Reminder/Recall Text Replies: Using Machine Learning for Focused Follow-up

Kaiyue Wu, MPH
New York City Department of Health and Mental Hygiene

AIRA 2026 National Meeting
Pittsburgh, PA
April 28, 2026

NYC Health

Citywide Immunization Registry (CIR)

Citywide Immunization Registry (CIR) is the Immunization Information System (IIS) for New York City (NYC)

- Began citywide in 1997
- As of March 2026, CIR contains
  - > 15.4 million patients
  - > 188 million immunizations
- Mandatory reporting of immunizations for children < 19 years
  - Reporting for adults ≥ 19 years requires consent
- Population-based

NYC Health

Reminder/Recall (R/R)

- Reminder/Recall (R/R) is the process of identifying and notifying people who are:

due soon for an immunization (reminder)

OR

overdue for an immunization (recall)

R/R in NYC

- **Provider-based R/R** is available through the CIR's Online Registry (OR) application at no cost to providers
  - 2010: Create patient lists, letters, and mailing labels for R/R
  - 2015: Text messaging R/R became available

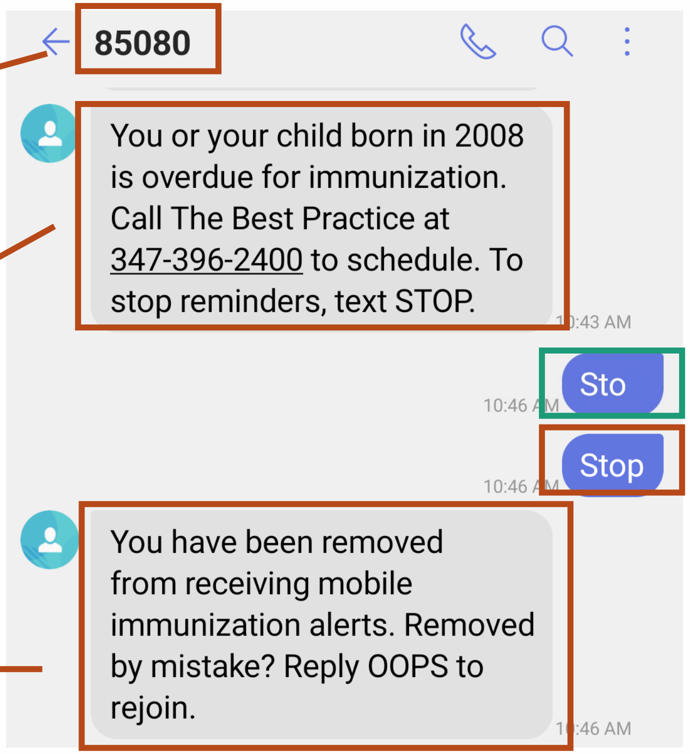
- **Centralized R/R** delivers messages directly from NYC Health Department to a large number of patients based on demographic and immunization information in the CIR, regardless of their medical care

