Keeping Informed: Automatic Processing of Residual Functional Capacity Form Images

JULIA PORCINO AND CHUNXIAO ZHOU

HIP'19
SEPTEMBER 20-21, 2019

NIH National Institutes of Health
Clinical Center

SOCIAL SECURITY ADMINISTRATION USA

Acknowledgements

This research was supported by the Intramural Research Program of the National Institutes of Health and the US Social Security Administration

All opinions expressed here are the authors and not those of the US government.

We have no conflicts of interest to disclose.

Background

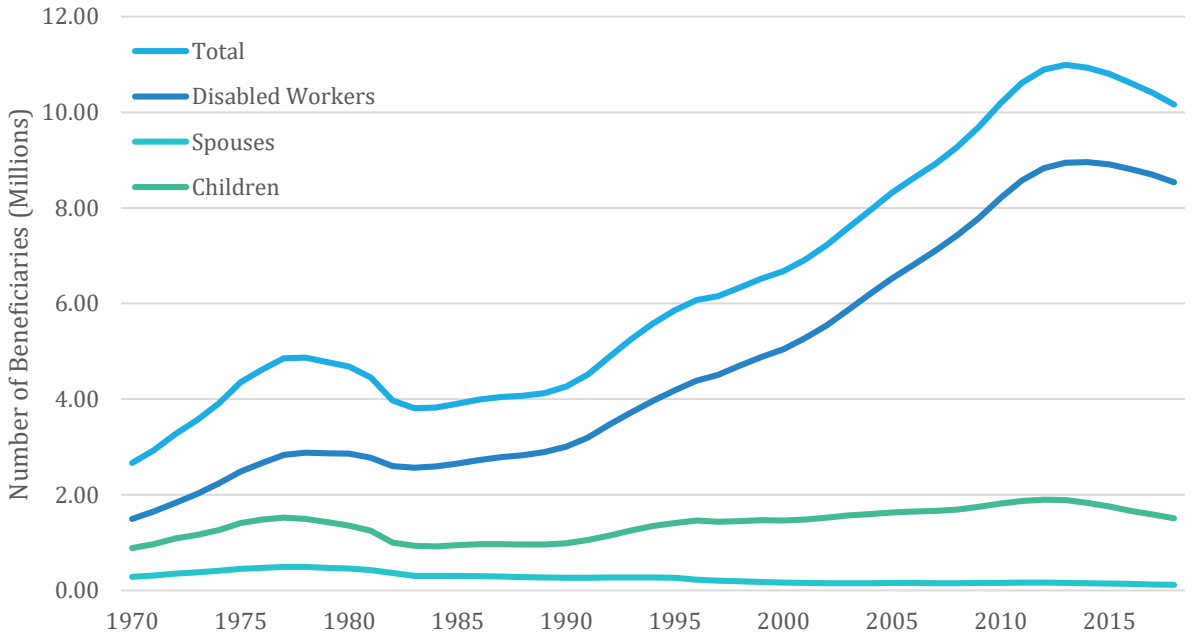
US Social Security Administration (SSA)

Disability Programs:

- Work disability
  - Cash & Health Insurance
- >10 million beneficiaries
- 2-3 million new applications

Adjudication Process:

- Manual review
  - External medical records and evidence
  - Internal administrative & case processing data

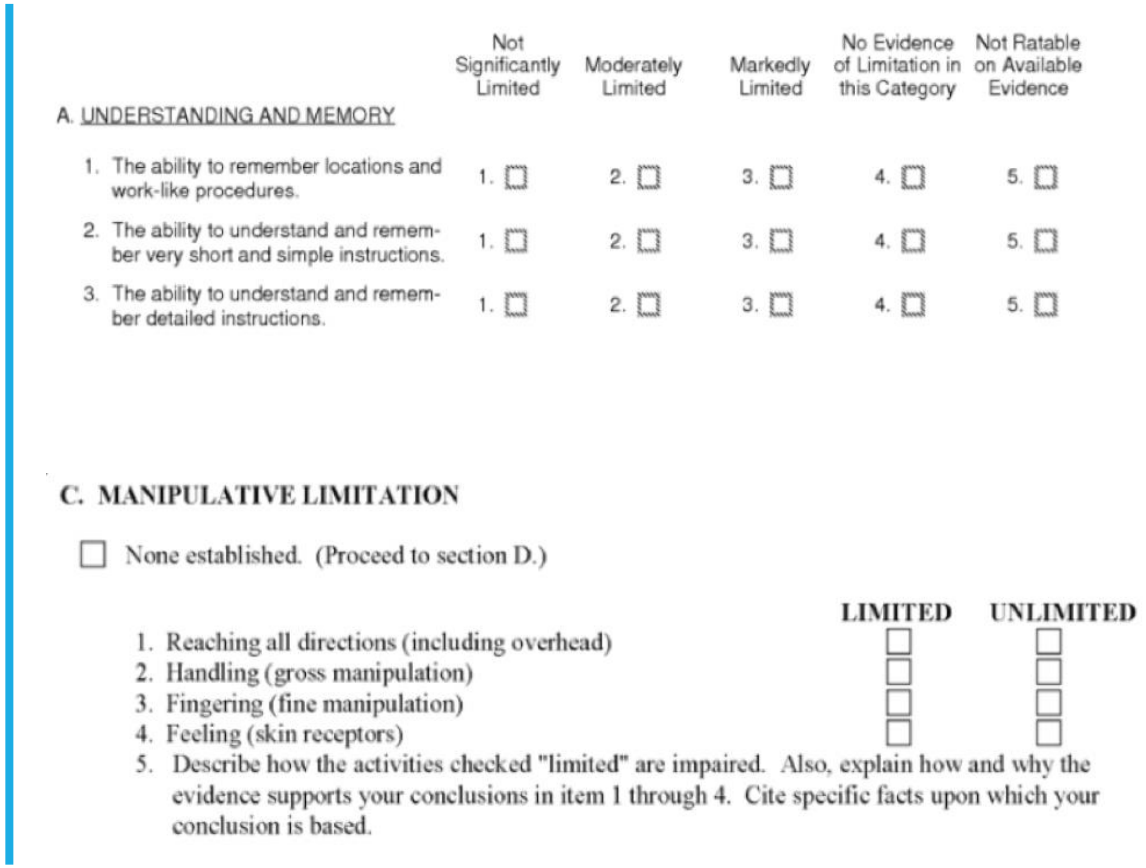


SSA Office of the Chief Actuary: https://www.ssa.gov/oact/STATS/DIbenies.html

Residual Functional Capacity (RFC) Forms

Function as relates to work

- Mental and Physical RFCs
- Checkboxes and free text
- Currently: electronic database
- Historically: “paper” form



	Not Significantly Limited	Moderately Limited	Markedly Limited	No Evidence of Limitation in this Category	Not Ratable on Available Evidence
A. UNDERSTANDING AND MEMORY					
1. The ability to remember locations and work-like procedures.	1. ☐	2. ☐	3. ☐	4. ☐	5. ☐
2. The ability to understand and remember very short and simple instructions.	1. ☐	2. ☐	3. ☐	4. ☐	5. ☐
3. The ability to understand and remember detailed instructions.	1. ☐	2. ☐	3. ☐	4. ☐	5. ☐

C. MANIPULATIVE LIMITATION

☐ None established. (Proceed to section D.)

	LIMITED	UNLIMITED
1. Reaching all directions (including overhead)	☐	☐
2. Handling (gross manipulation)	☐	☐
3. Fingering (fine manipulation)	☐	☐
4. Feeling (skin receptors)	☐	☐

5. Describe how the activities checked "limited" are impaired. Also, explain how and why the evidence supports your conclusions in item 1 through 4. Cite specific facts upon which your conclusion is based.

