Wolters Kluwer

APPLIED MEDICAL CODE MAPPING WITH CHARACTER-BASED DEEP LEARNING MODELS AND WORD-BASED LOGIC

Krishna Srihasam, Lead data scientist

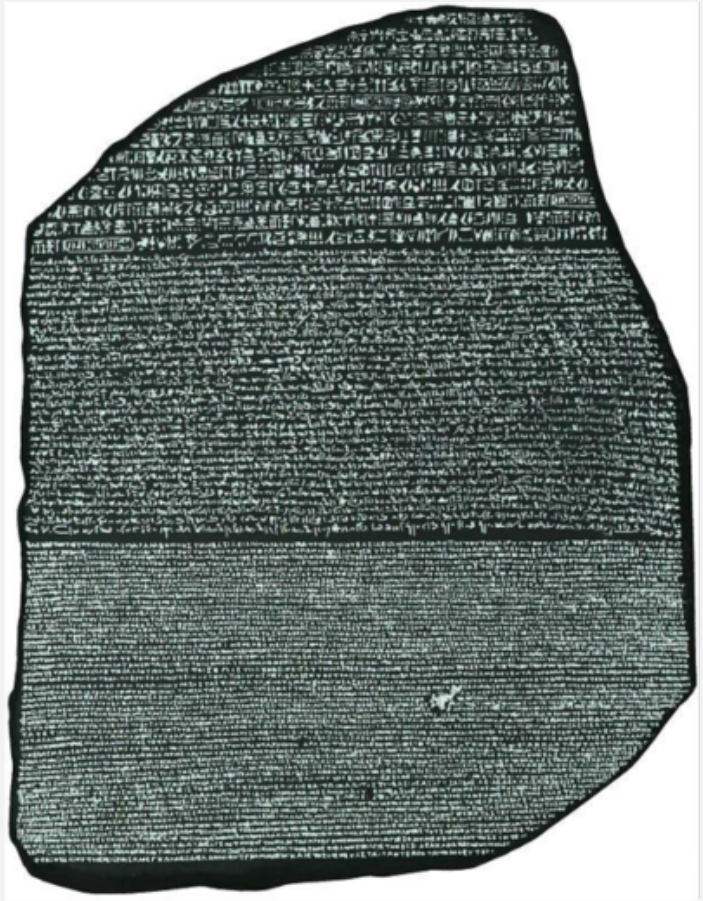
John Langton, Director of applied data science

Wolters Kluwer Health Data Science

Wolters Kluwer

MAPPING FROM ONE SET OF TERMS TO ANOTHER: A PROBLEM AS OLD AS TIME

Rosetta Stone



Phonetic	Ideographic
	
F	To carry
	
M	To Breath
	
D	To Grasp
	
O	People/Company

Wolters Kluwer

STANDARD MEDICAL CODES

- Medical codes are used to record, analyze, and communicate patient information for diagnosis, treatment, billing and more
- Healthcare providers often use custom codes that are not interoperable with other organizations' systems
- Mapping from custom codes to standard codes is necessary for information exchange and analytics

MULTIPLE USE CASES REQUIRE CODE MAPPINGS

Code Set	Description	Use Case
SNOMED	Clinical terminology polyarchy used primarily for diagnosis	Doctor records a particular disease condition for a patient in an EMR using a SNOMED Code for easy reference and searchability.
LOINC	Code set for medical laboratory observations	Doctor orders blood glucose test by sending a request to an outside lab and referencing the specific LOINC code for that test
RxNorm	Drug reference	Rules system operates on drug codes and fires an alert if two drugs that interact poorly are prescribed to the same patient

WHAT IS LOINC?

- LOINC is a common language (set of identifiers, names, and codes) for identifying health measurements, observations, and documents

- **Logical Observation Identifiers Names and Codes** (**LOINC**) is a database for identifying medical lab observations

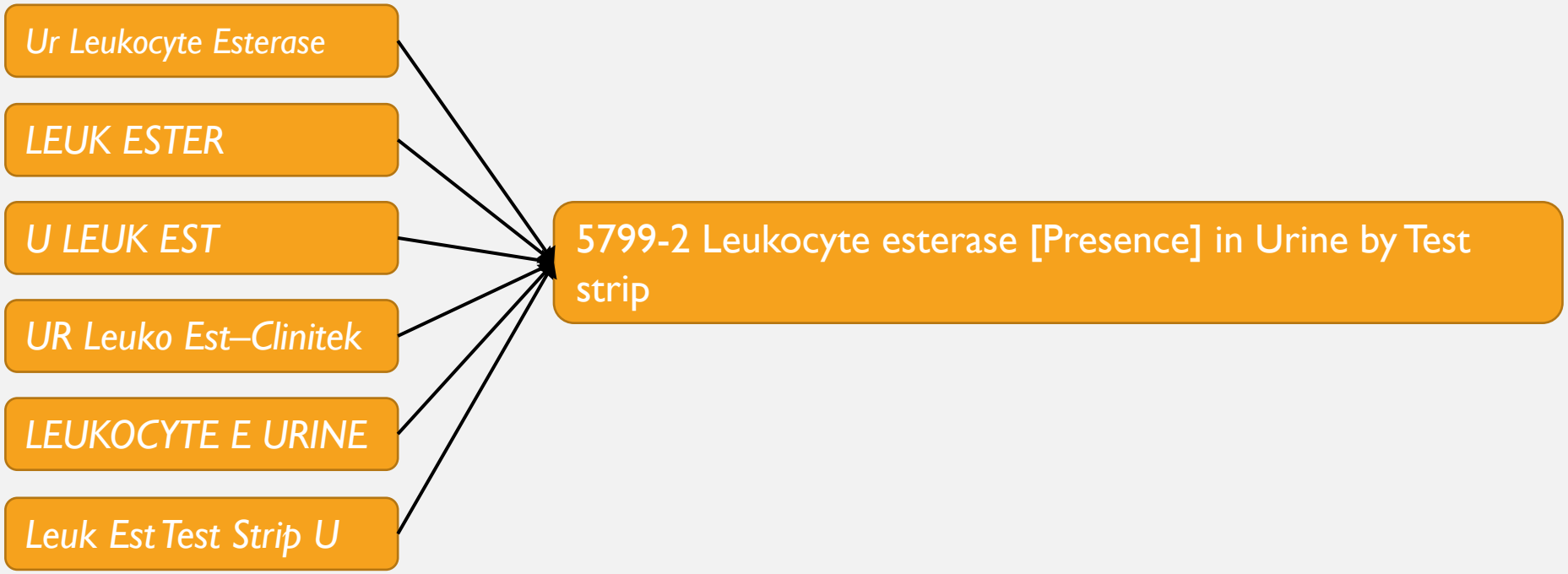
- Approximately 41000 LOINC codes exist

THE BIG PROBLEM

Local systems have different ways of identifying the same test or measurement. Fragmentation has created islands of isolation.

