 Pipeline Architecture

1. **Layout Detection:** Detectron2 model identifies bounding boxes for text blocks, tables, figures, and captions
2. **Content Extraction:** Parser converts detected regions into structured JSON with text content and metadata
3. **Claim Processing:** User-provided claims are processed through orchestrated LLM pipeline
4. **LLM Execution:** Four specialized LLM calls analyze documents with specific prompts
5. **Result Caching:** All intermediate outputs stored in hierarchical cache structure for reproducibility

3 Implementation Details

3.1 Layout Detection Pipeline

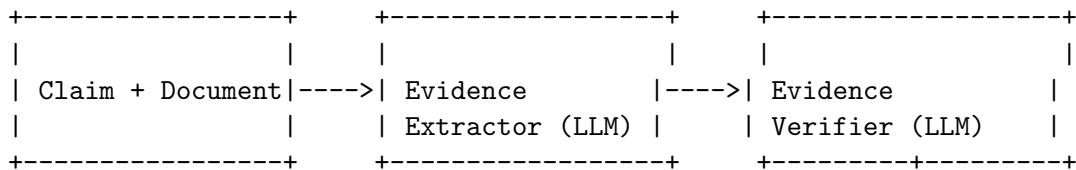
The layout detection uses Detectron2 with ResNet-50 backbone. The model outputs bounding boxes with class labels:

The detection pipeline:

- Processes PDFs page-by-page at 400 DPI resolution
- Runs inference with confidence threshold of 0.2
- Merges overlapping boxes to reduce duplicates
- Outputs JSON with bbox coordinates and class labels

3.2 Multi-Stage LLM Processing

The claim verification uses four LLM calls (Claude-3.5-Sonnet):



Total elements: 15
Reading order: 15 elements

Vaccine 29 (2011) 7733–7739



Protective efficacy of a trivalent recombinant hemagglutinin protein vaccine (FluBlok®) against influenza in healthy adults: A randomized, placebo-controlled trial^a

John Treanor^{a,*}, Hana El Sahly^b, James King^c, Irene Graham^d, Ruvim Izikson^e, Robert Kohberger^f, Peter Patriarca^g, Manon Cox^e

^a University of Rochester, Rochester, NY, United States
^b Baylor College of Medicine, Houston, TX, United States
^c University of Maryland, Baltimore, MD, United States
^d St. Louis University, St. Louis, MO, United States
^e Protein Sciences Corporation, Meriden, CT, United States
^f Blair & Co, Greenwich, CT, United States
^g Biologics Consulting Group, Inc., Bethesda, MD, United States

ARTICLE INFO

Article history:
Received 20 May 2011
Received in revised form 27 July 2011
Accepted 27 July 2011
Available online 9 August 2011

Keywords:
Influenza vaccine
Baculovirus expression
Recombinant protein
Clinical trials

ABSTRACT

Background: Development of influenza vaccines that do not use embryonated eggs as the substrate for vaccine production is a high priority. We conducted this study to determine the protective efficacy of a recombinant, baculovirus-expressed seasonal trivalent influenza virus hemagglutinin (rHA0) vaccine (FluBlok®).
Methods: Healthy adult subjects at 24 centers across the US were randomly assigned to receive a single injection of saline placebo (2304 subjects), or trivalent FluBlok containing 45 mcg of each rHA0 component (2344 subjects). Serum samples for assessment of immune responses by hemagglutination-inhibition (HAI) were taken from a subset of subjects before and 28 days after immunization. Subjects were followed during the 2007–2008 influenza season and combined nasal and throat swabs for virus isolation were obtained from subjects reporting influenza-like illness.
Results: Rates of local and systemic side effects were low, and the rates of systemic side effects were similar in the vaccine and placebo groups. HAI antibody responses were seen in 78%, 81%, and 52% of FluBlok recipients to the H1, H3, and B components, respectively. FluBlok was 44.6% (95% CI, 18.8%, 62.6%) effective in preventing culture-confirmed influenza meeting the CDC influenza-like illness case definition despite significant antigenic mismatch between the vaccine antigens and circulating viruses.
Conclusions: Trivalent rHA0 vaccine was safe, immunogenic and effective in the prevention of culture confirmed influenza illness, including protection against drift variants.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Though embryonated hen's eggs have been used to generate effective influenza vaccines for many years, this system does have several important drawbacks. Vaccine manufacturing using eggs requires specialized facilities, and the ability to scale up egg production rapidly in response to an emergency is limited. In addition, poultry are potentially vulnerable to the same subtypes of influenza

that might also be responsible for pandemic influenza. It is usually necessary to adapt candidate vaccine viruses for high yield growth in eggs, a process that can be time consuming, is not always successful, and which can select receptor variants that may not be optimally representative of circulating influenza strains [1,2].