Motivation

Why are we interested in historical RFC Forms?

- Update current databases with historical form data
- Assess change in function over time
- Comparison to other sources of function

Millions of paper forms

- Forms used since 1980s
- Want automatic way to extract information

Challenges

SSA Data

SSA stores all documents as TIF images

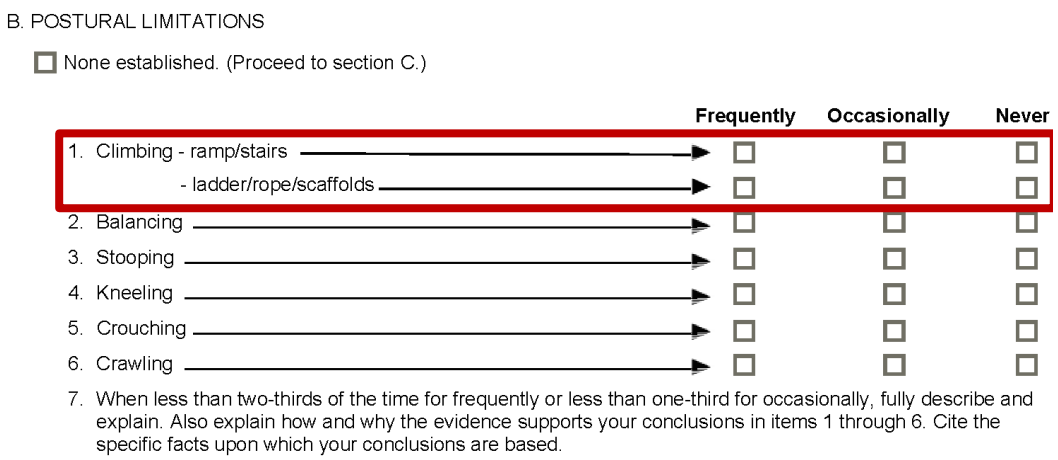
- Limitations with existing software

RFC forms come from templates that can be edited

- Base content (generally) remains consistent
- Layout varies greatly

RFC Form Variation

Number of checkboxes per section:

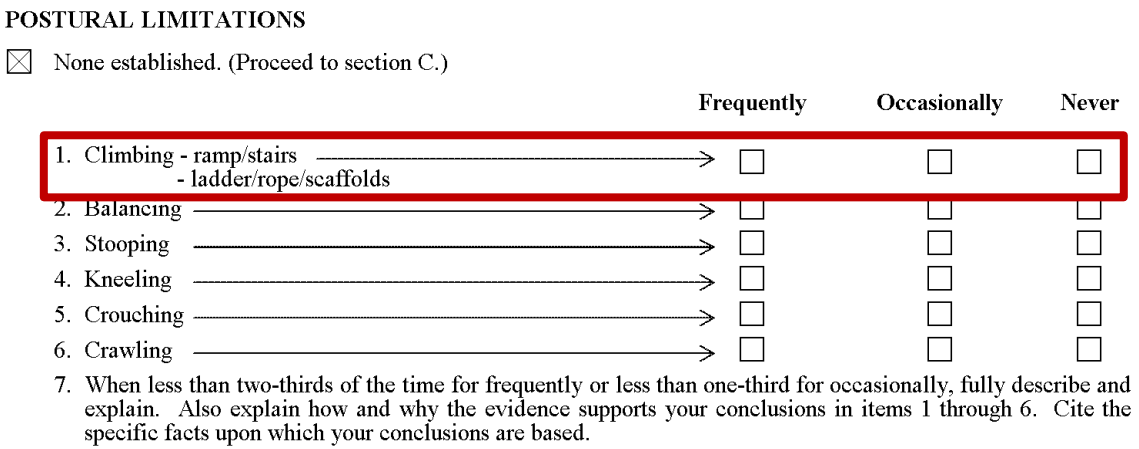


B. POSTURAL LIMITATIONS

☐ None established. (Proceed to section C.)

	Frequently	Occasionally	Never
1. Climbing - ramp/stairs	☐	☐	☐
- ladder/rope/scaffolds	☐	☐	☐
2. Balancing	☐	☐	☐
3. Stooping	☐	☐	☐
4. Kneeling	☐	☐	☐
5. Crouching	☐	☐	☐
6. Crawling	☐	☐	☐

7. When less than two-thirds of the time for frequently or less than one-third for occasionally, fully describe and explain. Also explain how and why the evidence supports your conclusions in items 1 through 6. Cite the specific facts upon which your conclusions are based.



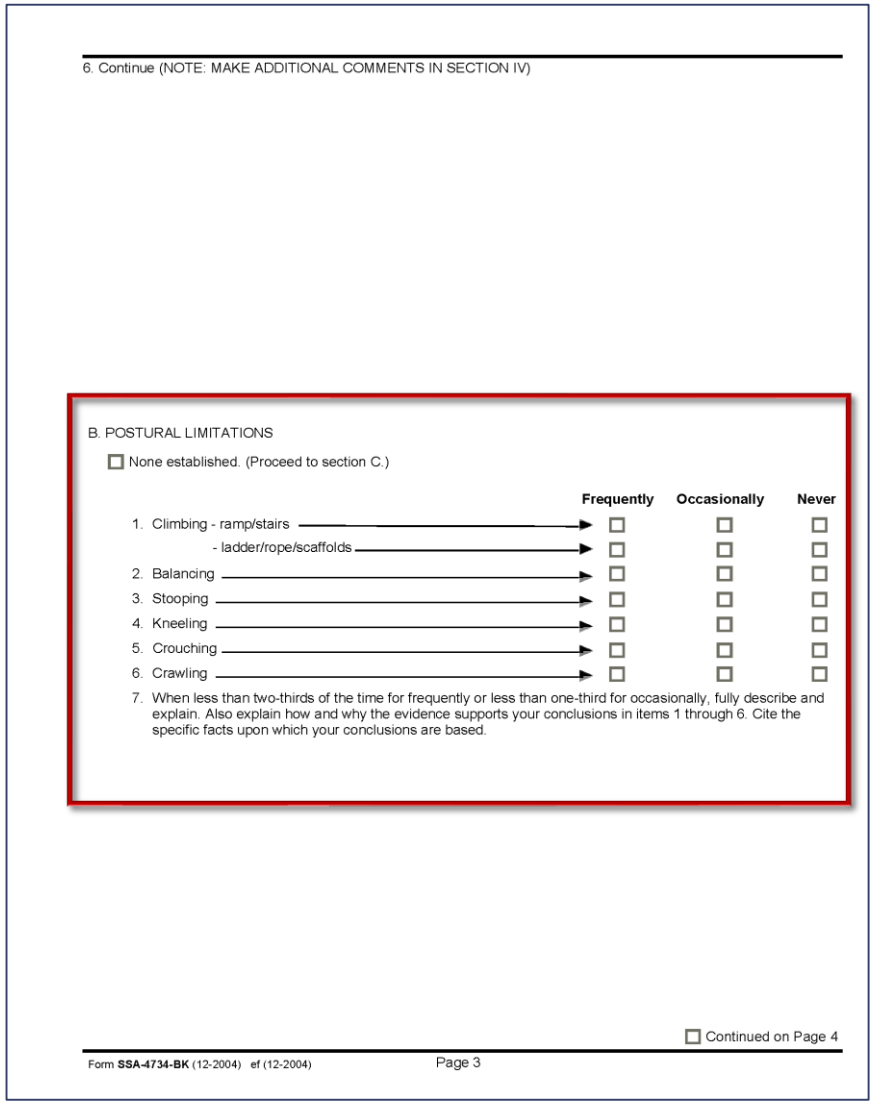
B. POSTURAL LIMITATIONS

☒ None established. (Proceed to section C.)