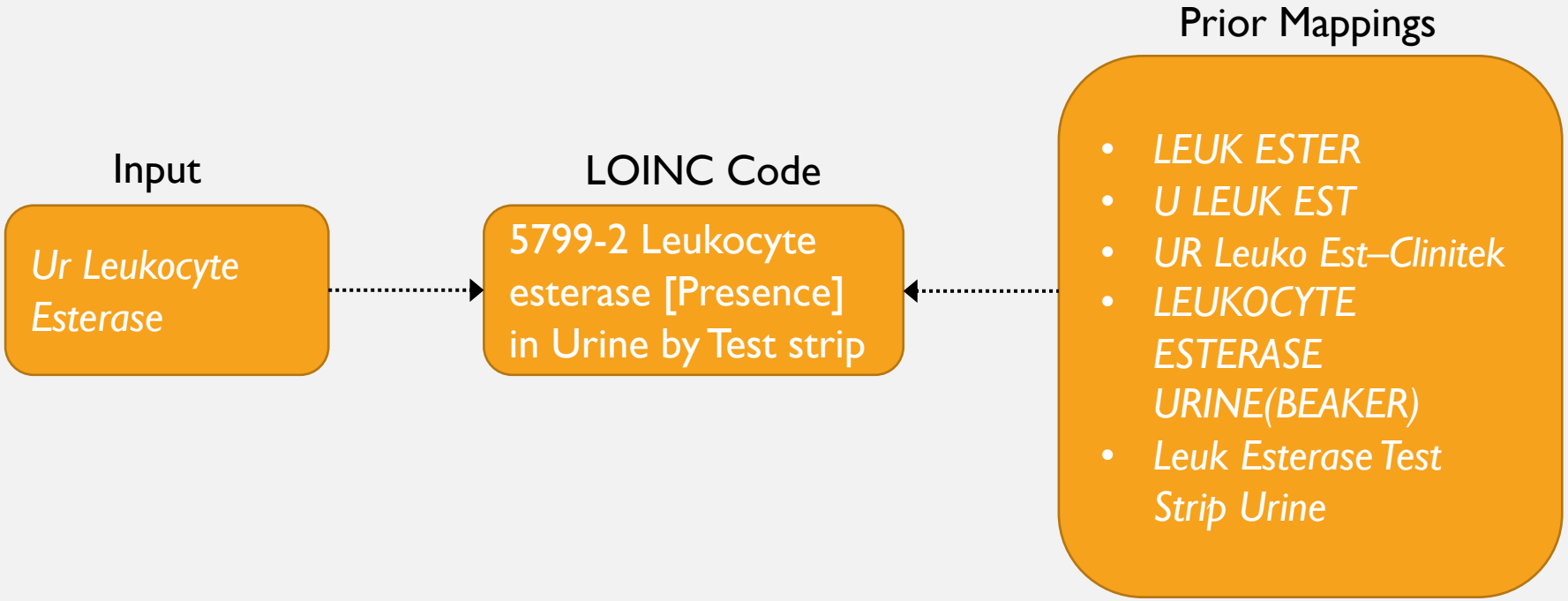
Map your local test codes to a LOINC for inter-operability

MAPPING TEST NAMES TO LOINC



LOINC MAPPING AS ML PROBLEM

Prior manual mappings served as labeled data for training



Wolters Kluwer

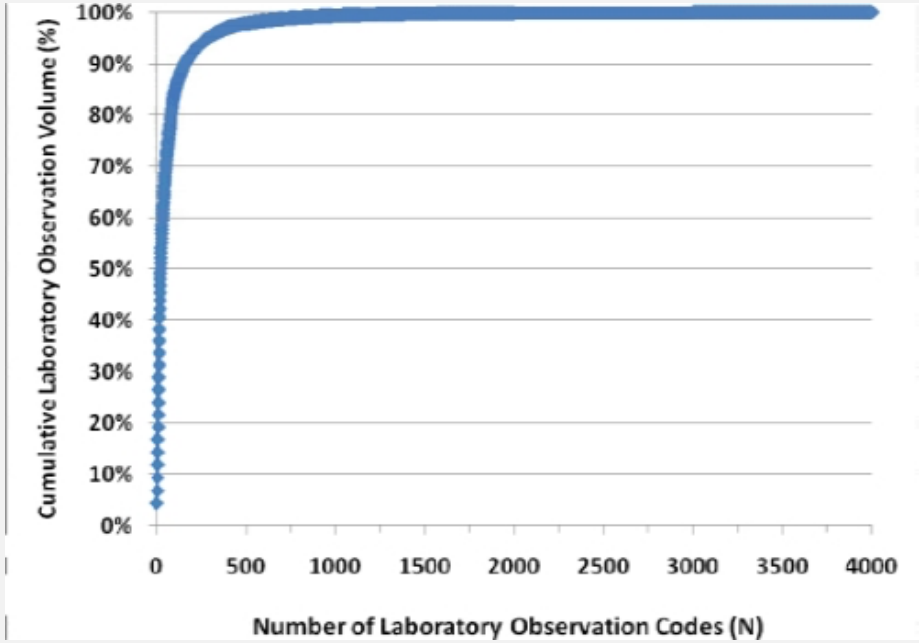
NOISY INPUTS

- Random acronyms leaving out letters: **Lkct vs *Leukocyte***
- Concatenating strings: ***UrLeukocyteEsterase*** *vs* ***Ur Leukocyte Esterase***
- Parsing and I/O errors: **00001**

