Machine Learning application to doctor-patient interaction

FiloBlu: Sentiment Analysis

andrea.ciardiello@roma1.infn.it

INFN
Istituto Nazionale
di Fisica Nucleare

Outline

- **FiloBlu** is an industrial research project funded by POR FERS.

- The collaboration is composed by local industrial partners, hospital structures and research institutes.

- The Aim of the collaboration is to implement ICT technology to the support of long term Domiciliary Care

- Within this collaboration the INFN has developed a machine learning based text classification method.

Motivation: domiciliary care of oncologic patients

**Long Term Domiciliary Care** for oncologic patients is preferred
- Higher standard of living
- Psychological wellbeing
- Cheaper than hospitalization

To successfully follow therapy during the domiciliary care the patient/Caregiver is in constant contact with healthcare professionals
- Hospitalization should be considered the "last resort"
- Clinicians, psychologists, nurses have an active role
- Frequent monitoring on therapy quality and general health
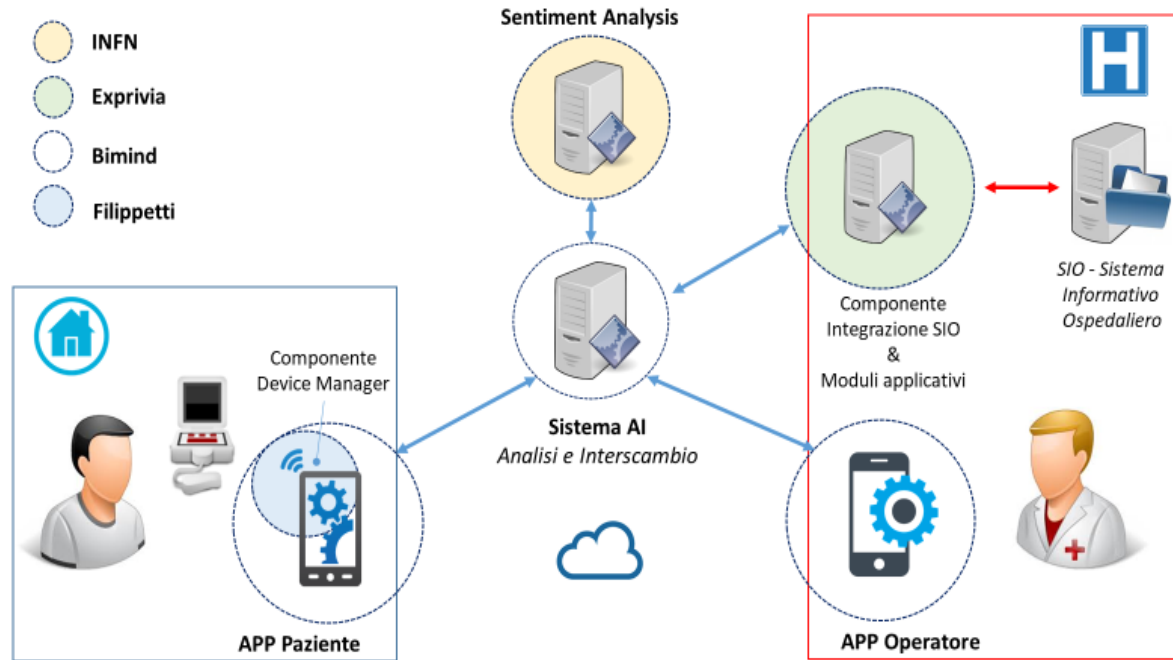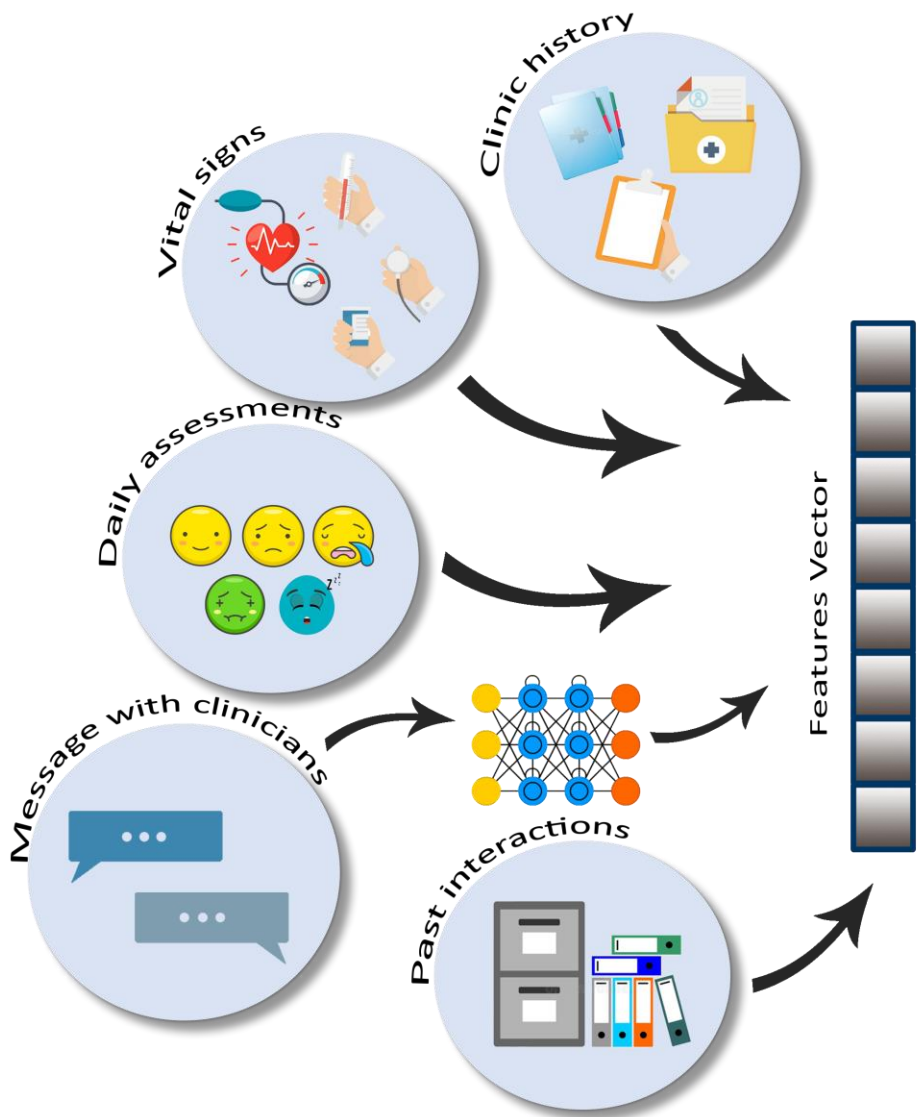
When the FiloBlu project will be completed, the automated classification method will take advantage of all the information gathered by the application
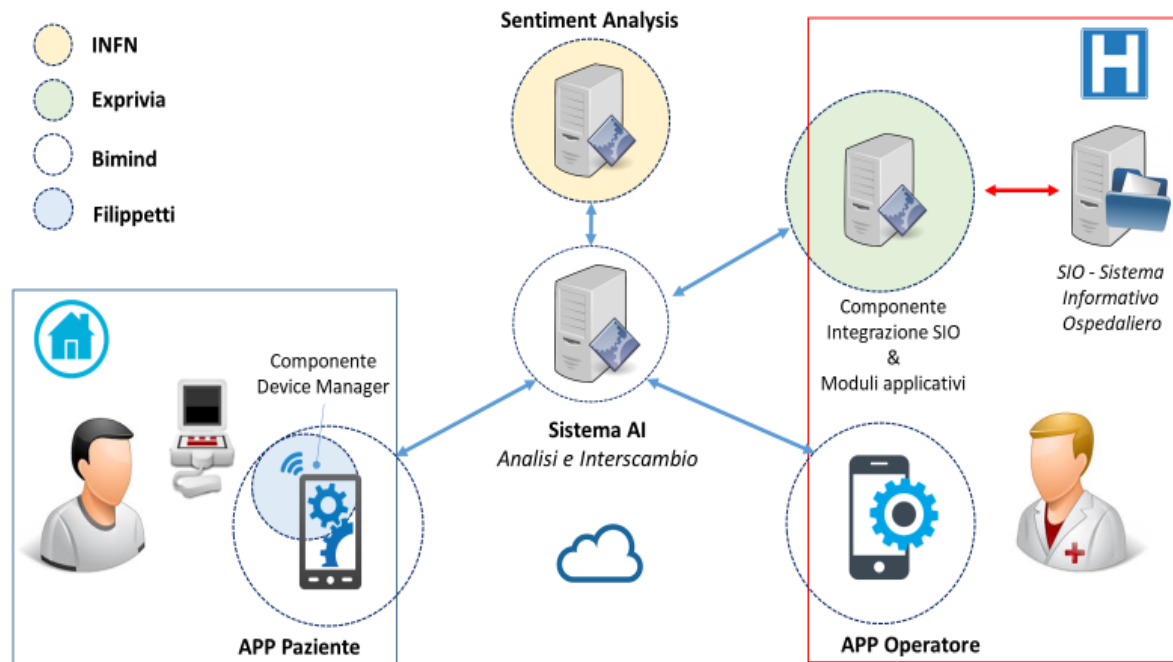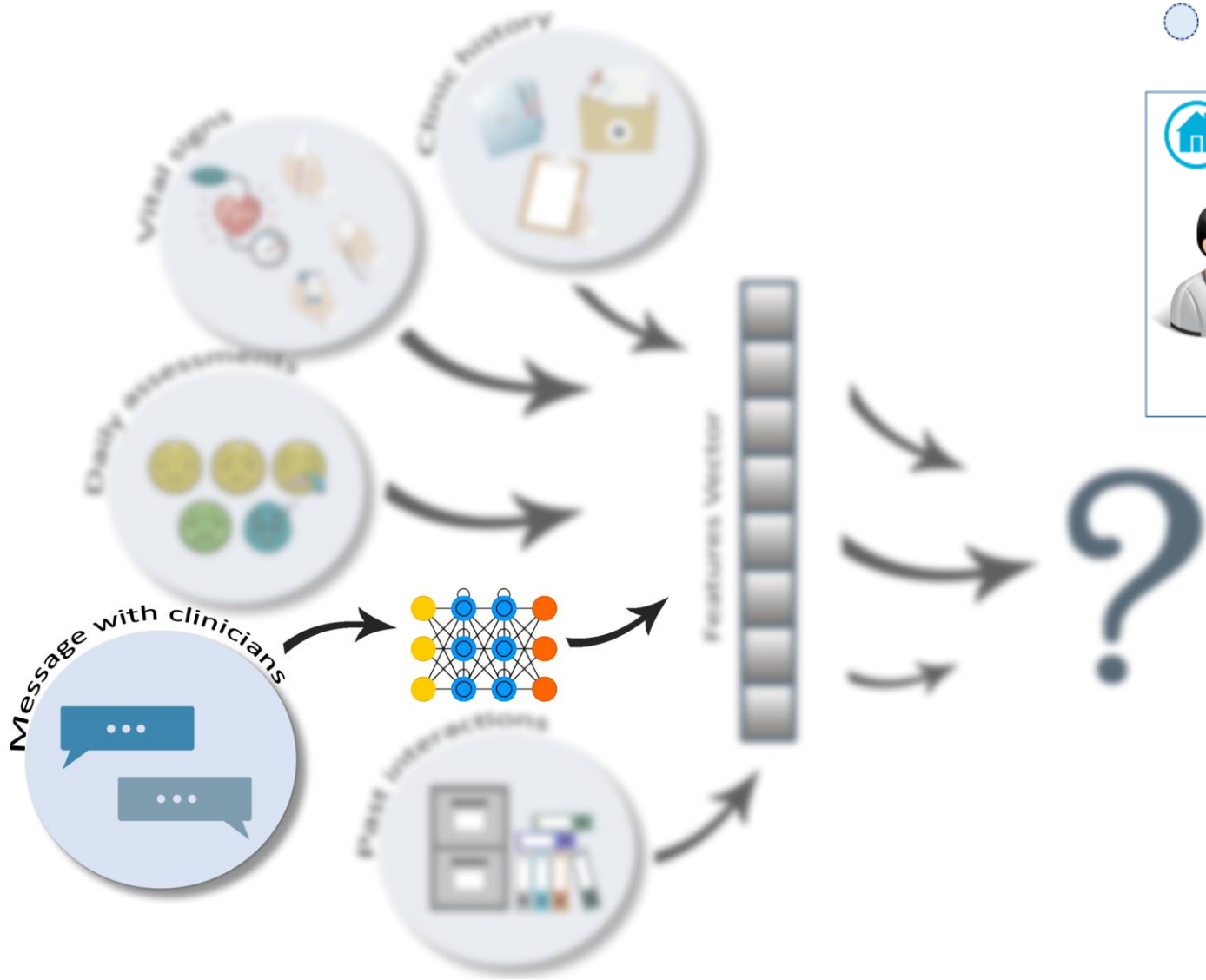
This presentation is about the NLP technique implemented for the short message analysis