R/R Text Messages

Short Code 85080 is the number that recipients will see when you text them

Example of a reminder/recall message

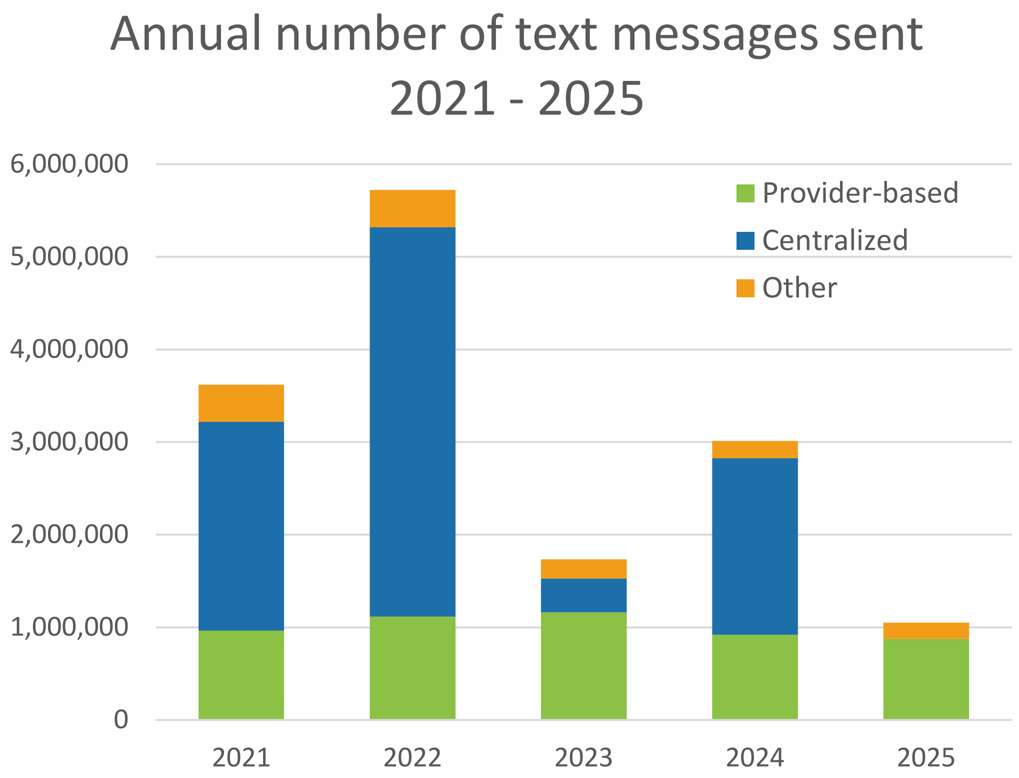
Auto-reply confirming patient has opted out of receiving text messages after replying “STOP”. Patient can opt back in to receive text messages by replying, “OOPS”.

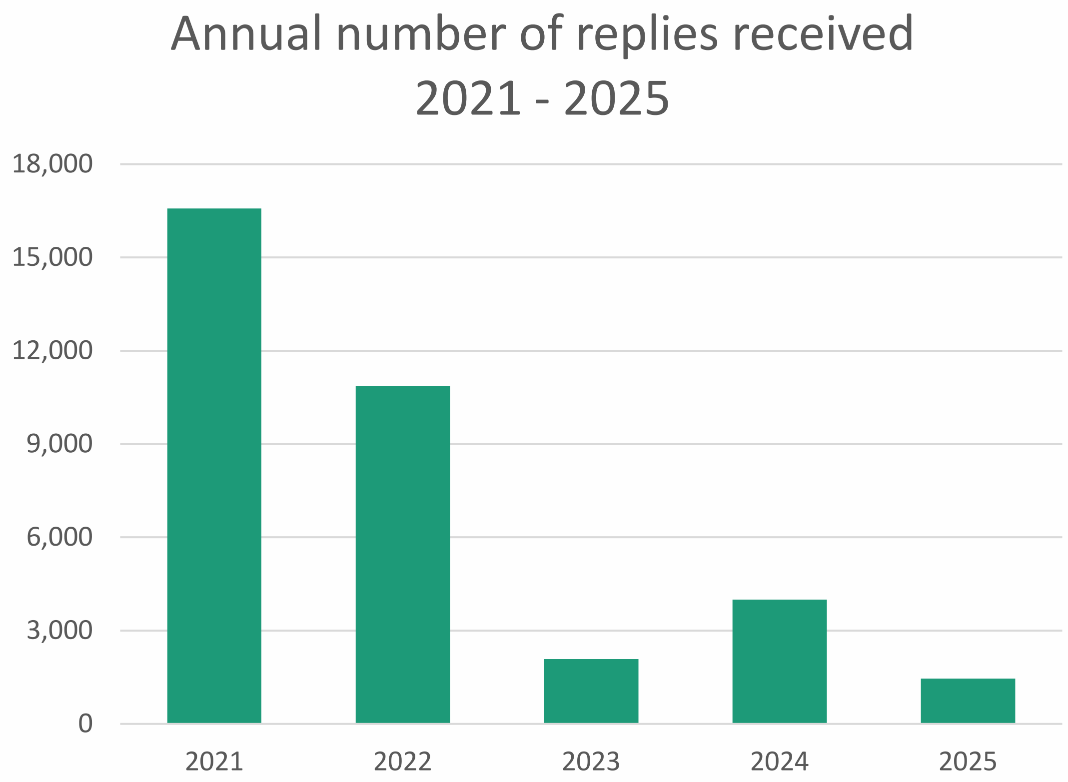


Example of a text message reply that doesn’t match a keyword

CIR text messaging is defaulted as opt-in. The recipient always has the option of continuing to receive messages or stopping them by texting “STOP”

R/R Text Messaging Monitoring and Evaluation





- Although most of the recipients do not respond to the text messages, the volume of replies is high

Examples of Replies to R/R Text Messages

No inglés, español.

eggs, milk, bread

Can we make an appointment?

