

## Team 4

Remi Flicoteaux – Lead  
Pierre Zweigenbaum  
Vassilina Nikoulina  
Yohan Taille  
Sameh Megrhi  
Claire Vinet  
Cyril Grouin  
Namik Taright  
Thomas Séjourné  
Michael Elbaz

## Dat-ICU 2018

**Automated text classification  
for ICD 9 diagnosis code  
assignment from MIMIC  
Database**

**Hopital Rothschild - Campus Picpus**



## ICD coding

Internationally agreed standard from WHO

High value information for hospitals

- For billing
- For epidemiological studies
- For clinical trials
- For hospital activity measures
- For anti-biological stewardship

## Challenges

Time consuming activity

Coding performed by physicians / coders

Not easily reproducible between similar patients

High inaccuracy rate of records requiring rework

Automated coding of medical acts and diagnosis would be a high improvement to provide better care and follow-ups to patients, enhance knowledge of hospitals of their own activity and support researchers with common definitions of comparable events between facilities

## Aim of study

MIMIC contains both textual notes from discharge records completed by treating physicians and human-assigned ICD codes

Can this provide relevant data to train a system that provides automated ICD coding?

## Cohort Selection

All visits with an hospital record

All visits with progress notes (to do)

Select ICD codes with at least 1,000 occurrences (from 59,652 to 51,300 discharge summaries)



# Method

## Data selection:

- SparkSQL
- Pandas

## Feature extraction:

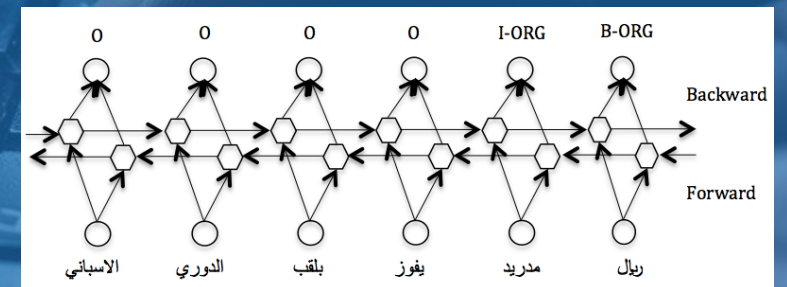
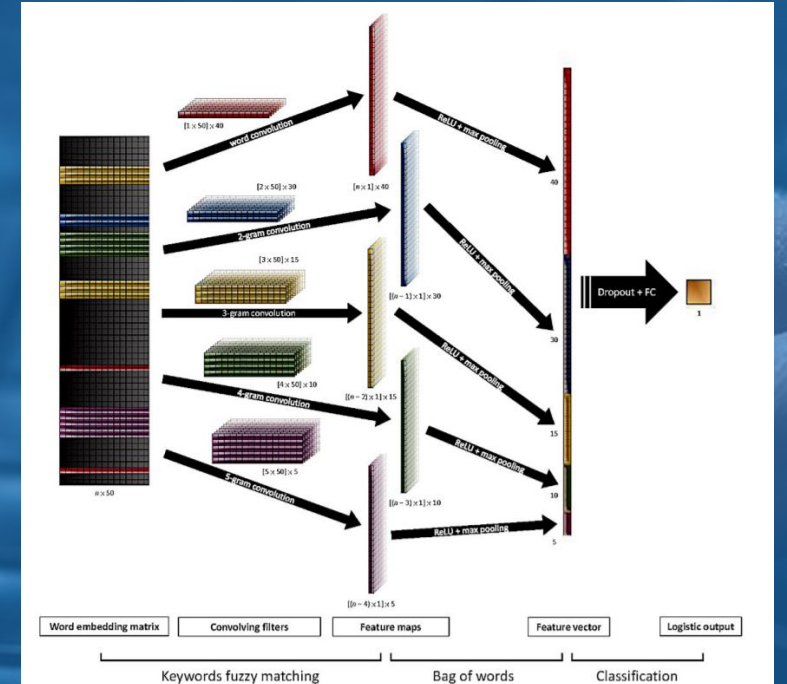
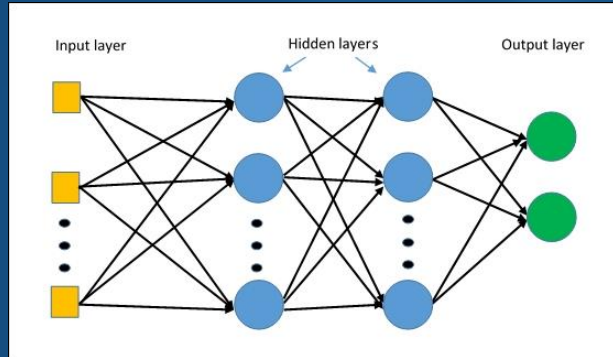
- bag of tokenized words: known ICD terms
- GloVe word embeddings trained on MIMIC texts