	Frequently	Occasionally	Never
1. Climbing - ramp/stairs	☐	☐	☐
- ladder/rope/scaffolds			
2. Balancing	☐	☐	☐
3. Stooping	☐	☐	☐
4. Kneeling	☐	☐	☐
5. Crouching	☐	☐	☐
6. Crawling	☐	☐	☐

7. When less than two-thirds of the time for frequently or less than one-third for occasionally, fully describe and explain. Also explain how and why the evidence supports your conclusions in items 1 through 6. Cite the specific facts upon which your conclusions are based.

Sections per page:



6. Continue (NOTE: MAKE ADDITIONAL COMMENTS IN SECTION IV)

B. POSTURAL LIMITATIONS

☐ None established. (Proceed to section C.)

	Frequently	Occasionally	Never
1. Climbing - ramp/stairs	☐	☐	☐
- ladder/rope/scaffolds	☐	☐	☐
2. Balancing	☐	☐	☐
3. Stooping	☐	☐	☐
4. Kneeling	☐	☐	☐
5. Crouching	☐	☐	☐
6. Crawling	☐	☐	☐

7. When less than two-thirds of the time for frequently or less than one-third for occasionally, fully describe and explain. Also explain how and why the evidence supports your conclusions in items 1 through 6. Cite the specific facts upon which your conclusions are based.

☐ Continued on Page 4

Form **SSA-4734-BK** (12-2004) ef (12-2004) Page 3

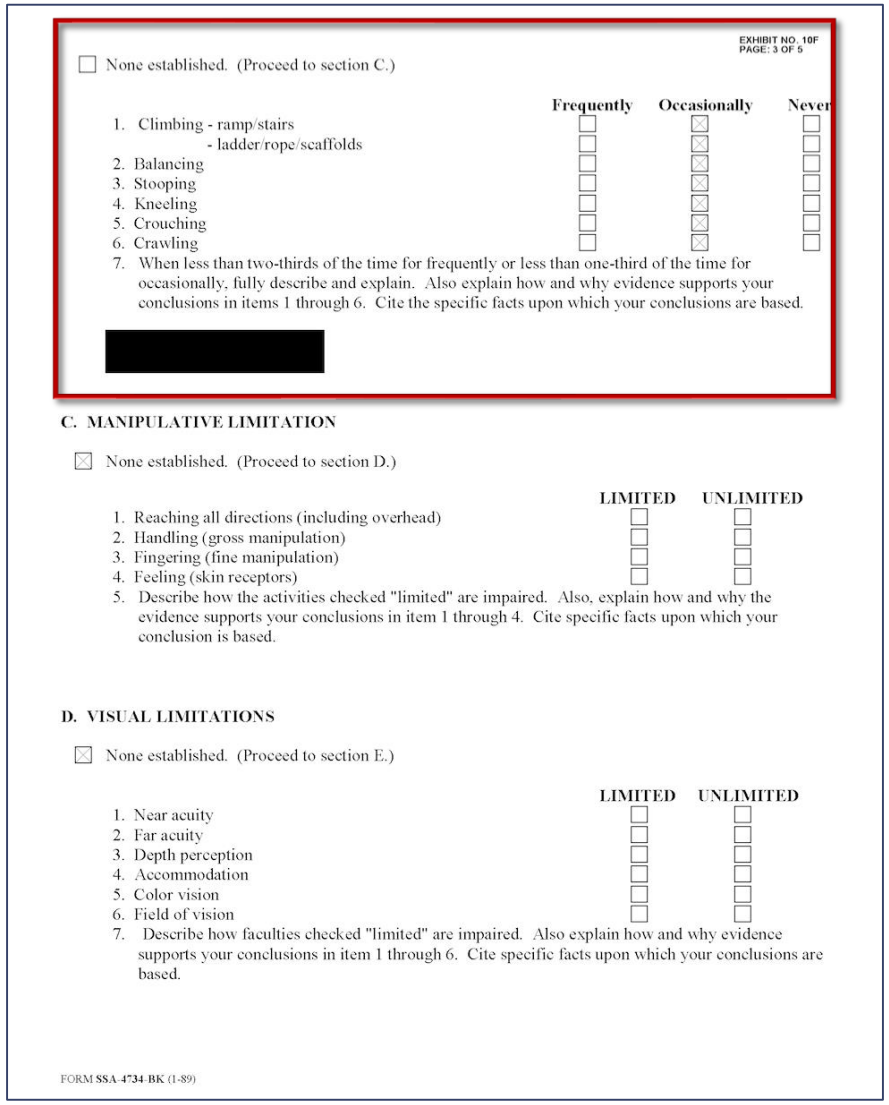


EXHIBIT NO. 10F
PAGE: 3 OF 5

☐ None established. (Proceed to section C.)

	Frequently	Occasionally	Never
1. Climbing - ramp/stairs	☐	☒	☐
- ladder/rope/scaffolds	☐	☒	☐
2. Balancing	☐	☒	☐
3. Stooping	☐	☒	☐
4. Kneeling	☐	☒	☐
5. Crouching	☐	☒	☐
6. Crawling	☐	☒	☐

7. When less than two-thirds of the time for frequently or less than one-third of the time for occasionally, fully describe and explain. Also explain how and why evidence supports your conclusions in items 1 through 6. Cite the specific facts upon which your conclusions are based.

[redacted]

C. MANIPULATIVE LIMITATION

☒ None established. (Proceed to section D.)

	LIMITED	UNLIMITED
1. Reaching all directions (including overhead)	☐	☐
2. Handling (gross manipulation)	☐	☐
3. Fingering (fine manipulation)	☐	☐
4. Feeling (skin receptors)	☐	☐

5. Describe how the activities checked "limited" are impaired. Also, explain how and why the evidence supports your conclusions in item 1 through 4. Cite specific facts upon which your conclusion is based.

D. VISUAL LIMITATIONS

☒ None established. (Proceed to section E.)

	LIMITED	UNLIMITED
1. Near acuity	☐	☐
2. Far acuity	☐	☐
3. Depth perception	☐	☐
4. Accommodation	☐	☐
5. Color vision	☐	☐
6. Field of vision	☐	☐

7. Describe how faculties checked "limited" are impaired. Also explain how and why evidence supports your conclusions in item 1 through 6. Cite specific facts upon which your conclusions are based.

FORM SSA-4734-BK (1-89)

Section Spans Two Pages:

B. **POSTURAL LIMITATIONS**

☐ None established. (Proceed to section C.)

	Frequently	Occasionally	Never
1. Climbing - ramp/stairs	☒	☐	☐
- ladder/rope/scaffolds	☐	☒	☐
2. Balancing	☒	☐	☐
3. Stooping	☐	☒	☐
4. Kneeling	☐	☒	☐
5. Crouching	☐	☒	☐
6. Crawling	☐	☒	☐

7. When less than two-thirds of the time for frequently or less than one-third of the time for occasionally, fully describe and explain. Also explain how and why evidence supports your conclusions in items 1 through 6. Cite the specific facts upon which your conclusions are based.

C. MANIPULATIVE LIMITATION

☒ None established. (Proceed to section D.)