CLASS IMBALANCE AND SAMPLE SCARCITY

2000 (5 %) codes account for 99% of all the tests

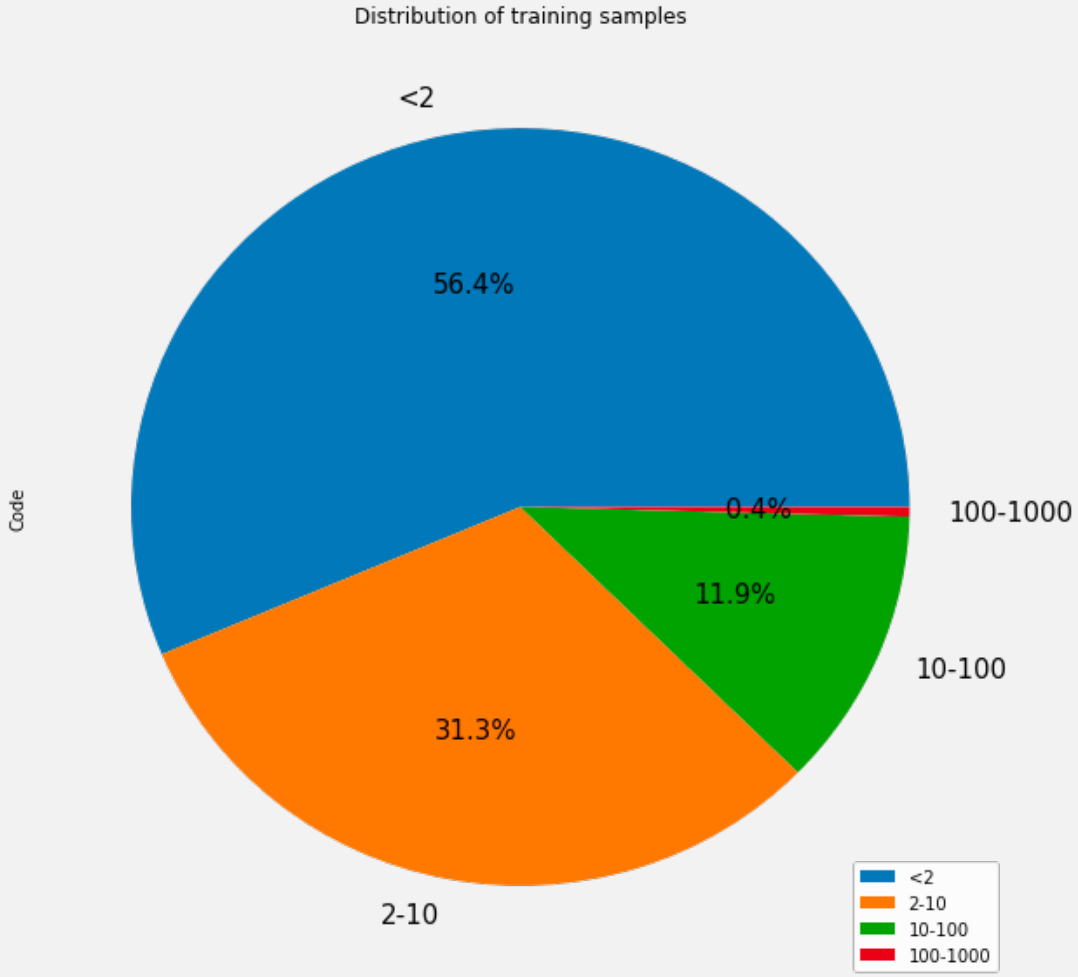
- 80 codes (2%) accounted for 80% of the total volume from all institutions
- 784 codes (19%) accounted for 99% of the volume from all institutions.



AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium. 2007 ;():771-5.

Wolters Kluwer

CLASS IMBALANCE AND SAMPLE SCARCITY



- Most common LOINC has 582 samples in the data
- Least common LOINC has only ONE sample in the data
- ~56% of LOINCs have only one sample in the data
- ~0.5% of LOINCs have more that 100 samples

Wolters Kluwer

MAPPING LOINC CODES

- Each LOINC Code has 6 distinct parts

Code: '5799-2'
Component: LEUKOCYTE ESTERASE
Specimen: URINE
Scale: ORDINAL
Timing: PT
Method: TEST STRIP
Property: Presence

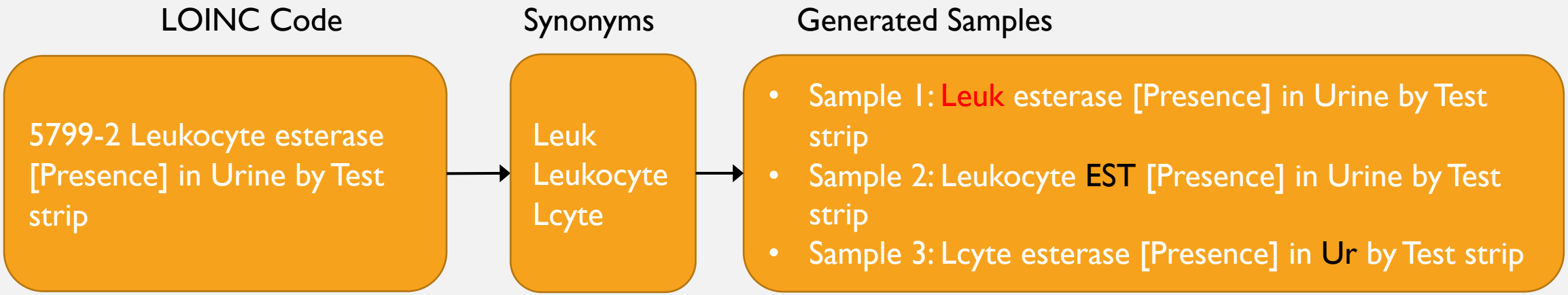
'Leukocyte esterase [Presence] in Urine by Test strip'

CLASS IMBALANCE AND SAMPLE SCARCITY

Name	Unique Values in LOINC Database	Unique Values in our training data	% Coverage
Component	19507	4783	25
System	344	143	42
Units	668	380	57
Method	504	212	42
Scale	6	6	100
Property	116	91	78
LOINC Codes	46156	11190	24

SAMPLE GENERATION VIA SYNONYM REPLACEMENT

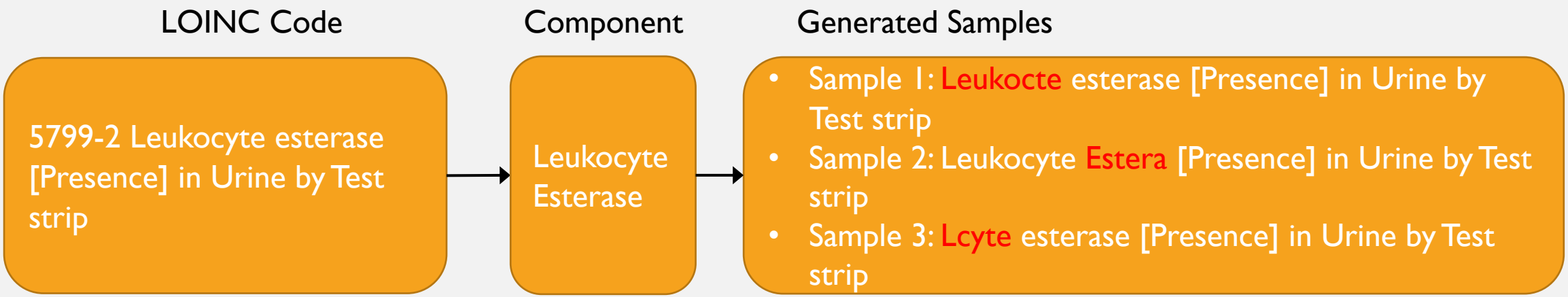
- Synonyms provided by prior mappings and medical informaticists
- Create new samples by replacing terms in code with known synonyms
- Multiple permutations possible