Already got it

Wrong number.

Thank you

We have an appointment next week.

I'm Driving.
-Sent from my Car

No Thanks

Stoop

We moved out of NYC

Hola

qwertyuiop

onglay imetay onay eesay

Oops

å¥½çš„è°¢è°¢

NYC Health

Reply-Driven Follow-up

- No direct two-way communication to senders
- CIR staff can view replies in vendor dashboard and determine if follow-up is needed for certain cases, for example:
  - Person has moved out of NYC -> Update patient active status in CIR
  - Incorrect phone number -> Remove phone number from CIR
  - Threatening message -> Opt-out phone number
- Automating text reply categorization can assist in triaging and prioritizing follow-up activities

Objectives

- Build a supervised machine learning model to classify replies to R/R text messages
  - Identify text replies that require follow-ups
  - Optimize workflow with reduced manual review efforts
- Prioritize actions that can improve CIR data quality and de-escalate sensitive communications
- Be more prepared to respond during emergencies

Methods

1. Define action-oriented reply message categories using an inductive, consensus-based coding approach by five CIR reviewers
2. Manually review and label text message replies from 2023-2025 (N = 7366)
3. Develop a text classifier using Multinomial Naïve Bayes as a baseline model with the *scikit-learn*[1] package in Python
4. Explore model refinement based on observed patterns and evaluate alternative models for comparison
5. Select the best-performing model for future automated classification

1. Pedregosa, Fabian, et al. 2011."Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research 12: 2825–2830.

Health

Reply Categories

Category	Example	Action	Priority
Legal issue	“I will sue you”	Notify supervisor and potential opt-out	High
Deceased patient	“He passed away…”	Potential removal of the mobile phone number from CIR	
Foul language	“Go to hell”	Potential opt-out	Medium
Possible record correction children	“I don’t have a child”	Review CIR record	
Possible record correction other	“I already got the flu shot”	Review CIR record	
Relocation	“No longer in NY”	Potential removal of the mobile phone number from CIR	
Seeking more info	“What time can I go?”	Potential follow up	
Other relevant texts	“I have an appointment for it tmr”	Potential follow up	
Non-textable phone number	“This phone cannot receive texts.”	Potential follow up	
Simple acknowledgement	“OK”	None	Low
Other auto replies	“Driving. Can’t text.”		
Irrelevant/Meaningless texts	“qwerty”		

Reply Categories

Category	Example	Action	Priority
Legal issue	“I will sue you”	Notify supervisor and potential opt-out	High
Deceased patient	“He passed away…”	Potential removal of the mobile phone number from CIR	
Foul language	“Go to hell”	Potential opt-out	Medium
Possible record correction children	“I don’t have a child”	Review CIR record	
Possible record correction other	“I already got the flu shot”	Review CIR record	
Relocation	“No longer in NY”	Potential removal of the mobile phone number from CIR	
Seeking more info	“What time can I go?”	Potential follow up	
Other relevant texts	“I have an appointment for it tmr”	Potential follow up	
Non-textable phone number	“This phone cannot receive texts.”	Potential follow up	
Simple acknowledgement	“OK”	None	Low
Other auto replies	“Driving. Can’t text.”		
Irrelevant/Meaningless texts	“qwerty”		

Reply Categories

Category	Example	Action	Priority
Legal issue	"I will sue you"	Notify supervisor and potential opt-out	High
Deceased patient	"He passed away…"	Potential removal of the mobile phone number from CIR	
Foul language	"Go to hell"	Potential opt-out	Medium
Possible record correction children	"I don't have a child"	Review CIR record	
Possible record correction other	"I already got the flu shot"	Review CIR record	
Relocation	"No longer in NY"	Potential removal of the mobile phone number from CIR	
Seeking more info	"What time can I go?"	Potential follow up	
Other relevant texts	"I have an appointment for it tmr"	Potential follow up	
Non-textable phone number	"This phone cannot receive texts."	Potential follow up	
Simple acknowledgement	"OK"	None	Low
Other auto replies	"Driving. Can't text."		
Irrelevant/Meaningless texts	"qwerty"		