Screenshot from the APP

1/28/2020   ARPG   6

Physiological Parameter are registered in the database

The visualization helps the clinicians express their feedback taking them into accounts

Method Summary

Messages are gathered from the app.
Multiple message in a short timeframe are merged.

Preprocessing → Complexity Reduction:
Remove capitalization, punctuation, numbers …
Assumption: capitalization, punctuation does not provide useful information.
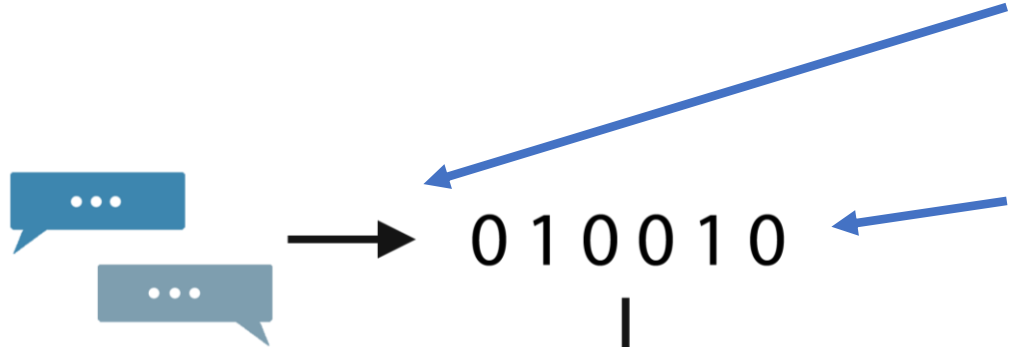
From text to machine readable data:
Message vectorialization

0 1 0 0 1 0

A Neural Network is used for the classification
Dense layers (fully connected)

Topic classification   Priority
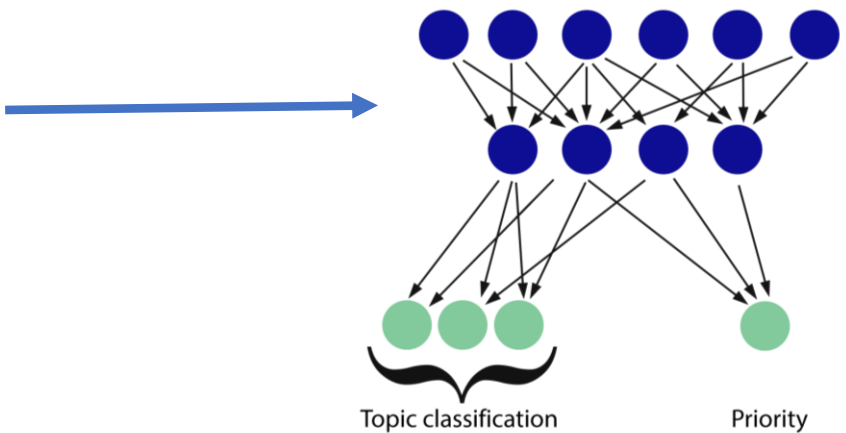
Multiple classification:
Topics and priority score

Topics Classification:
To whom the message should be sent

Priority score:
Is used to set an "attention level" for the message

From text to Machine readable Data

Original message : Buongiorno dott. Giuliani, Oggi mi duole la testa e mi sento molto fiacco

Complexity Reduction: Making Assumptions

Remove capitalization, punctuation, numbers
Assumption: capitalization, punctuation does not provide useful information.

Stop Words: Common words that are not informative   **Stemming**: Substitute a word with its radical
- Remove sparse terms (rare words)
- Remove other terms (e.g. proper nouns).

Stemming/SW : duol*, testa, sent*, molt*, fiacco,

From text to Machine readable Data

Original message : Buongiorno dott. Giuliani, Oggi mi duole la testa e mi sento molto fiacco

Tokenize: From text to numerical ID

Tokenize: duol* → 123, testa → 45, sent* → 3452, molt* →7300, fiacco → 168 ,

$$X = \begin{bmatrix} [1, 0, 0, 0, 0, 0, 0, \dots], \ \#testa \\ [0, 1, 0, 0, 0, 0, 0, \dots], \ \#duol* \\ [0, 0, 1, 0, 0, 0, 0, \dots], \ \#feel \\ [0, 0, 0, 1, 0, 0, 0, \dots], \ \#very \\ [0, 0, 0, 0, 1, 0, 0, \dots], \ \#sick \end{bmatrix}$$

One-hot encoding
duole + testa is not fiacco

X = N x P matrix
- N = Number of elements in message
- P = Number of words in dictionary

X = main input for many computational text analysis applications.

Complexity Reduction: Making Assumptions
Discard Word Order (Bag of Words)
Assumption: Word Order Doesn't Matter.

$$X_{BW} = [1, 1, 1, 1, 1, 0, 0, \dots],$$

•Bag of word models using character ngrams can be very efficient. **Do not underestimate them!**. They are relatively cheap to compute, and also easy to interpret.

Recipe to Supervised Learning Method

Supervised methods: Hand coding is used to train statistical models to classify texts in pre-determined categories.