	LIMITED	UNLIMITED
1. Reaching all directions (including overhead)	☐	☐
2. Handling (gross manipulation)	☐	☐

266

3. Fingering (fine manipulation) ☐ ☐
4. Feeling (skin receptors) ☐ ☐

EXHIBIT NO. 8F
PAGE: 5 OF 8

5. Describe how the activities checked "limited" are impaired. Also, explain how and why the evidence supports your conclusions in item 1 through 4. Cite specific facts upon which your conclusion is based.

D. VISUAL LIMITATIONS

☒ None established. (Proceed to section E.)

	LIMITED	UNLIMITED
1. Near acuity	☐	☐
2. Far acuity	☐	☐
3. Depth perception	☐	☐
4. Accommodation	☐	☐
5. Color vision	☐	☐
6. Field of vision	☐	☐

7. Describe how faculties checked "limited" are impaired. Also explain how and why evidence supports your conclusions in item 1 through 6. Cite specific facts upon which your conclusions are based.

E. COMMUNICATIVE LIMITATION

☒ None established. (Proceed to section F.)

	LIMITED	UNLIMITED
1. Hearing	☐	☐
2. Speaking	☐	☐

3. Describe how the faculties checked in "limited" are impaired. Also, explain how and why the evidence supports your conclusions in items 1 and 2. Cite the specific facts upon which your conclusions are based.

F. ENVIRONMENTAL LIMITATIONS

☐ None established. (Proceed to section II.)

	UNLIMITED	AVOID CONCENTRATED EXPOSURE	AVOID EVEN MODERATE EXPOSURE	AVOID ALL EXPOSURE
1. Extreme cold	☐	☒	☐	☐
2. Extreme heat	☒	☐	☐	☐
3. Wetness	☐	☒	☐	☐
4. Humidity	☒	☐	☐	☐

267

5. Noise ☒ ☐ ☐
6. Vibration ☐ ☒ ☐
7. Fumes, odors, dusts, gases, poor ventilation, etc. ☒ ☐ ☐
8. Hazards (machinery, heights, etc.) ☐ ☒ ☐
9. Describe how these environmental factors impair activities and identify hazards to be avoided. Also, explain how and why evidence supports your conclusions in items 1 through 8. Cite the specific facts upon which your conclusions are based.

EXHIBIT NO. 8F
PAGE: 6 OF 8

II. SYMPTOMS

For symptoms alleged by the claimant to produce physical limitations, and for which the following have not previously been addressed in section I, discuss whether:

A. The symptom(s) is attributable, in your judgment, to a medically determinable impairment.

B. The severity or duration of the symptom(s), in your judgment, is disproportionate to the expected severity or expected duration on the basis of the claimant's medically determinable impairment(s).

C. The severity of the symptom(s) and it's alleged effect on function is consistent, in your judgment, with the total medical and non-medical evidence, including statements by the claimant and others, observations regarding activities of daily living, and alterations of usual behavior or habits.

III. TREATING OR EXAMINING SOURCE STATEMENT(S)

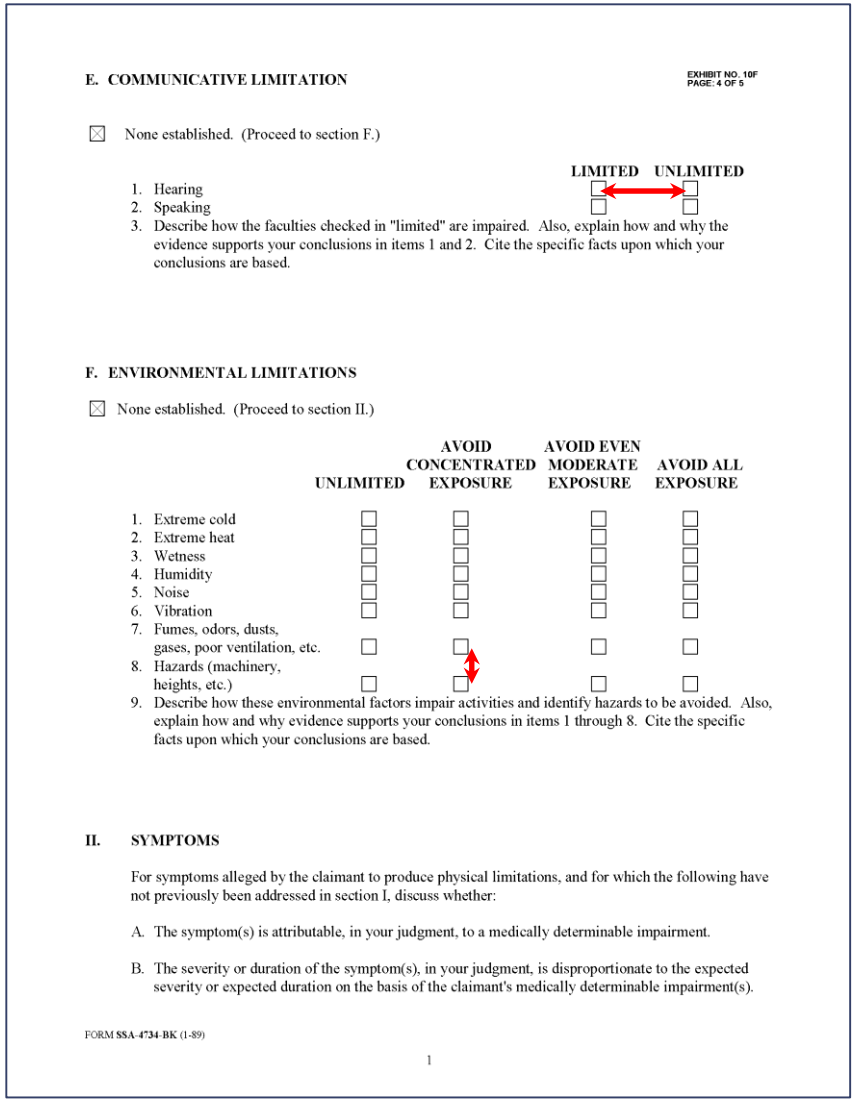
A. Is treating or examining source statement(s) regarding the claimant's physical capacities in file?

☐ Yes ☒ No (includes situations which there was no source or when the source(s) did not provide a statement regarding the claimant's physical capacities.)

B. If yes, are there treating/examining source conclusions about the claimant's limitations or restrictions which are significantly different from your findings?

268

Distance between rows and columns:



E. COMMUNICATIVE LIMITATION

EXHIBIT NO. 10F
PAGE: 4 OF 5

☒ None established. (Proceed to section F.)

	LIMITED	UNLIMITED
1. Hearing	☐	☐
2. Speaking	☐	☐

3. Describe how the faculties checked in "limited" are impaired. Also, explain how and why the evidence supports your conclusions in items 1 and 2. Cite the specific facts upon which your conclusions are based.

F. ENVIRONMENTAL LIMITATIONS

☒ None established. (Proceed to section II.)

	UNLIMITED	AVOID CONCENTRATED EXPOSURE	AVOID EVEN MODERATE EXPOSURE	AVOID ALL EXPOSURE
1. Extreme cold	☐	☐	☐	☐
2. Extreme heat	☐	☐	☐	☐
3. Wetness	☐	☐	☐	☐
4. Humidity	☐	☐	☐	☐
5. Noise	☐	☐	☐	☐
6. Vibration	☐	☐	☐	☐
7. Fumes, odors, dusts, gases, poor ventilation, etc.	☐	☐	☐	☐
8. Hazards (machinery, heights, etc.)	☐	☐	☐	☐

9. Describe how these environmental factors impair activities and identify hazards to be avoided. Also, explain how and why evidence supports your conclusions in items 1 through 8. Cite the specific facts upon which your conclusions are based.