Reply Categories

Category	Example	Action	Priority
Legal issue	“I will sue you”	Notify supervisor and potential opt-out	High
Deceased patient	“He passed away…”	Potential removal of the mobile phone number from CIR	
Foul language	“Go to hell”	Potential opt-out	Medium
Possible record correction children	“I don’t have a child”	Review CIR record	
Possible record correction other	“I already got the flu shot”	Review CIR record	
Relocation	“No longer in NY”	Potential removal of the mobile phone number from CIR	
Seeking more info	“What time can I go?”	Potential follow up	
Other relevant texts	“I have an appointment for it tmr”	Potential follow up	
Non-textable phone number	“This phone cannot receive texts.”	Potential follow up	
Simple acknowledgement	“OK”	None	Low
Other auto replies	“Driving. Can’t text.”		
Irrelevant/Meaningless texts	“qwerty”		

Baseline Model

Baseline model: Multinomial Naïve Bayes with word-level Term Frequency–Inverse Document Frequency (TF-IDF)

Patterns

1. Misspelling
2. Short messages
3. Bad encoding
4. Rare but important categories
5. Non-English texts

Solutions

1. Alternative TF-IDF approach
2. Text meta features
3. Character repair
4. Rule-based keyword matching
5. Translation

Health

Model Refinement (1)

1. Alternative TF-IDF approach
    - Word-level TF-IDF with bigrams
        - Extracts words and consecutive word pairs
        - Example: I got vaccination
    - Character-level TF-IDF
        - Extracts sequences of 2-5 characters
        - Example: vaccination

NYC Health

Model Refinement (2)

2. Text meta features
    - Character count
    - Word count
    - Average word length
    - Uppercase ratio
    - Punctuation ratio
    - Non-ASCII ratio
    - Flags for all-caps, very short, single-word, and multiline messages

Model Refinement (3)

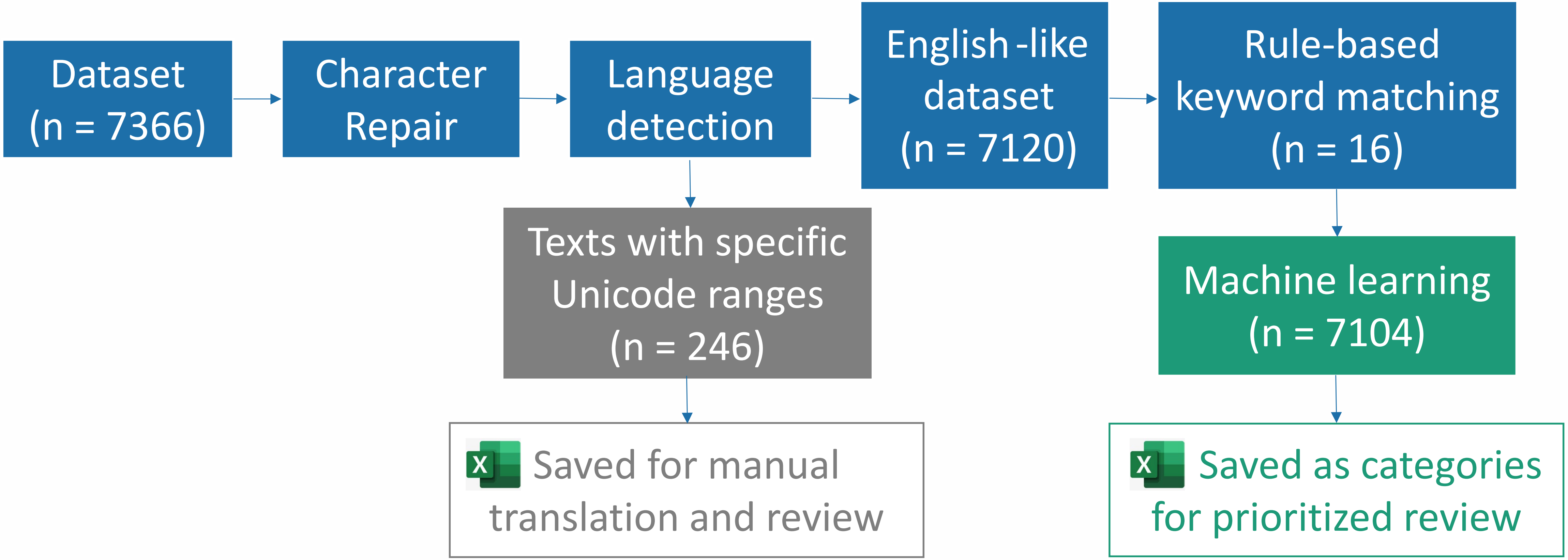
3. Character repair
    - Detect and fix double-encoded UTF-8 text to restore correct characters
        - Example: “å¥½çš„è°¢è°¢” --> “好的，谢谢” --> “OK, thank you”
4. Rule-based keyword matching
    - Apply keyword rules to automatically assign labels for categories with insufficient training data
        - Example: “sue”, “lawyer” -> legal issues
        - Example: “passed away”, “dead” -> deceased patients