1) Set of known categories
   - Urgent, Not Urgent
   - Attention Level
   - Topics

2) Set of hand-labeled messages
   - Labeling done by clinicians
   - Training Set: messages we'll use to learn how to classify
   - Validation Set: messages we'll use to learn how well we classify

3) Set of unlabeled messages that we want to classify

4) Method to extrapolate from hand coding to unlabeled messages

5) Validate by comparing predicted label to real (hand-coded) label.
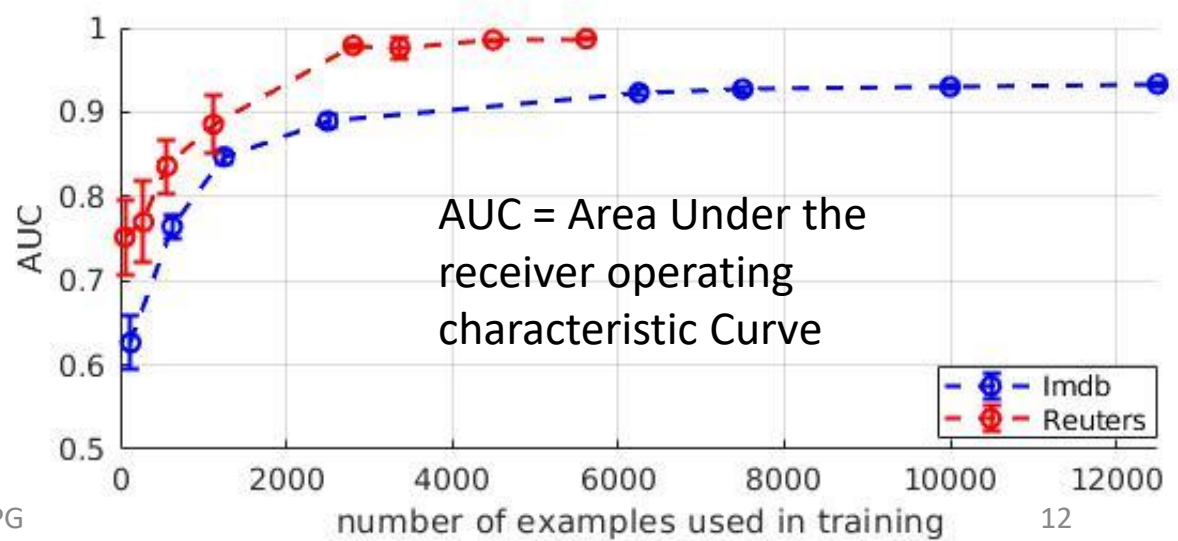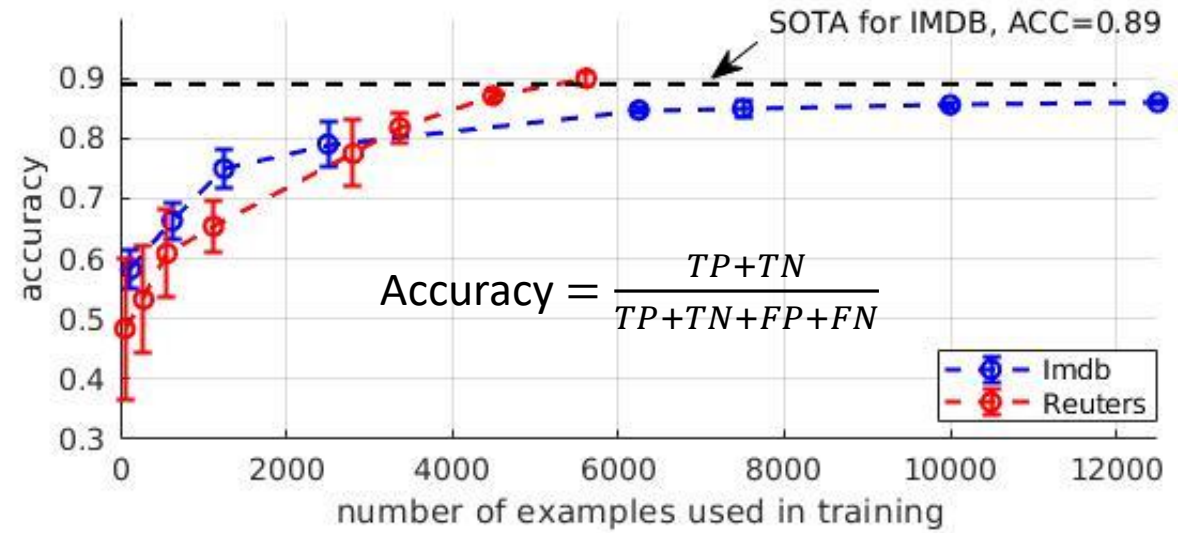
Model performance on Benchmarks

In the context of medical communication, data is scarce :

- Privacy concern (sharing dataset may be difficult)
- High topics specialization, low generalization between user groups

The use of simple model (small parameters space) that need less data to be trained is beneficial.