## Classification:

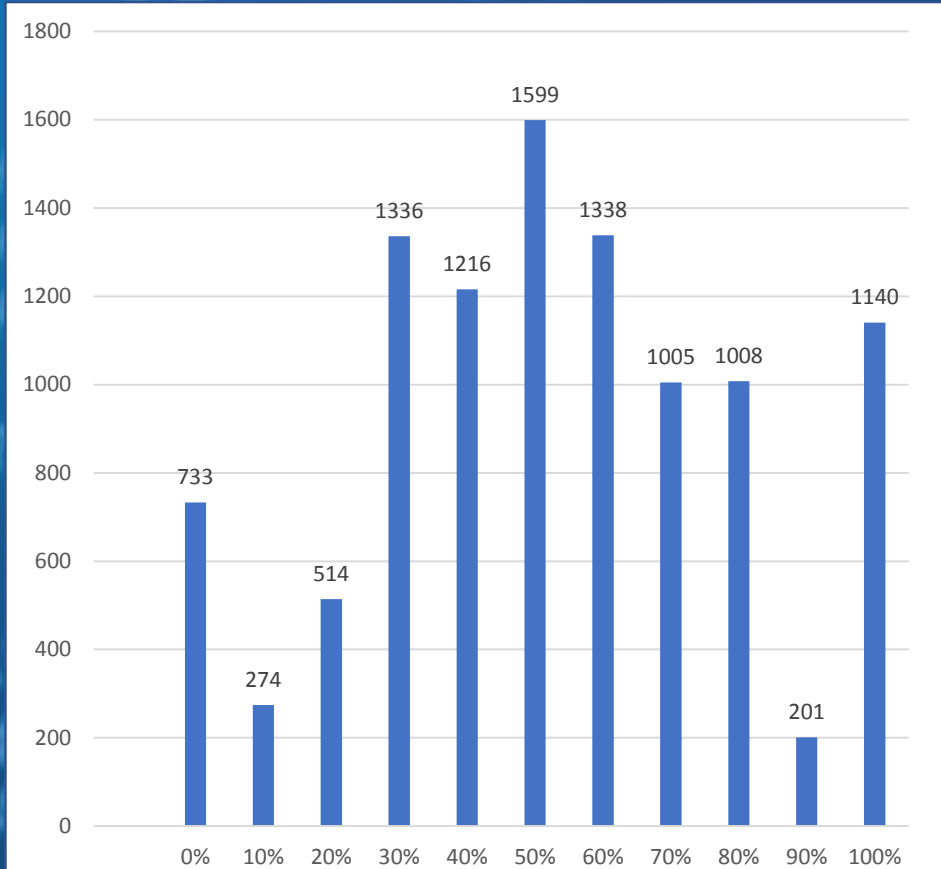
- Convolutional Neural Network
- Bidirectional LSTM
- Multilayer perceptron