Alternative Models

After refinement, the baseline model was compared to five alternative models:

- Linear SVC (SVC)
- Logistic Regression (LR)
- SGD Classifier (SGD)
- Complement Naïve Bayes (CNB)
- Random Forest (RF)

Workflow



Dataset (n = 7366)
Character Repair
Language detection
English-like dataset (n = 7120)
Rule-based keyword matching (n = 16)
Texts with specific Unicode ranges (n = 246)
Machine learning (n = 7104)
Saved for manual translation and review
Saved as categories for prioritized review

Results – Model Comparison

```
ML Train: 5683, ML Test: 1421

>>> BASELINE: MultinomialNB + word TF-IDF
    Accuracy: 0.751

>>> SVC   Accuracy: 0.861
>>> LR    Accuracy: 0.844
>>> SGD   Accuracy: 0.640
>>> CNB   Accuracy: 0.829
>>> RF    Accuracy: 0.852
```

- Baseline Multinomial Naïve Bayes model had an accuracy of 75.1%

NYC Health

Results – Best Model with Linear SVC

```
==============================================================
BEST: SVC – Accuracy: 0.861
==============================================================
                                   precision    recall  f1-score   support

                  Foul language       0.74      0.52      0.61        27
              Meaningless texts       0.85      0.95      0.90       616
     Non-textable phone number        1.00      1.00      1.00       144
             Other auto replies       0.96      1.00      0.98        26
            Other relevant texts      0.40      0.22      0.29        76
Possible record correction children   0.75      0.63      0.69        38
   Possible record correction other   0.86      0.84      0.85        87
                     Relocation       1.00      0.36      0.53        14
              Seeking more info       0.68      0.67      0.68        76
         Simple acknowledgement       0.93      0.90      0.91       317

                       accuracy                           0.86      1421
                      macro avg       0.82      0.71      0.74      1421
                   weighted avg       0.85      0.86      0.85      1421

>>> 5-Fold Cross-Validation (SVC):
    Mean: 0.868 (+/- 0.007)
    Folds: ['0.870', '0.881', '0.863', '0.866', '0.863']
```

- Linear SVC achieved the highest accuracy at 86.1%, validated with 5-fold stratified cross-validation (86.8% ± 0.7%)