Simple models can achieve good results on test datasets compared to SOTA

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

SOTA for IMDB, ACC=0.89

AUC = Area Under the receiver operating characteristic Curve

<div>

IMDB dataset
Movie review (Sentiment)
Binary classification:
Positive / negative
25,000 examples
12,500 used for testing

</div>

<div>

Reuters newswire dataset
Topics classification
Multi class:
46 topics
11,228 examples
5,614 used for testing

</div>

How to get data? Web Scraping

- Many web sites offer thousands of already written interaction

- Real word usage (spelling error occurrence, low level vocabulary)

- Week labelling (mainly for legal reason)

Message from patient

Label ?

Diagnosi prostata ingrossata

Buongiorno ho 71 anni dopo varie visite ecografie prostatiche ,e flusso urinario la diagnosi prostata ingrossata, l'urologo mi prescritto omnic da 04 mg,e proscar da 5 ,mg visto il scarso risultato. Le ha cambiate con urorec da 5 mg ma niente e cambiato urino poco e male e ora con il caldo non ho neanche lo stimolo x urinare puo darmi un consiglio sono molto preoccupato. Grazie della gentile risposata buona giornata

RISPOSTA DI:
DR. GIOVANNI
Specialista in Andrologia e Patologia della riproduzione umana

Guarda il profilo

Gentile lettore , il consiglio in questi casi è sempre quello di risentire in diretta sempre il suo urologo di fiducia: purtroppo situazioni cliniche particolari, come la sua, richiedono sempre un primo passo decisivo e fondamentale, cioè una visita clinica diretta; senza questa noi, da questa postazione, nulla le possiamo dire di preciso e mirato. Detto questo si ricordi comunque che sempre la visita medica specialistica rappresenta il solo strumento valido per poterle dare un'indicazione diagnostica mirata e poi eventualmente una prospettiva terapeutica corretta e che le informazioni fornite via internet vanno sempre intese come meri suggerimenti clinici e di comportamento. Un cordiale saluto.

Data Labelling :

- Dataset scraped from social networks:
  - ~ 2500 examples from different sources
  - Hand – labelled in 8 Semantic cases
    - t - Terapia oncologica
    - m - Misure, osservazioni:
    - n - Nutrizione, anche via flebo
    - f - Farmaci non oncologici, integratori
    - d - Diagnosi oncologiche, metastasi, mutazioni
    - g - Gestione
    - s - Stato d'animo ,preoccupazioni
    - a - Questioni amministrative

From the 8 main Semantic classes the dataset is divided in 3 main topics; priority messages are flagged separately

- 3 topics
  - A - Administrative (a, g)
  - C – Clinical information (t, m, n, f, d)
  - S – Sentiment, psicological well-beeing (s)

- P – Priority score, applied to all classes

Results on Test Set

ROC and AUC for the Test set classification
Curves for topics (A,C,S) and priority (P) are reported. Micro average is the mean on the entire dataset disregarding the class. Macro average in the mean performance between the classes.

- Train data:
  - ~ 2500 hand-labelled messages gathered from cancer support Facebook groups.

- Test data:
  - ~ 450 messages, independently labelled, from a different source.