## Evaluation:

- F-Score



## Bidirectional LSTM evaluation

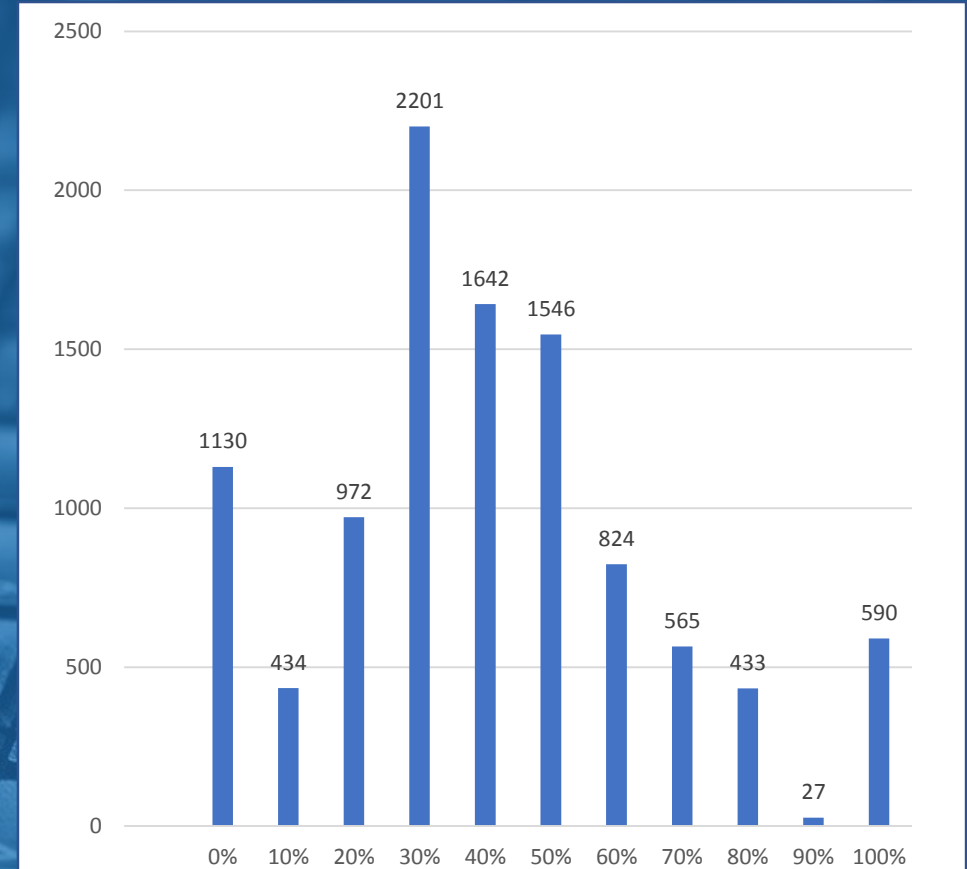


Threshold  
0,25

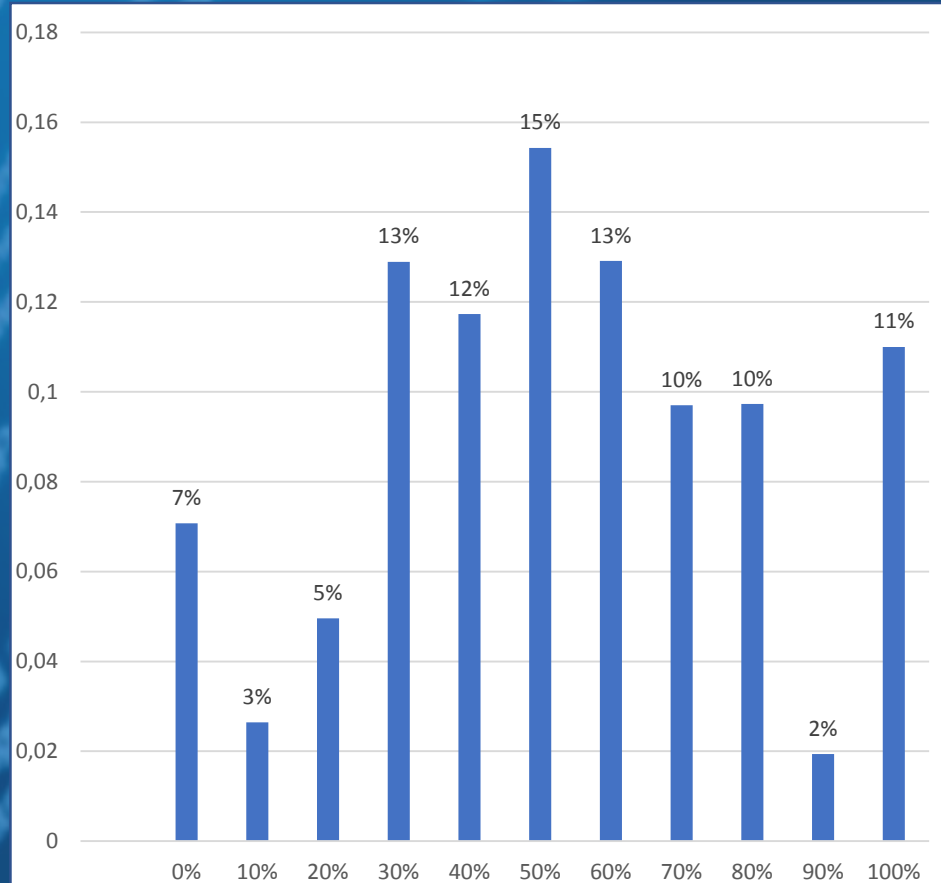
On  
10,354  
expected

Max  
F-Score  
with a 0,3  
thershold

## Convolutional Neural Network Evaluation



## Bidirectional LSTM evaluation

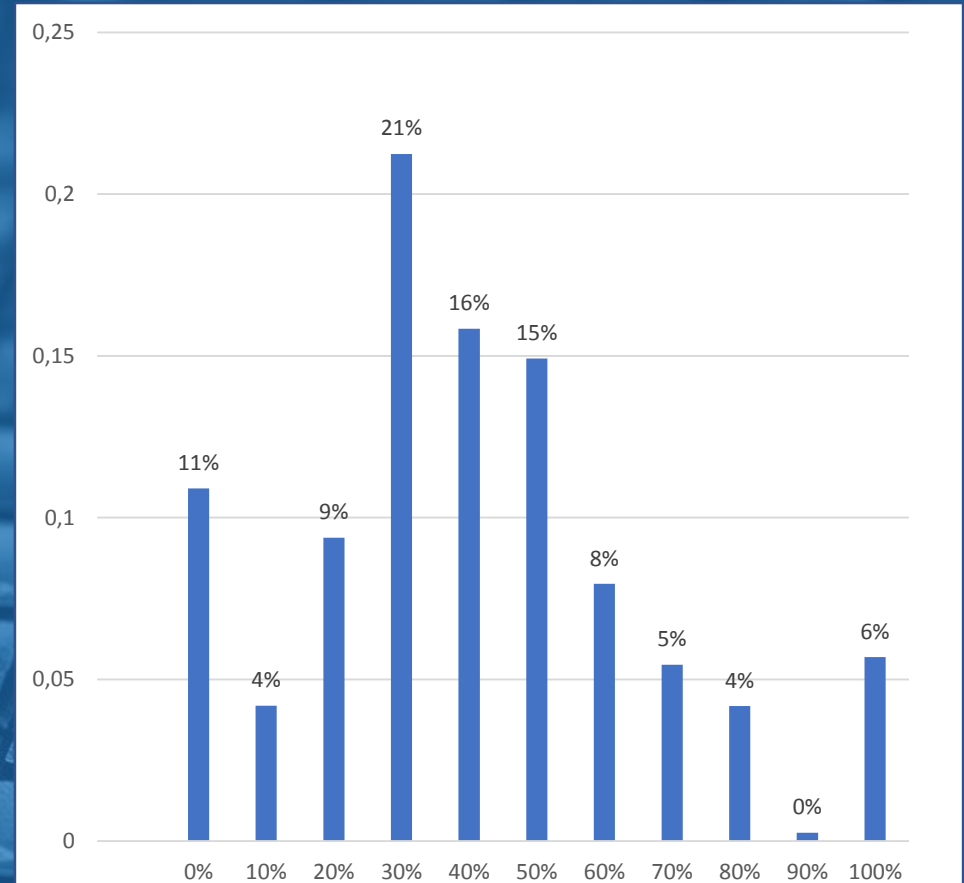


Threshold  
0,25

On  
10,354  
expected

Max  
F-Score  
with a 0,3  
thershold

## Convolutional Neural Network Evaluation





**Further developments  
&  
Discussion**

**Interaction with other  
teams**  
(active learning procedure for  
production...)

**Further improvements  
& testing**

**Web service  
deployment**



# Flask

web development,  
one drop at a time

http://127.0.0.1:5000/icu

Params

Send

Save

aders

Body

Pre-request Script

Tests

Cookies Code

www-form-urlencoded raw binary Text

Date: [\*\*2138-6-9\*\*] Discharge Date: [\*\*2138-6-12\*\*]\n\nDate of Birth: [\*\*2111-2-28\*\*] Sex:  
rvice: Cardiothoracic Surgery\n\nPREOPERATIVE DIAGNOSIS:\n1. Bicuspid aortic valve.\n2. Dilated aorta.\n3.  
Aortic insufficiency.\n\nHISTORY OF PRESENT ILLNESS: The patient has had a heart\nmurmur since childhood and found to  