II. SYMPTOMS

For symptoms alleged by the claimant to produce physical limitations, and for which the following have not previously been addressed in section I, discuss whether:

A. The symptom(s) is attributable, in your judgment, to a medically determinable impairment.

B. The severity or duration of the symptom(s), in your judgment, is disproportionate to the expected severity or expected duration on the basis of the claimant's medically determinable impairment(s).

FORM SSA-4734-BK (1-89)

1

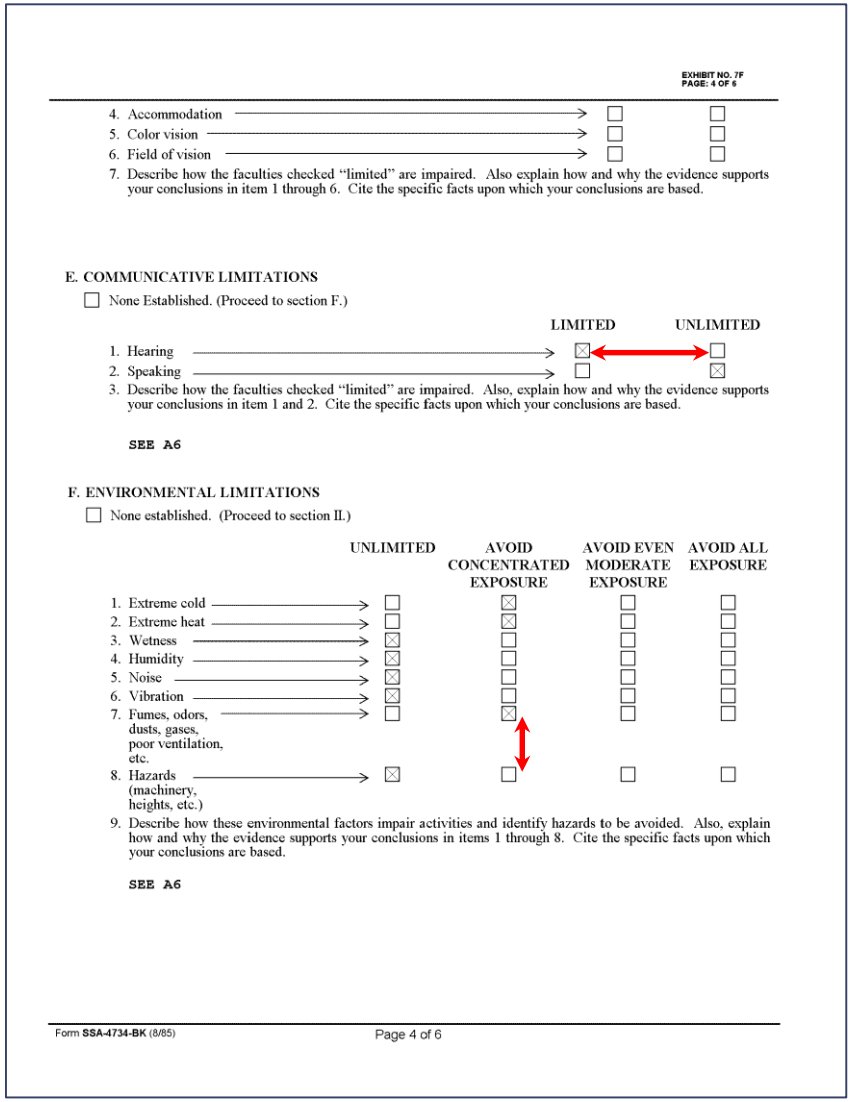


EXHIBIT NO. 7F
PAGE: 4 OF 6

4. Accommodation	☐	☐
5. Color vision	☐	☐
6. Field of vision	☐	☐

7. Describe how the faculties checked "limited" are impaired. Also explain how and why the evidence supports your conclusions in item 1 through 6. Cite the specific facts upon which your conclusions are based.

E. COMMUNICATIVE LIMITATIONS

☐ None Established. (Proceed to section F.)

	LIMITED	UNLIMITED
1. Hearing	☒	☐
2. Speaking	☐	☒

3. Describe how the faculties checked "limited" are impaired. Also, explain how and why the evidence supports your conclusions in item 1 and 2. Cite the specific facts upon which your conclusions are based.

SEE A6

F. ENVIRONMENTAL LIMITATIONS

☐ None established. (Proceed to section II.)

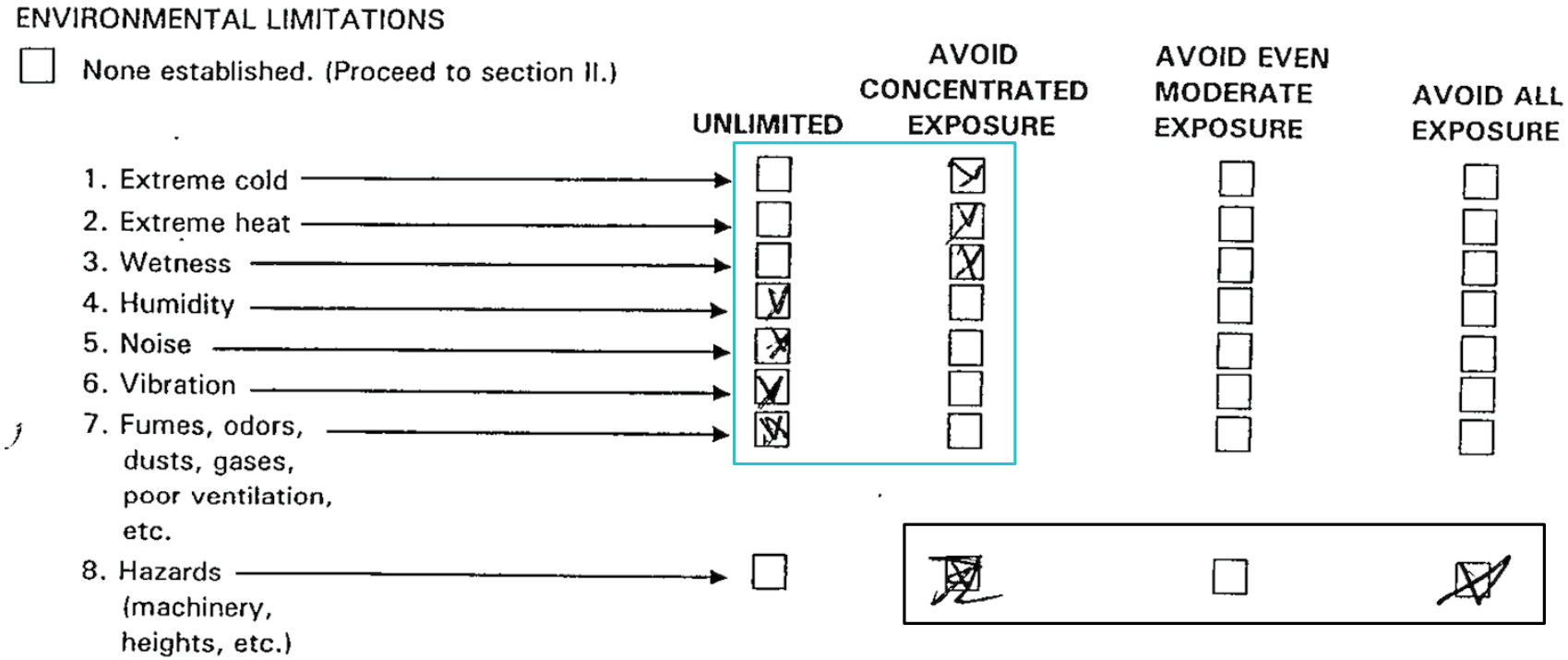
	UNLIMITED	AVOID CONCENTRATED EXPOSURE	AVOID EVEN MODERATE EXPOSURE	AVOID ALL EXPOSURE
1. Extreme cold	☐	☒	☐	☐
2. Extreme heat	☐	☒	☐	☐
3. Wetness	☒	☐	☐	☐
4. Humidity	☒	☐	☐	☐
5. Noise	☒	☐	☐	☐
6. Vibration	☒	☐	☐	☐
7. Fumes, odors, dusts, gases, poor ventilation, etc.	☐	☒	☐	☐
8. Hazards (machinery, heights, etc.)	☒	☐	☐	☐

9. Describe how these environmental factors impair activities and identify hazards to be avoided. Also, explain how and why the evidence supports your conclusions in items 1 through 8. Cite the specific facts upon which your conclusions are based.