Word usage in messages of different topics has many unique elements:
- Good results on Topics classification can be achieved with few examples of each class.
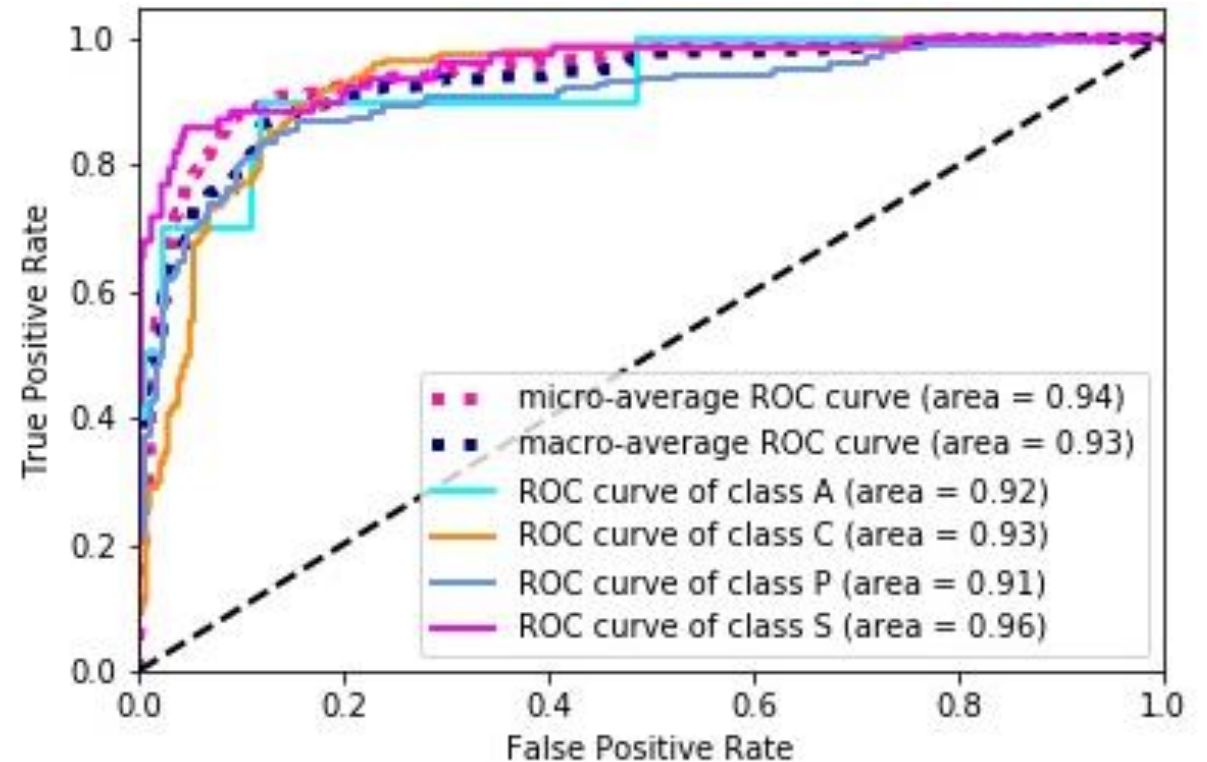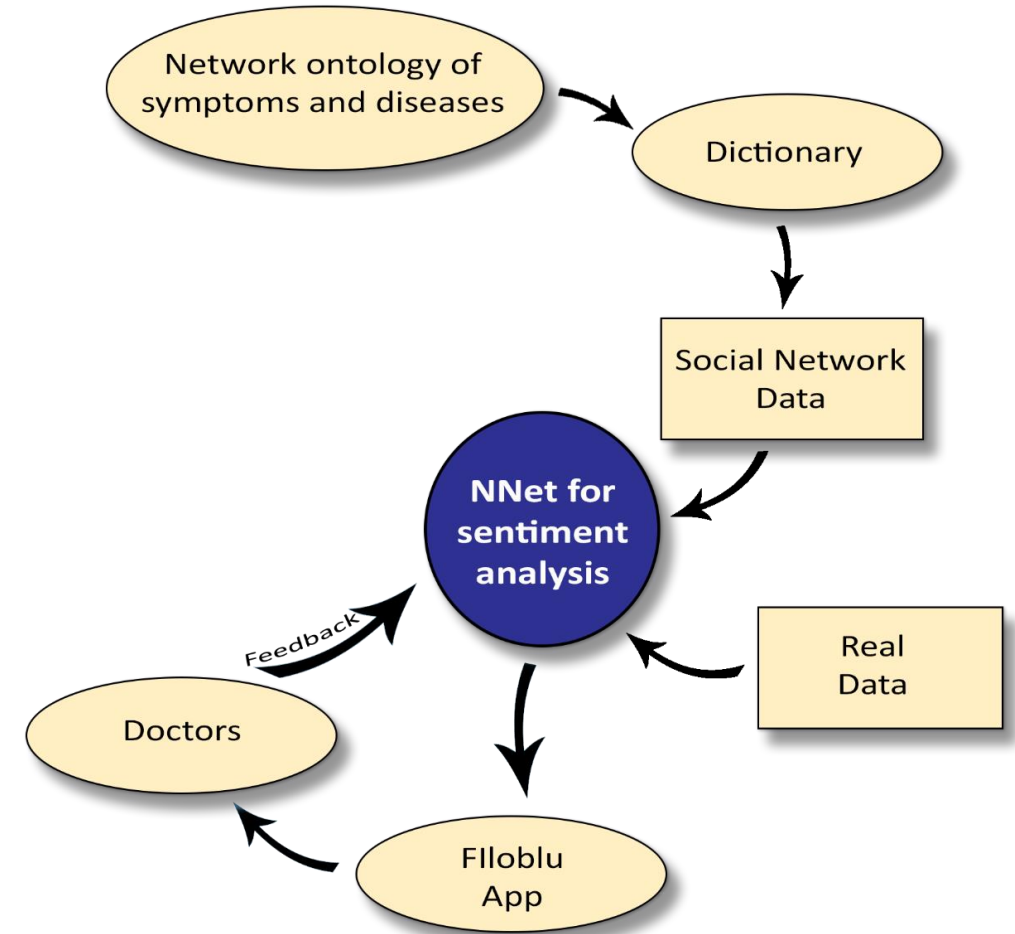
Retraining strategy and customization

Initial classification of the network is based on data gathered from the web. Dictionary of relevant word are composed from a "graph ontology of medical dictionary".