have a bicuspid aortic\nvalve on echocardiogram, and recently had an increase in the\nsize of the ascending aorta.  
Otherwise, the patient denies\nany other medical problems. [\*\*Name (NI) \*\*] did have surgery in [\*\*2124\*\*] for\nan  
undescended testicle.\n\nSOCIAL HISTORY: Denies a smoking history. Occasional\nalcohol, maybe once per week.\n\nFAMILY  
HISTORY: Noncontributory.\n\nMEDICATIONS ON ADMISSION: Prophylactic antibiotics.\n\nALLERGIES: No known drug  
allergies.\n\nLABORATORY ON ADMISSION: Preoperative vital signs were a\nheart rate of 78, blood pressure 102/68,  
respiratory rate\nof 18. He was a healthy, 27-year-old male. Lungs were\nclear. Heart had a 3/6 systolic ejection  
murmur. Otherwise\nthe examination was within normal limits.\n\nHISTORICAL COURSE: See on [\*\*2138-6-9\*\*] the

Body

Cookies

Headers (4)

Test Results

Status: 200 OK

Time: 4348 ms

Size: 2.82 KB

Pretty

Raw

Preview

HTML



Save Response

```
1 {"HADM_ID":{"0":"000"},"008":{"0":"0.0590254478"},"031":{"0":"0.1057657078"},"038":{"0":"0.1935391575"},"041":{"0":"0.1701635271"},
  "043":{"0":"0.0755844414"},"048":{"0":"0.1731723249"},"049":{"0":"0.1357786506"},"057":{"0":"0.044159025"},"070":{"0":"0.
  .0934861824"},"076":{"0":"0.0124489684"},"112":{"0":"0.052752845"},"162":{"0":"0.045799043"},"197":{"0":"0.0672039911"},"198"