SEE A6

Form SSA-4734-BK (8/85) Page 4 of 6

Handwriting



Methods

Automatic Data Extraction

Steps:

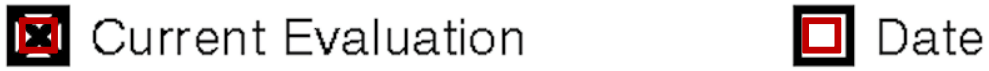
- Checkbox Detection
- Checkbox Matching
  - Templates
  - Template Matching Algorithm
- Record Output

Checkbox Detection

Use python's OpenCV to detect checkboxes based on size and shape


```
#Invert image (swap black and white pixels)
ret,th1 = cv2.threshold(frame,127,255,cv2.THRESH_BINARY_INV)
#Find all lines and shapes (contours) in the image
contours, hierarchy = cv2.findContours(th1,cv2.RETR_LIST,cv2.CHAIN_APPROX_SIMPLE)
#Only keep contours that match the area of CBs
cnt_area = [c for c in contours if 675 <= cv2.contourArea(c) <= 1285]
#Only keep contours that match the shape of CBs
cnt_circ = [c for c in cnt_area if 0.75 <= (4*np.pi*cv2.contourArea(c))/cv2.arcLength(c,True)**2 <= 0.8]
```

Ratio of black and white pixels at center of checkbox indicates marked checkboxes




```
non_zero = cv2.countNonZero(rect)
if non_zero > 0.1*nw*nh:
    #if CB is marked, keep track of it's coordinates
    page_mb.append([x,y,w,h])
    mb_i += 1
```

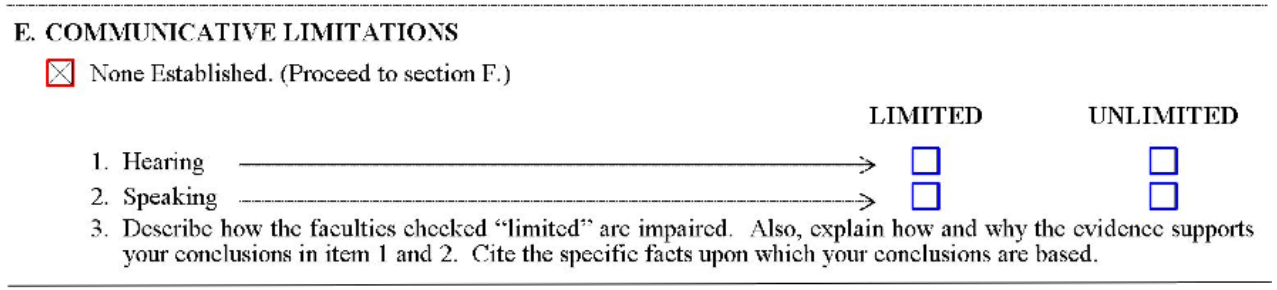
Checkbox Matching

Checkbox Position:

- Euclidean Coordinates
  - (x_i, y_i, p_i)
- Row-Column Coordinates (RCC)
  - (r_i, c_i)

Checkbox Alignment:

- $|x_i - x_j| < e_c \Rightarrow c_i = c_j$
- $|y_i - y_j| < e_r \Rightarrow r_i = r_j$



E. COMMUNICATIVE LIMITATIONS

☒ None Established. (Proceed to section F.)

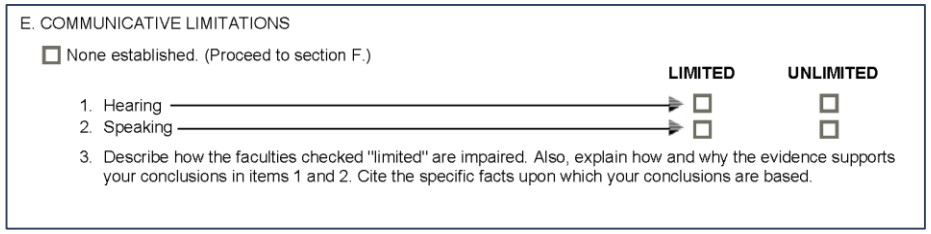
	LIMITED	UNLIMITED
1. Hearing	☐	☐
2. Speaking	☐	☐

3. Describe how the faculties checked "limited" are impaired. Also, explain how and why the evidence supports your conclusions in item 1 and 2. Cite the specific facts upon which your conclusions are based.

Euclidean Coordinates:
[(115, 202, 4), (1171, 310, 4), (1461, 310, 4), (1171, 353, 4), (1461, 353, 4)]

Row-Column Coordinates:
[(1, 1), (2, 2), (2, 3), (3, 2), (3, 3)]

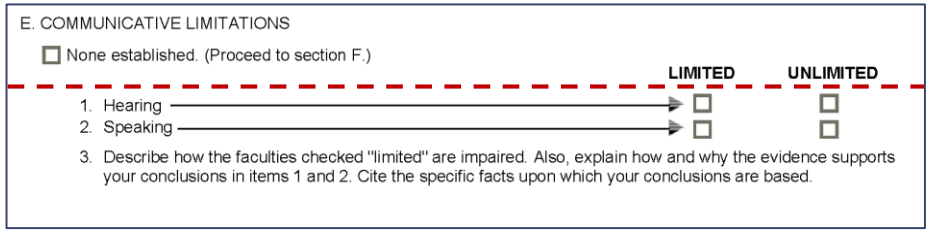
Section Break Row-Column Coordinates



RCC when no break occurs:

Before: [(1,1), (2,2), (2,3), (3,2), (3,3)]

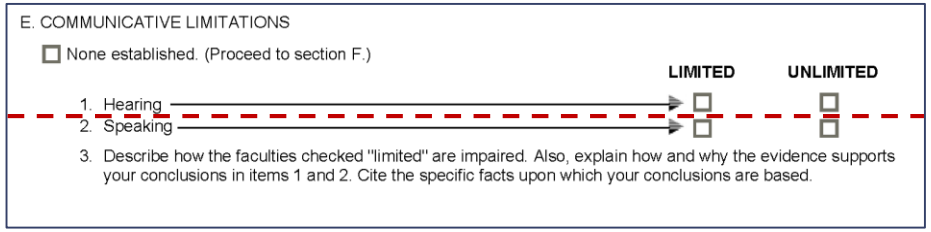
After: {}



RCC when break occurs after 1st row:

Before: [(1,1)]

After: [(1,1), (1,2), (2,1), (2,2)]



RCC when break occurs after 2nd row:

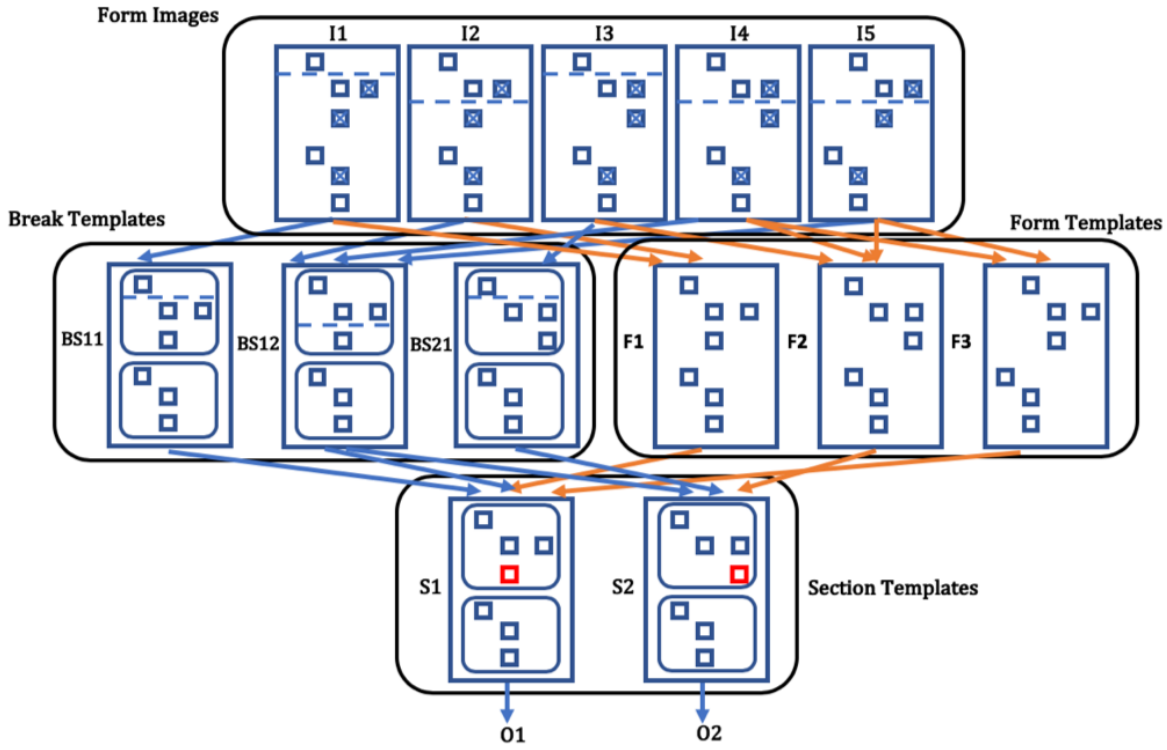
Before: [(1,1), (2,2), (2,3)]

After: [(1,1), (1,2)]

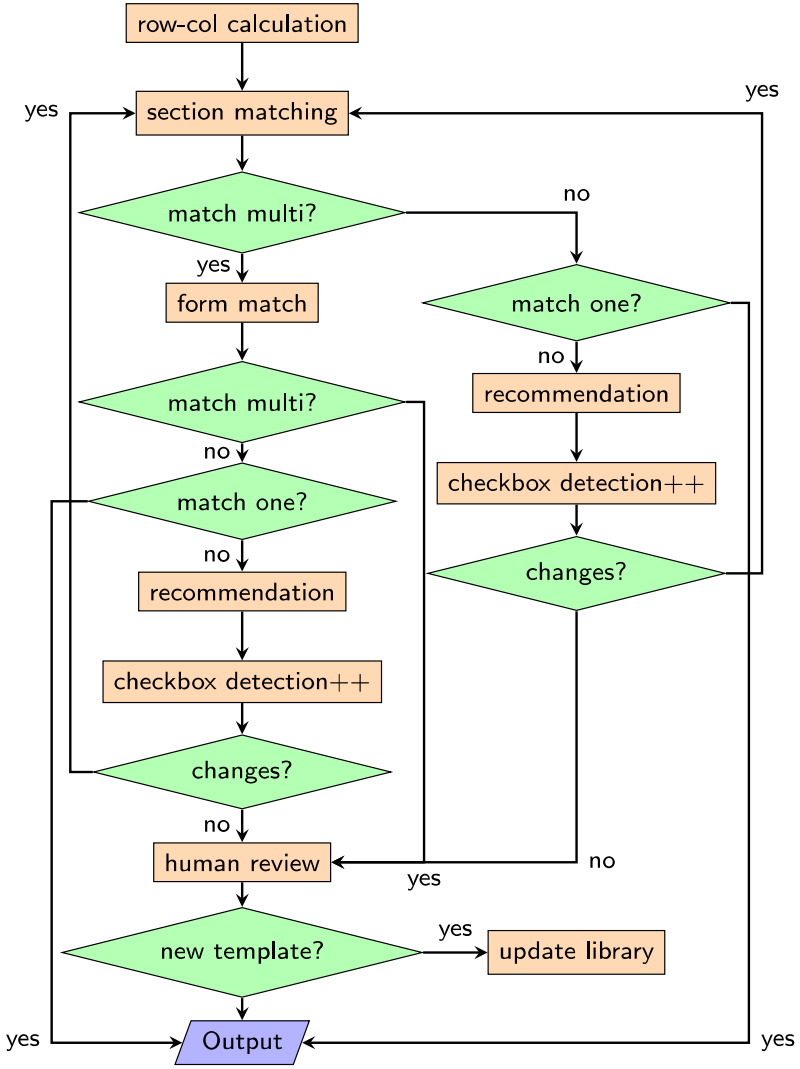
Templates

3 Types of Templates:

- Section Template T_S
  - Simplest type of template
  - Combined with other sections to match form
- Form Template T_F
  - Consider entire form F to be one section S
  - Reduces ambiguity across sections
- Break Template T_{SK}
  - Encodes all possible section breaks