Wolters Kluwer

SAMPLE GENERATION VIA ADDING NOISE

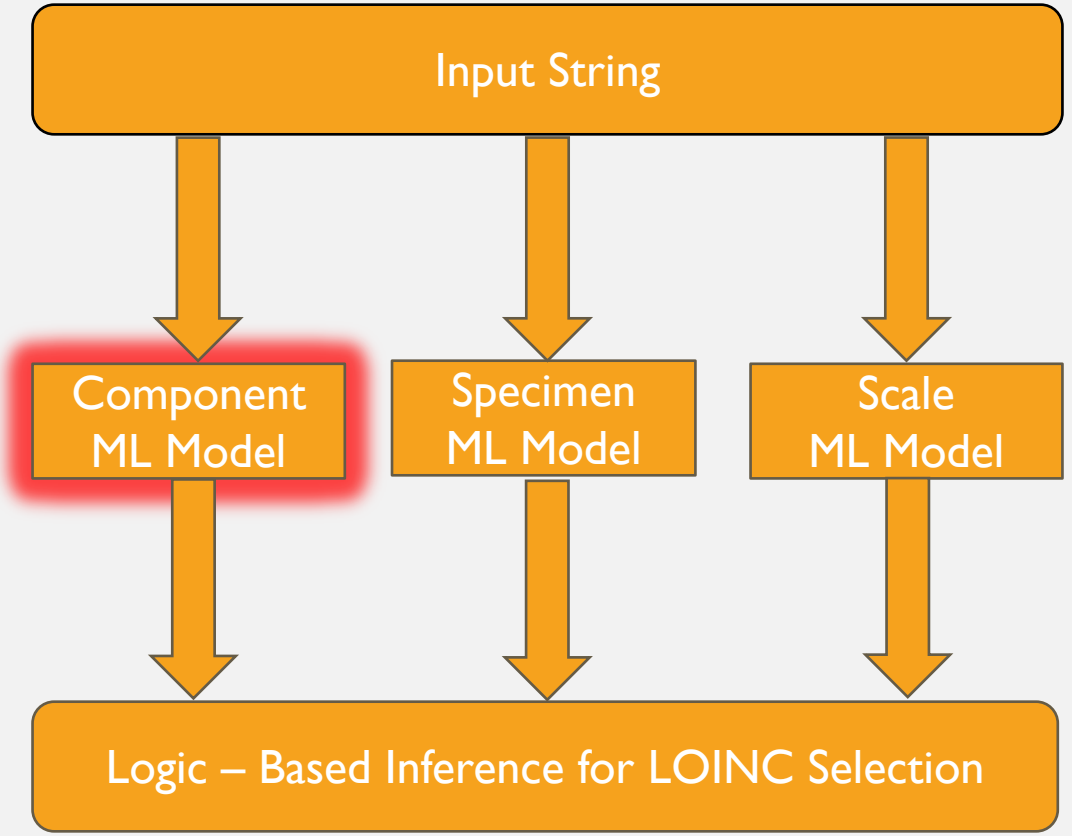
- Remove characters from the input string
- Add new characters to the input string
- Multiple permutations possible



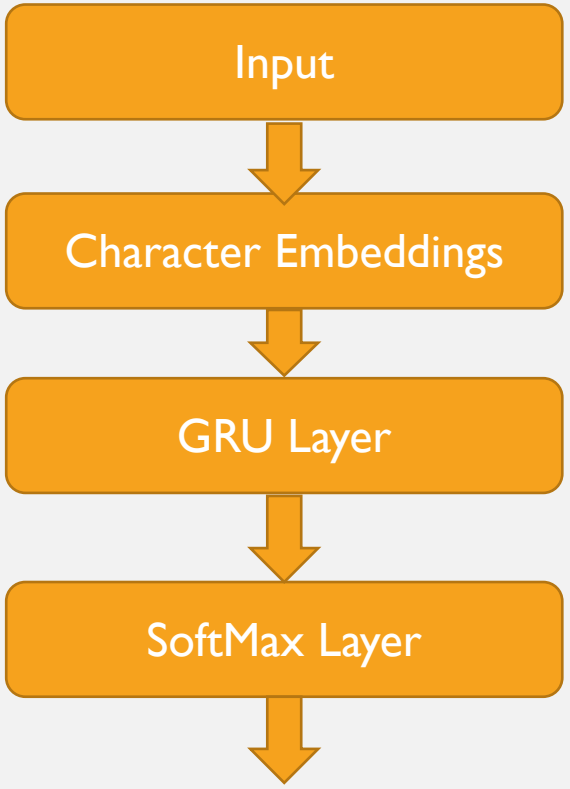
MAPPING APPROACH

- Around 3% of incoming strings could be mapped using standard ML techniques like fuzzy matching, word frequency analysis
- Character-based GRUs plateaued around 60% accuracy
- Hybrid Model using machine learning and logic based approaches achieved the best accuracy and coverage

HYBRID MODEL



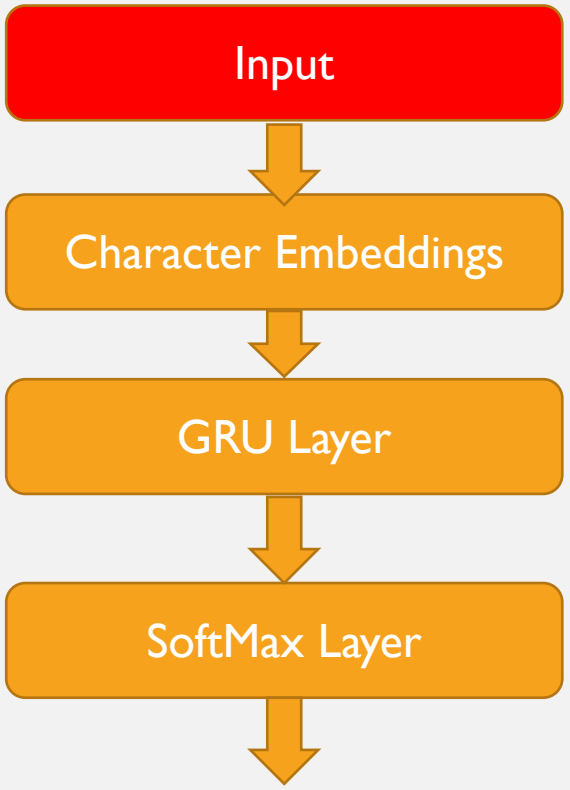
ML MODEL



- is a ***deep learning neural network model***.
- Learns to predict the output using character sequences instead of words
- Tokenize the data at character level and use information about character order
- Generate a character embedding vector
- Train a GRU on character embeddings to classify based on char sequence patterns
- Generates a prediction and a confidence score associated with this prediction

Wolters Kluwer

ML MODEL: INPUT

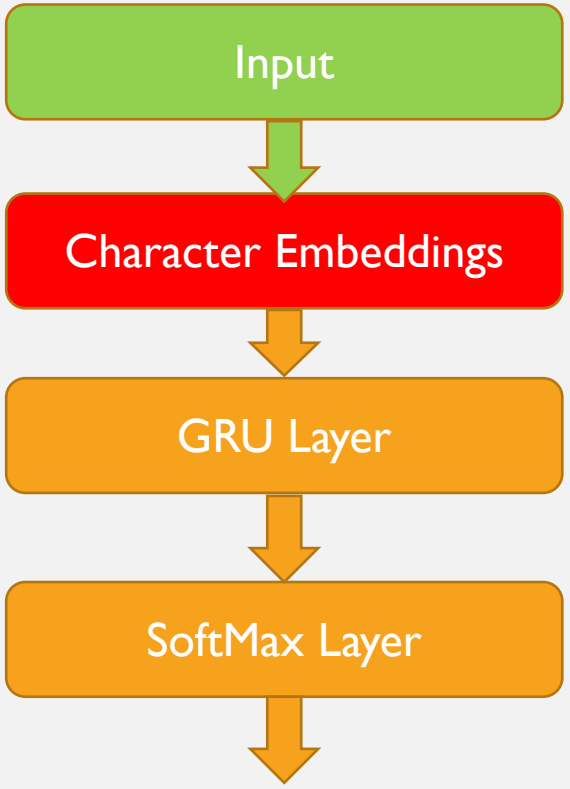


character	value
	0
(	1
)	2
a	3
b	4
c	5
e	6
i	7
k	8
l	9
n	10
o	11
p	12
r	13
s	14
t	15
u	16
y	17
-	18