Results – Hybrid Evaluation

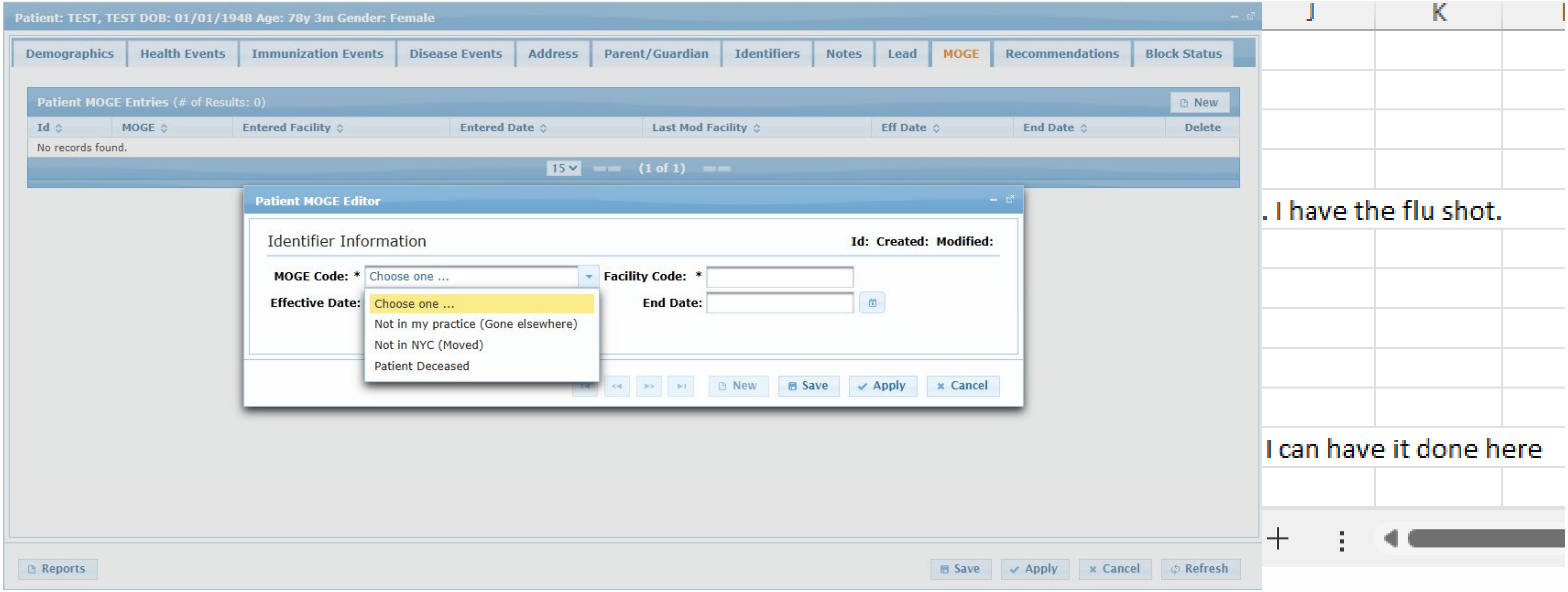
```
============================================================
HYBRID EVALUATION (Rule-based + Best ML on full test set)
============================================================
Hybrid Accuracy: 0.900
```

	precision	recall	f1-score	support
Deceased patient	0.25	1.00	0.40	1
Foul language	0.88	0.79	0.83	28
Legal issue	1.00	1.00	1.00	1
Meaningless texts	0.89	0.97	0.92	616
Non-textable phone number	1.00	1.00	1.00	144
Other auto replies	0.96	1.00	0.98	26
Other relevant texts	0.71	0.72	0.72	76
Possible record correction children	0.80	0.63	0.71	38
Possible record correction other	0.91	0.84	0.87	87
Relocation	1.00	0.36	0.53	14
Seeking more info	0.84	0.67	0.74	76
Simple acknowledgement	0.95	0.90	0.93	317
accuracy			0.90	1424
macro avg	0.85	0.82	0.80	1424
weighted avg	0.90	0.90	0.90	1424

- The combined hybrid pipeline achieved 90% accuracy
- Lower-priority categories can be filtered out with high accuracy, covering 77.5% of the test dataset

Health

Results – Export for Follow-up



Conclusions

- R/R text messages have been routinely used and play an important role in NYC vaccination campaign strategies
- The percentage of R/R text replies is low but the volume is high
- Linear SVC model shows the best performance for reply classification based on current dataset
- Hybrid model (rule-based + linear SVC) achieved an **90%** overall accuracy
- CIR staff can use the model to filter out categories with lower priorities
- Manual review can be potentially reduced by over **75%**

NYC Health

Limitations

- New tools require agency/IT approval
  - Language detection and translation
  - Pretrained language model
- Categories with infrequent replies or small sample sizes, are imbalanced for the model's analysis
- Model lacks ability to incorporate conversational context from multiple replies to a single message

Next steps

- Implement weekly automated text classification to generate spreadsheets of categorized replies for prioritized staff action
- Identify an IT-approved solution to enable translation functionality for non-English text analysis
- Enrich rare categories with synthetic examples
- Investigate new patterns that may emerge from replies

NYC Health

Acknowledgements

- Shirley Huie
- Emily Liang
- Minyi Ruan
- Iris Cheng

Thank you

Contact:

- Kaiyue Wu: kwu5@health.nyc.gov

- CIR website: www.nyc.gov/health/cir