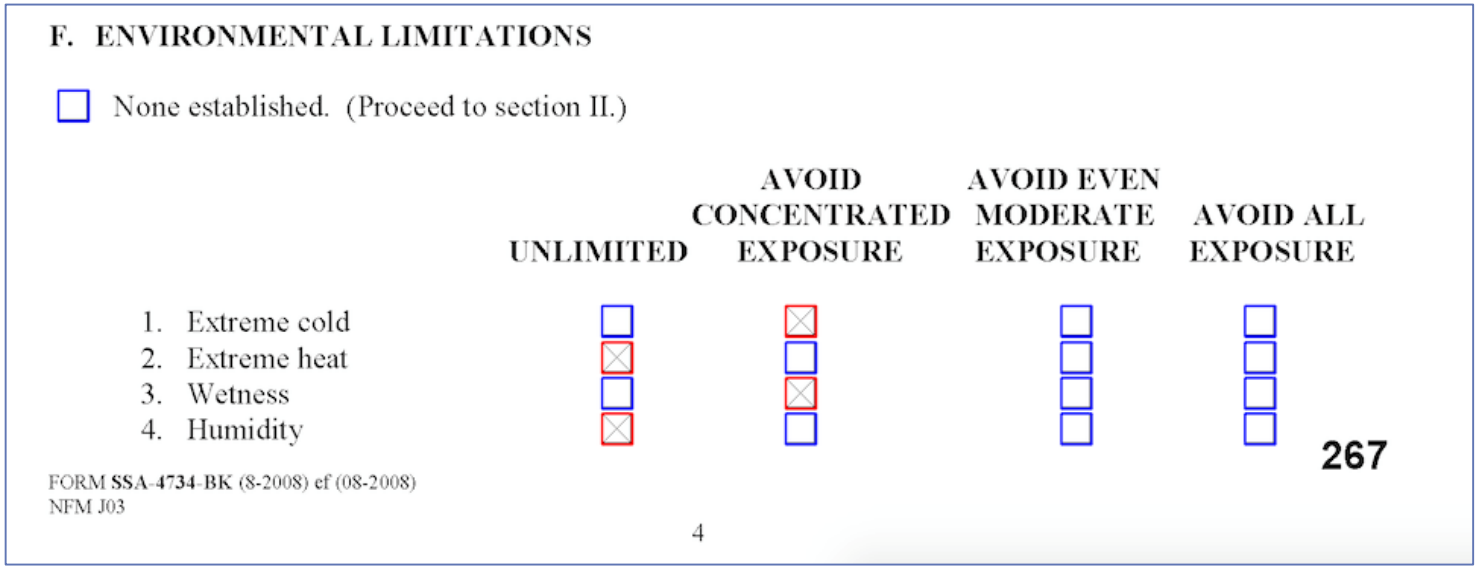


Template Matching Algorithm



Record Output

SAMPLE.tif:



F. ENVIRONMENTAL LIMITATIONS

☐ None established. (Proceed to section II.)

	UNLIMITED	AVOID CONCENTRATED EXPOSURE	AVOID EVEN MODERATE EXPOSURE	AVOID ALL EXPOSURE
1. Extreme cold	☐	☒	☐	☐
2. Extreme heat	☒	☐	☐	☐
3. Wetness	☐	☒	☐	☐
4. Humidity	☒	☐	☐	☐

267

FORM SSA-4734-BK (8-2008) ef (08-2008)
NFM J03

4

File Name	Environmental Limitations	Extreme Cold	Extreme Heat	Wetness	Humidity
SAMPLE		Avoid Concentrated	Unlimited	Avoid Concentrated	Unlimited

Tasks

TASK	PURPOSE	PHYSICAL RFCs*	MENTAL RFCs*
Validation	Evaluate templates and matching algorithm performance against original form images	10000	5000
Comparison	Evaluate template matching (RCC) against location matching (Euclidean)	4914	2364
Sample Generation	Perform data entry for entire sample	497646	98408

*Refers to number of images in sample

Results

Performance Metrics

Performance across 3 tasks for Physical RFC (PRFC) and Mental RFC (MRFC)

	Validation		Comparison		Full	
	PRFC	MRFC	PRFC	MRFC	PRFC	MRFC
Images	10000	5000	4914	2364	497646	98408
RFC Forms	9128	4281	4862	2342	448907	85268
Matched	8276	4167	4739	2278	410195	81391
Correct	8207	4151	4714	2260	406831	81025
Recall (%)	0.907	0.973	0.975	0.973	0.914	0.955
Precision (%)	0.992	0.996	0.995	0.992	0.992	0.996
F_1 (%)	0.947	0.985	0.985	0.982	0.951	0.975

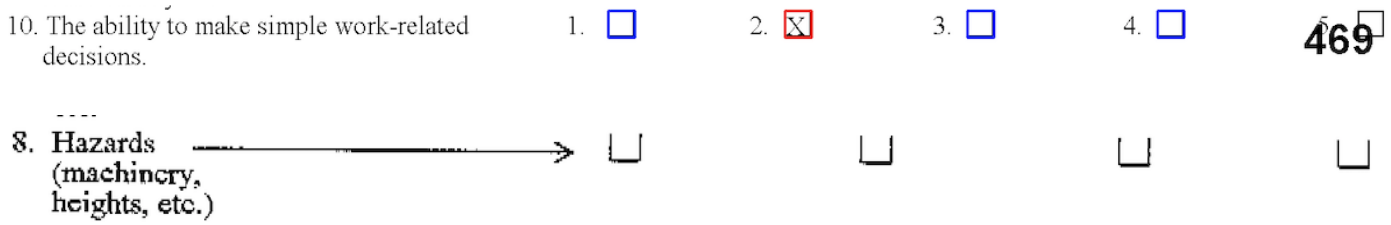
Comparison of Template vs. Location Matching

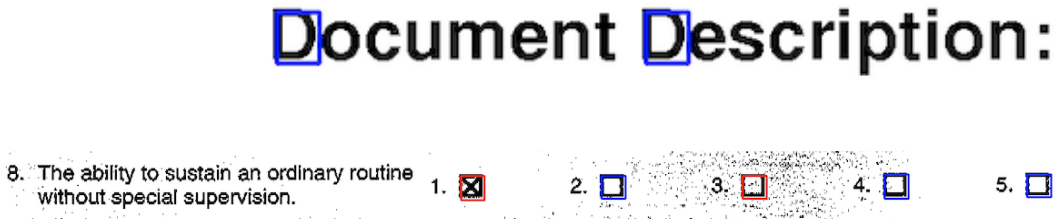
	Physical	Mental
RFC Forms	4862	2342
Template Matched	4739	2278
Location Matched	4662	2272
Both Matched	4621	2258
Agree	4381	2042
Agreement (%)	94.81	90.43

Error Analysis

Recall Errors:

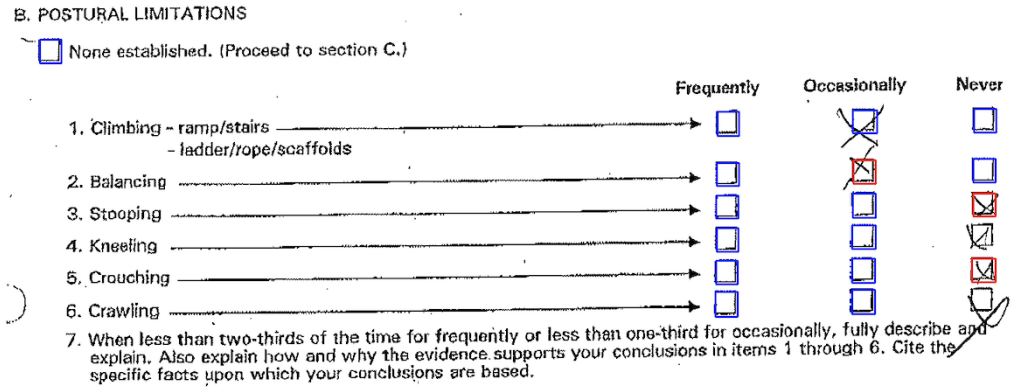
- Missed checkboxes
  - Image interference
  - Scan noise
  - Handwriting
- False positives





Precision Errors:

- Checkboxes appear marked when not
  - Image interference/Scan noise
- Checkboxes not marked in center
  - Handwriting



Next Steps

Checkbox Identification:

- Train models to identify checkboxes
  - Deep learning models

Checkbox Matching:

- Add automation to template generation
  - Learn to identify column/row headings

Generalization:

- Apply methods to other data
  - Checkboxes in medical records

Conclusion

Successfully used novel templates to extract checkbox data

Good performance comes from specificity of task and strong assumptions

- Grid-like structure of checkboxes
- No ambiguity in forms

Able to achieve good performance with basic computer vision

- Necessitated based on limited computing resources
- Errors came from missing checkboxes (handwriting, scan noise, etc.)
- More advanced methods (e.g., deep learning) could help improve checkbox identification or may be necessary for other applications (e.g., medical records)

Thank you!
Questions?

Contact Information: julia.porcino@nih.gov