'LEUK ESTER' : [9, 6, 16, 8, 0, 6, 14, 15, 6, 13]
'U LEUK EST' : [16, 0, 9, 6, 16, 8, 0, 6, 14, 15]
'Ur Leukocyte Esterase': [16, 13, 0, 9, 6, 16, 8, 11, 5, 17, 15, 6, 0, 6, 14, 15, 6, 13, 3, 14, 6]

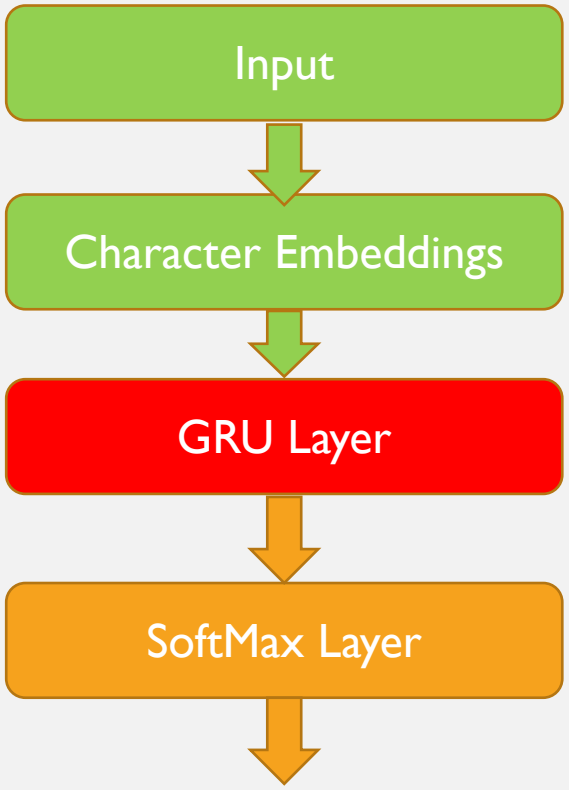
Wolters Kluwer

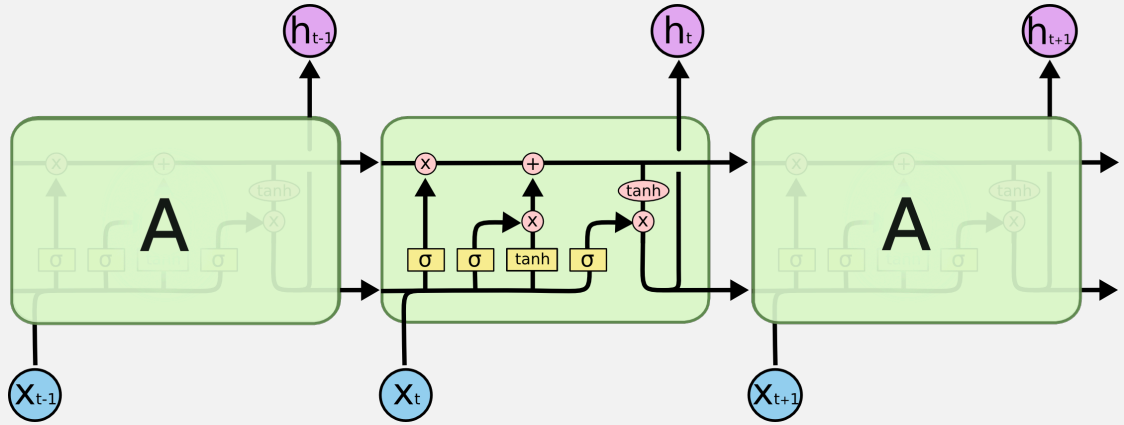
ML MODEL: CHARACTER EMBEDDING



- Neural network used for character embedding
- All vectors padded with 0s to meet max string length
- Each input string is a 2D array within a 3D array
- Fewer output nodes to input nodes for dimensionality reduction
- Similar to word2vec but at the character level