  :{"0":"0.071580708"},"244":{"0":"0.133835271"},"250":{"0":"0.3010516465"},"263":{"0":"0.050171297"},"272":{"0":"0.2639911175"},
  "274":{"0":"0.0731169134"},"275":{"0":"0.0484229922"},"276":{"0":"0.3398755789"},"278":{"0":"0.069890894"},"280":{"0":"0.
  .081481725"},"284":{"0":"0.0393487029"},"285":{"0":"0.264660418"},"286":{"0":"0.0740066618"},"287":{"0":"0.104768455"},"288":{"0"
  :0.0560539663},"291":{"0":"0.0451536588"},"293":{"0":"0.0624726228"},"294":{"0":"0.085593164"},"296":{"0":"0.0544287451"},"300"
  :{"0":"0.0753911287"},"303":{"0":"0.0739566609"},"305":{"0":"0.1426698118"},"327":{"0":"0.0559952073"},"331":{"0":"0.0579498485"},
  "338":{"0":"0.0338873677"},"345":{"0":"0.0497993678"},"348":{"0":"0.089221552"},"357":{"0":"0.0518827587"},"362":{"0":"0.
  .0403765589"},"401":{"0":"0.4099258482"},"403":{"0":"0.1689960361"},"410":{"0":"0.1668179929"},"411":{"0":"0.0528151654"},"413"
  :{"0":"0.0329842679"},"414":{"0":"0.2568119168"},"415":{"0":"0.0499068573"},"416":{"0":"0.0807388127"},"424":{"0":"0.1329770833"},
  "425":{"0":"0.076716356"},"426":{"0":"0.0622364283"},"427":{"0":"0.3601903617"},"428":{"0":"0.3217853308"},"433":{"0":"0
```



# Thank You to our hosts