Expression of proteins in insect cells using recombinant baculovirus has emerged as a promising technology for vaccine production. New recombinant baculoviruses can be generated quickly from sequence data, protein expression is very efficient under the control of the baculovirus polyhedrin promoter, and post-translational modifications of the protein are generally similar to other eukaryotic systems. In previous studies, we have evaluated baculovirus-expressed recombinant influenza virus hemagglutinins (rHA0s) as influenza vaccines in humans. Monovalent and bivalent rHA0s have been well tolerated and immunogenic in

^a [Trials.gov](http://www.trials.gov) Identifier: NCT00539981.

* Corresponding author at: University of Rochester Medical Center, Room 3-6308, 601 Elmwood Avenue, Rochester, NY 14642, United States. Tel.: +1 585 275 5871; fax: +1 585 442 9328.

E-mail address: John.Treanor@urmc.rochester.edu (J.J. Treanor).

13: Text X/s – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.vaccine.2011.07.128

Figure 1: Detectron2 output showing bounding box detection on clinical trial paper. Colors indicate detected classes: text (blue), table (green), figure (red), caption (yellow).


```
claim_{id}/
    evidence_extractor/output.json
    evidence_verifier/output.json
    completeness_checker/output.json
    image_evidence_analyzer/
        {figure_id}/output.json
visualizations/          # Layout detection overlays
```

3.4 Marketing Pipeline Adaptation

Marketing materials require modified processing due to different layout patterns:

Page 1 - Layout Detection

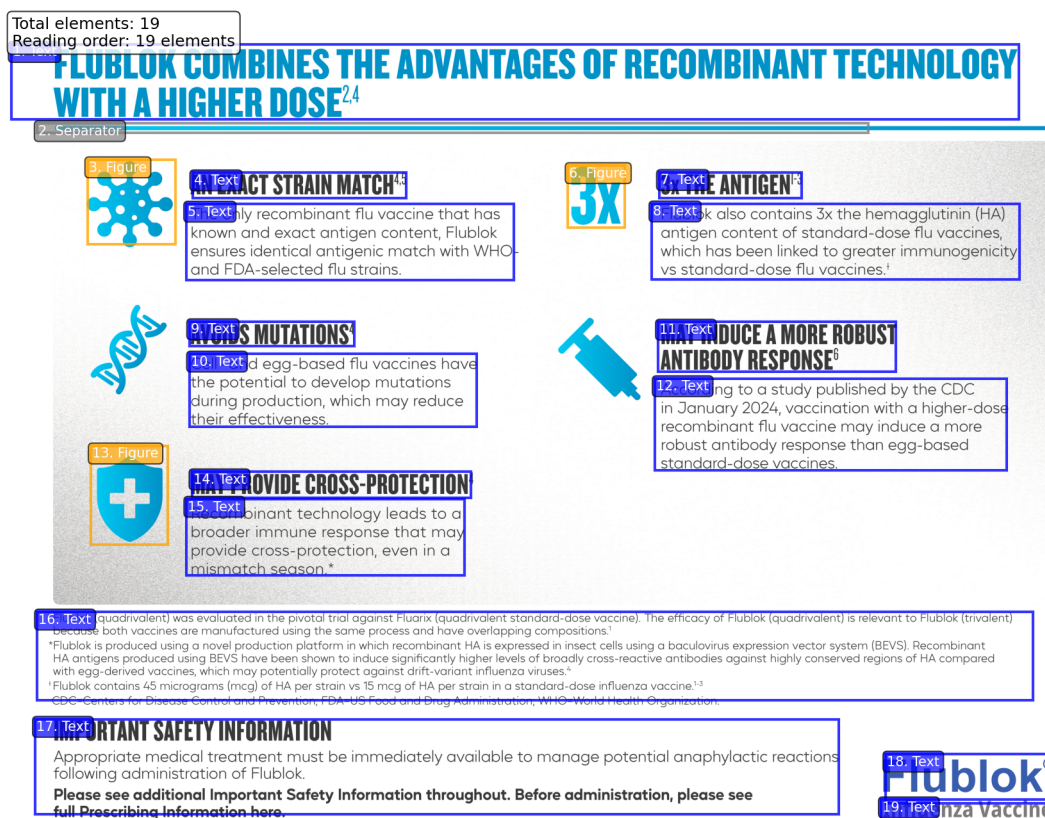


Figure 2: Marketing layout detection shows emphasis on visual hierarchy and promotional elements. The pipeline adapts confidence thresholds for better detection of design-heavy layouts.

Key differences:

- Lower confidence threshold (0.1 vs 0.2) for creative layouts
- Additional post-processing for overlapping design elements
- Separate cache directory (data/marketing_cache/)
- Same LLM prompts used for both document types