ML MODEL: GRU



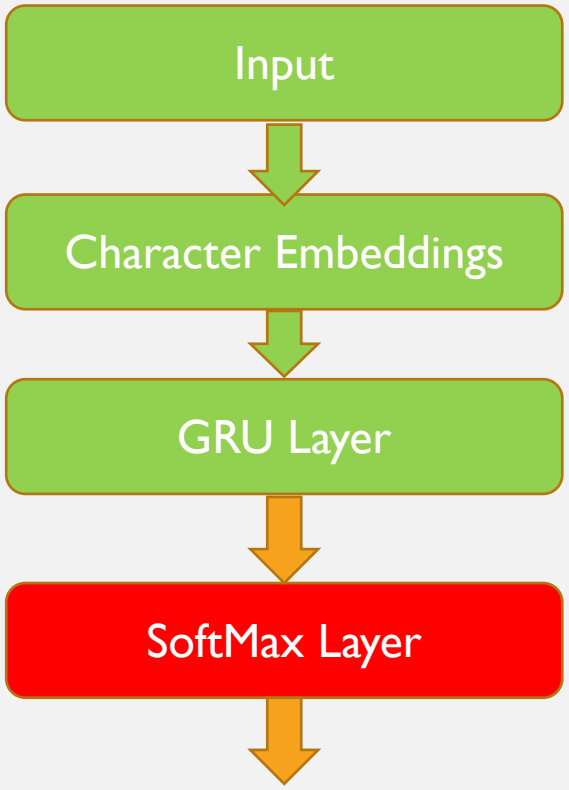


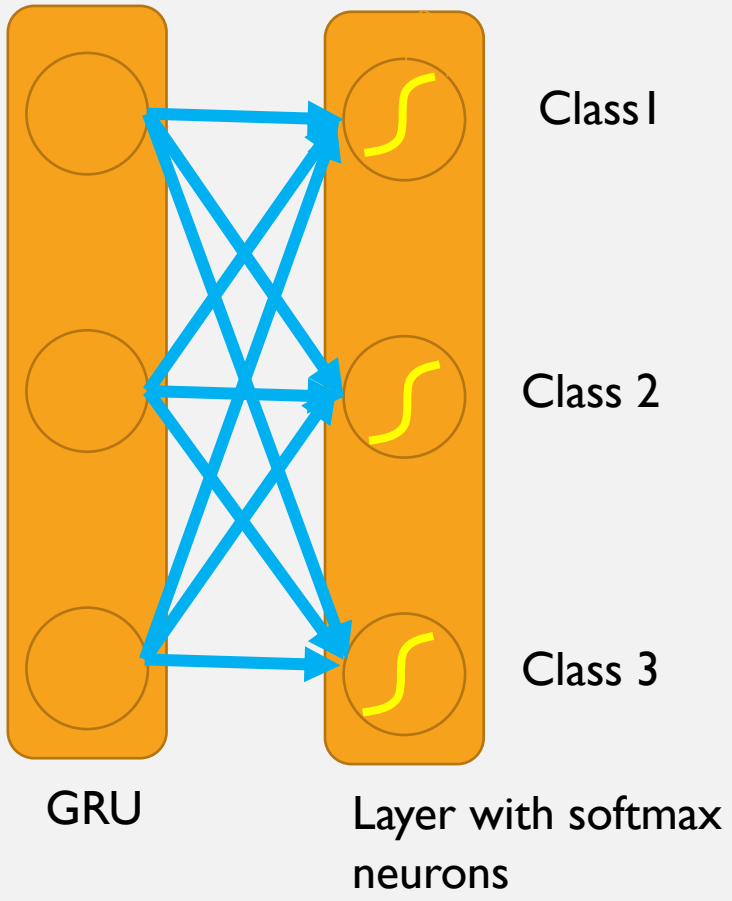
Embedded vector for L | Embedded vector for E | Embedded vector for U

Bi-directional GRU layer was used to learn the all interactions.

Wolters Kluwer

ML MODEL: SOFTMAX LAYER





Wolters Kluwer

EXAMPLES

- Example :"***Ur Leukocyte Esterase***"

  ML model prediction for component:

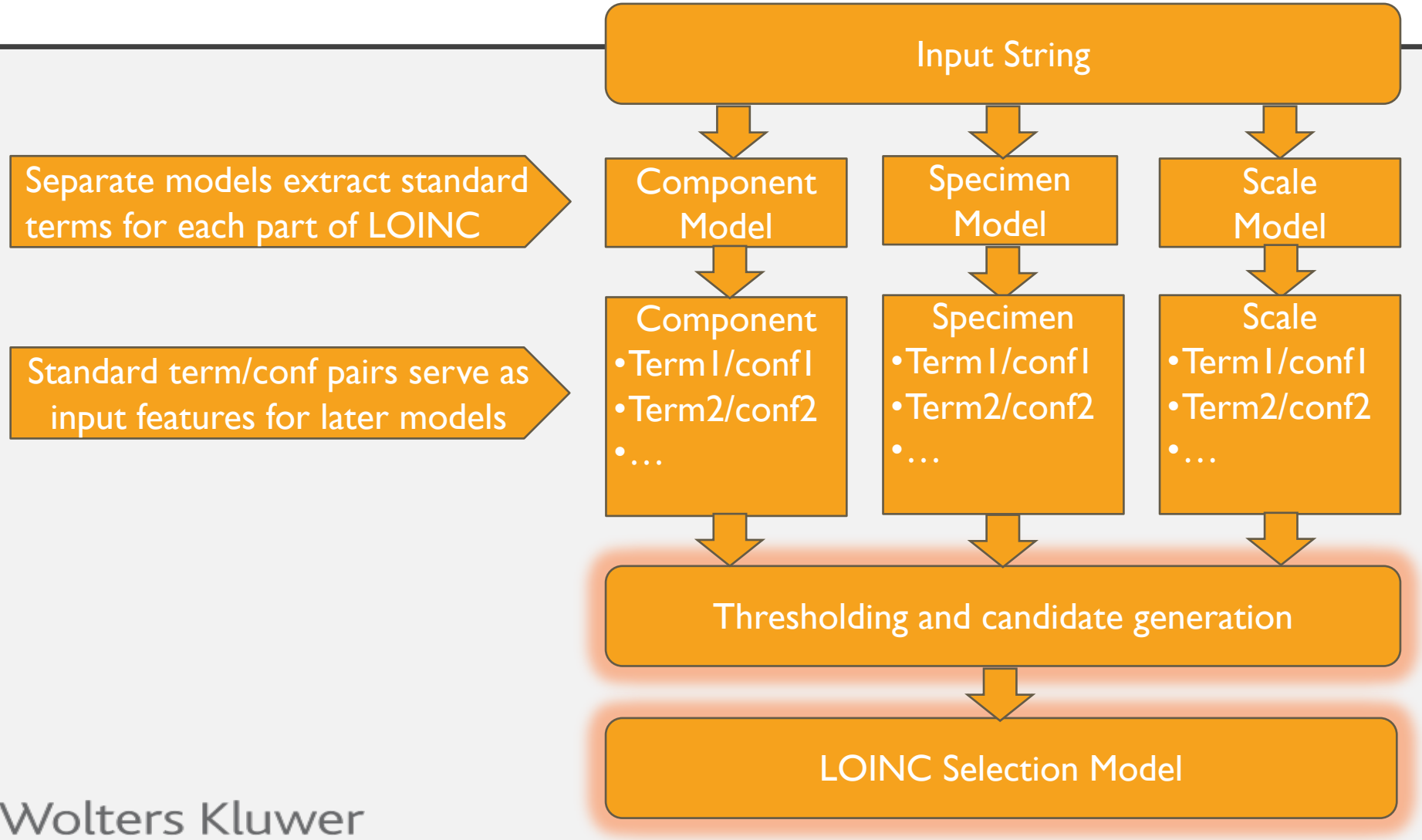
  - ('LEUKOCYTE ESTERASE', 0.945),
  - ('LEUKOCYTES, 0.021),
  - (ALBUMIN', 0.002)

- Example :"***UrLeukocyteEsterase***"

  ML model prediction for component:

  - ('LEUKOCYTE ESTERASE', 0.985),
  - ('LEUKOCYTES', 0.013),
  - (QUERCUS RUBRA AB.IGE', 0.002)

HYBRID MODEL



Wolters Kluwer

HYBRID MODULE: LOGIC-BASED INFERENCE

- Input :"*Ur Leukocyte Esterase*"

Wolters Kluwer

HYBRID MODULE: LOGIC-BASED INFERENCE

Input

"*Ur Leukocyte Esterase*"

Predictions

PART	VALUE	CONFIDENCE
COMPONENT	LEUKOCYTE ESTERASE	0.945
SYSTEM	URINE	0.95
SCALE	ORD	0.86
METHOD	NONE	0.84
PROPERTY	PRTHR	0.763
TIMING	PT	0.988

Wolters Kluwer

HYBRID MODULE: LOGIC-BASED INFERENCE

41K LOINCs Total

component = LEUKOCYTE ESTERASE

6 LOINCs

'2563-5',
'27297-1',
'5799-2',
'59262-6',
'60026-2',
'77563-5'

- Input :"*Ur Leukocyte Esterase*"

PART	VALUE	CONFIDENCE
COMPONENT	LEUKOCYTE ESTERASE	0.945
SYSTEM	URINE	0.95
SCALE	ORD	0.86
METHOD	NONE	0.84
PROPERTY	PRTHR	0.763
TIMING	PT	0.988

HYBRID MODULE: LOGIC-BASED INFERENCE

46K LOINCs Total

COMPONENT = LEUKOCYTE ESTERASE

6

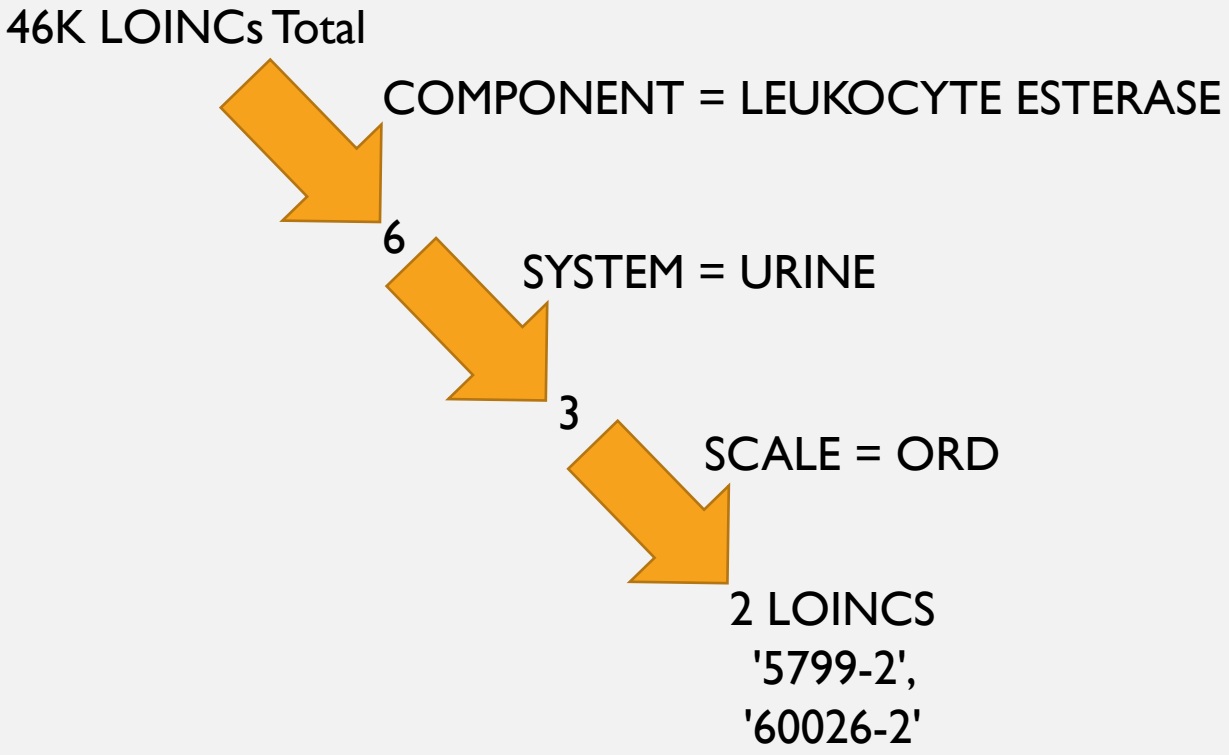
SYSTEM = URINE

3 LOINCs
'27297-1',
'5799-2',
'60026-2'

- Input :"*Ur Leukocyte Esterase*"

PART	VALUE	CONFIDENCE
COMPONENT	LEUKOCYTE ESTERASE	0.945
SYSTEM	URINE	0.95
SCALE	ORD	0.86
METHOD	NONE	0.84
PROPERTY	PRTHR	0.763
TIMING	PT	0.988

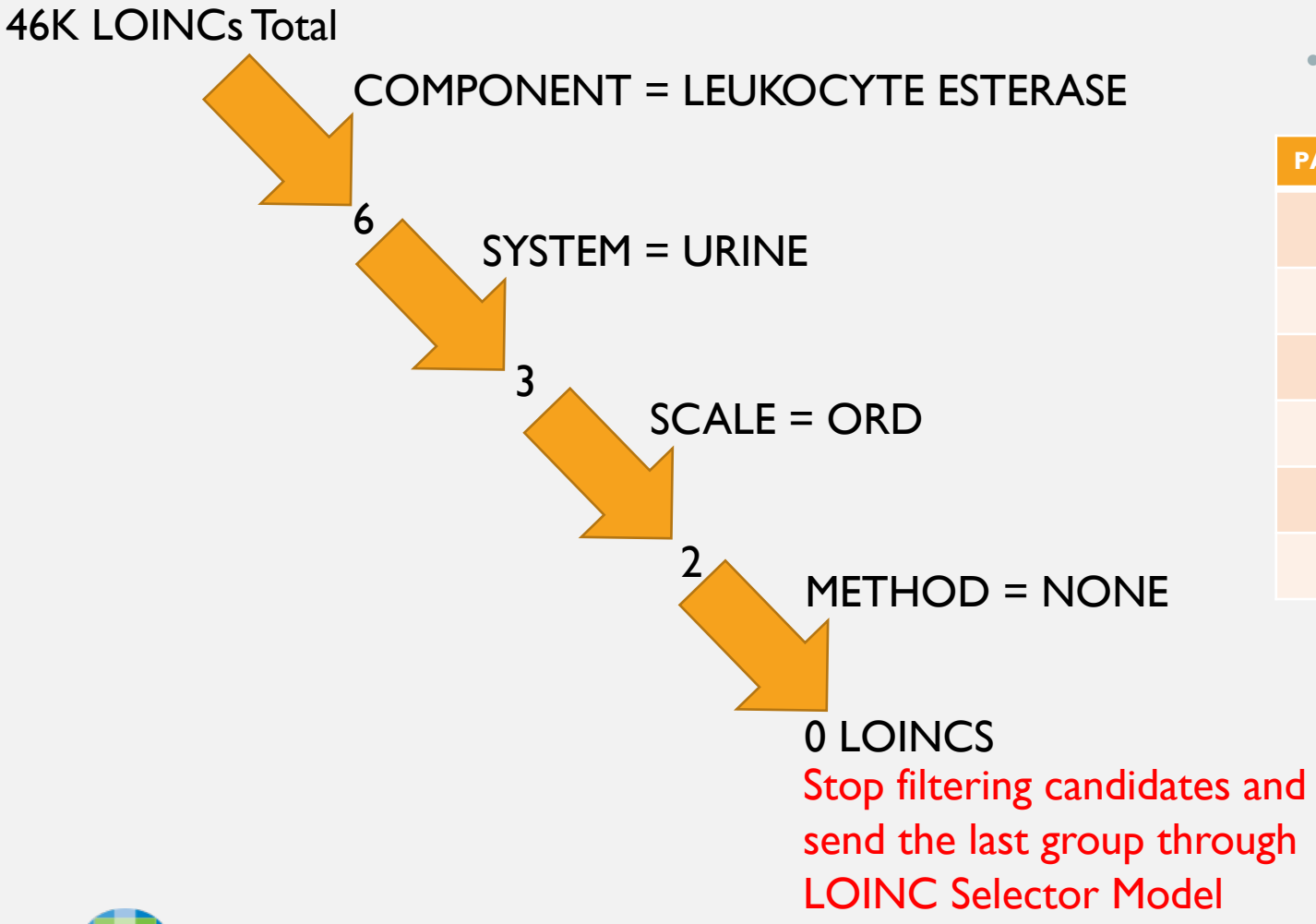
HYBRID MODULE: LOGIC-BASED INFERENCE



- Input :"*Ur Leukocyte Esterase*"

PART	VALUE	CONFIDENCE
COMPONENT	LEUKOCYTE ESTERASE	0.945
SYSTEM	URINE	0.95
SCALE	ORD	0.86
METHOD	NONE	0.84
PROPERTY	PRTHR	0.763
TIMING	PT	0.988

HYBRID MODULE: LOGIC-BASED INFERENCE



- Input :"*Ur Leukocyte Esterase*"

PART	VALUE	CONFIDENCE
COMPONENT	LEUKOCYTE ESTERASE	0.945
SYSTEM	URINE	0.95
SCALE	ORD	0.86
METHOD	NONE	0.84
PROPERTY	PRTHR	0.763
TIMING	PT	0.988

Wolters Kluwer

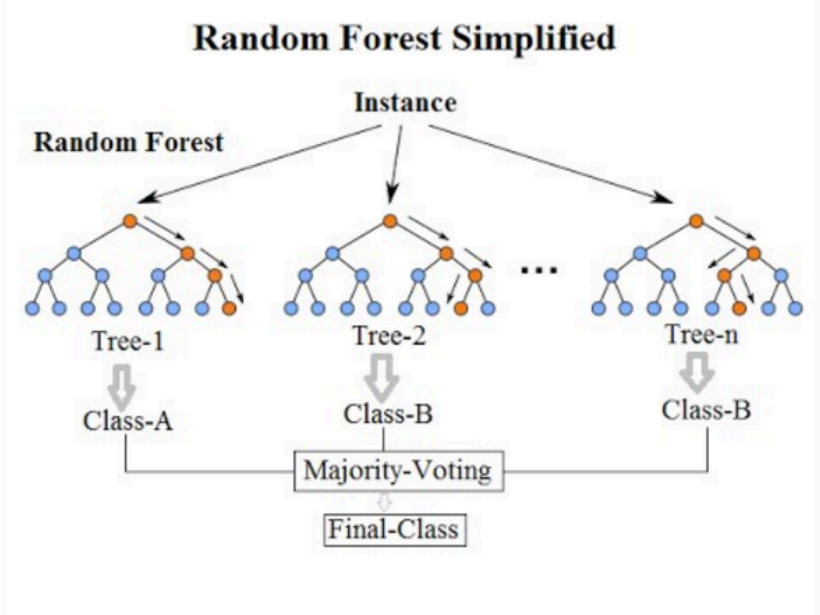
SELECTOR MODEL

1. List of candidates with predicted probabilities for each of the 6 parts
2. *Common Test* Rank the LOINC
3. Length of the LOINC Name
4. Whether the candidate is correct [output]

LOINC Pred Part1	Conf Part1	LOINC Pred Part2	Conf Part2	…	Rank of common test	Length	…	Correct
Leuk	.8	Strip	.7		3000	38		False
Leuk	.8	Ur	.9		65	23		True
Leuko	.5	Strip	.7		120	52		False

Training Data

Input string -> candidate

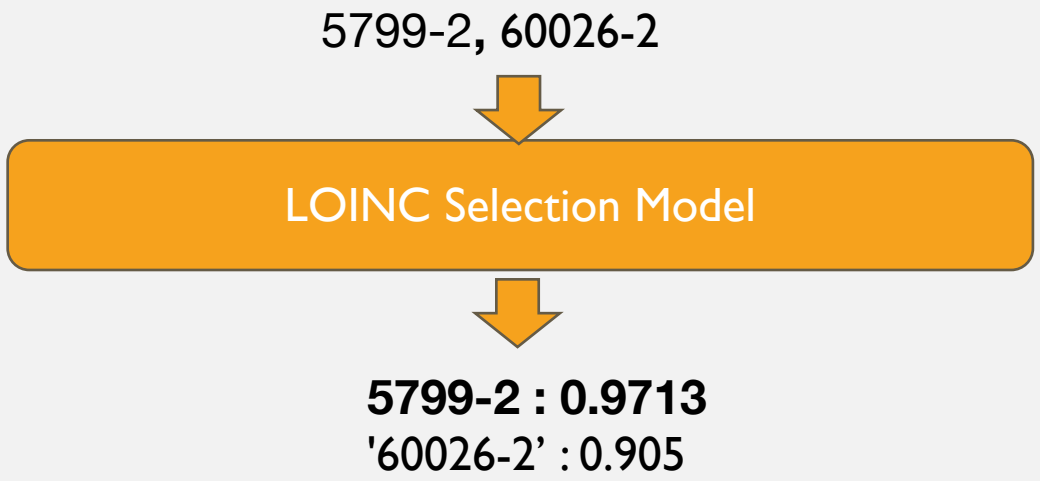


LOINC prediction + confidence

Wolters Kluwer

SELECTOR MODEL

PART	5799-2	60026-2
COMPONENT	0.945	0.945
SYSTEM	0.95	0.95
SCALE	0.86	0.86
METHOD	0.0	0.0
EX. UNITS	0.856	0.856
PROPERTY	0.763	0.763
TIMING	0.988	0.988
RANK	65	3000
LENGTH	52	62



Wolters Kluwer

EVALUATION

Bin	Accuracy%		Coverage %	
	Hybrid	Rules	Hybrid	Rules
p > 0.99	90.4	70.6	10	4.1
0.99 <= p <0.75	85.4	62.2	80.5	62.5
0.75 <= p < 0.5	80.1	37.4	7.5	22.9
P <= 0.5	55.8	18.3	2	10.5

EVALUATION

- Clinical informaticists average 80% accuracy in a completely manual mapping process
- Approaches using purely machine learning scored high for element prediction but were less than 70% accurate at predicting final LOINC codes.
- A hybrid approach combining logic and machine learning provided a dramatic increase in accuracy and coverage
- By accepting predictions with a confidence higher than .5, we can achieve human performance of 80% accuracy on a combined coverage of the top three bins or 98% of all incoming custom codes.

CONCLUSION

- Practical applications of artificial intelligence often require an ensemble of approaches.
- Combining the multiple approaches can overcome their respective weaknesses in particular use cases.
- We found that machine learning approaches were best equipped to extract LOINC elements from noisy text inputs, whereas logic-based methods were better at com-bining those elements into final LOINC codes.

THANK YOU

Wolters Kluwer