We develop a Retraining methods to help tailoring the decision capabilities of the networks to the everyday use in the hospital settings

- Clinicians that use the APP have the possibility to send a feedback on the classification performance.

- Data from real patient is gathered and saved together with the label assigned by the healthcare professional.

- The network is retrained on the real data

FEEDBACK!

Conclusions

- An automated text analysis method to classify short messages sent to clinicians has been developed
- A simple model may have good performance when data is scarce

- A retraining routine to customize the network classification to clinicians needs is implemented

- If BiMind successfully manage to gather enough data the development of a state of the art NLP system will be possible

andrea.ciardiello@roma1.infn.it

How to get data?
From App Usage

- Clinicians can select the attention level of an interaction directly from the APP

- The algorithm remember the choice and learn from the medic decision

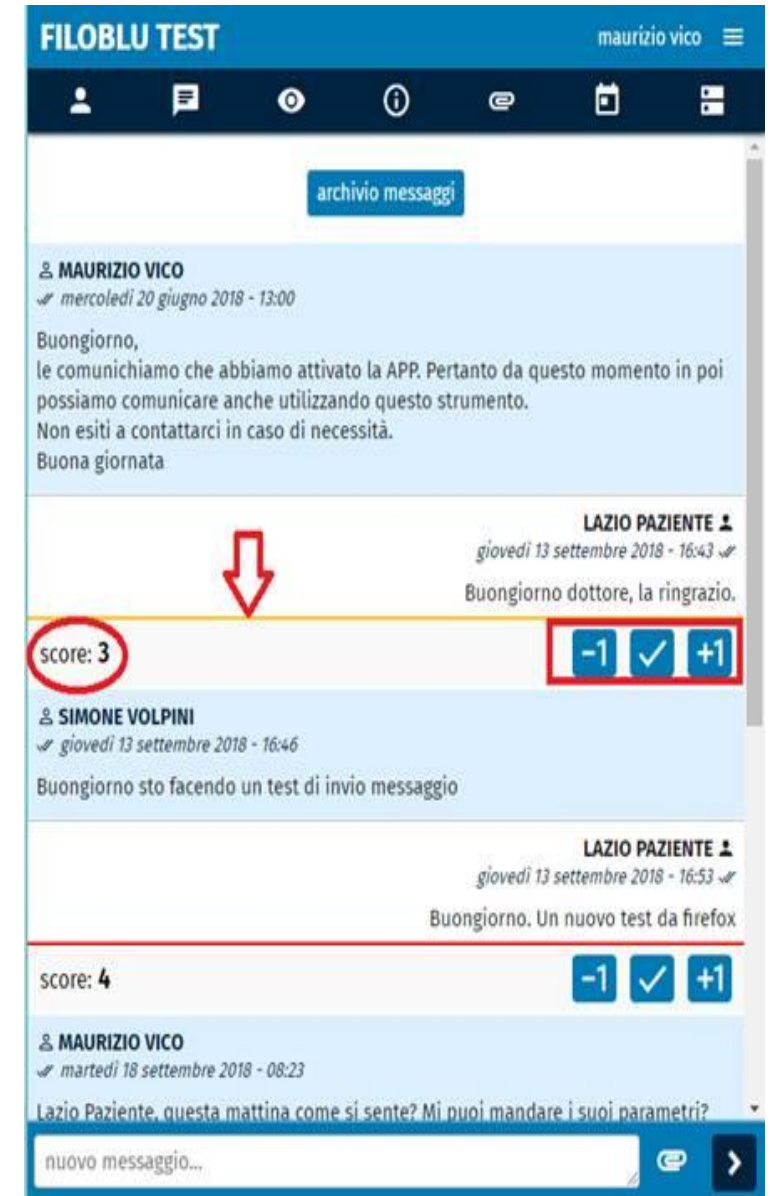- Similar cases (same context, same message) can be easily recognized

Interactions are scored automatically by the network.
The App allows the users from the medical side to express a feedback on the classification.
At the next Retrain Phase, weights of the network are modified accordingly.

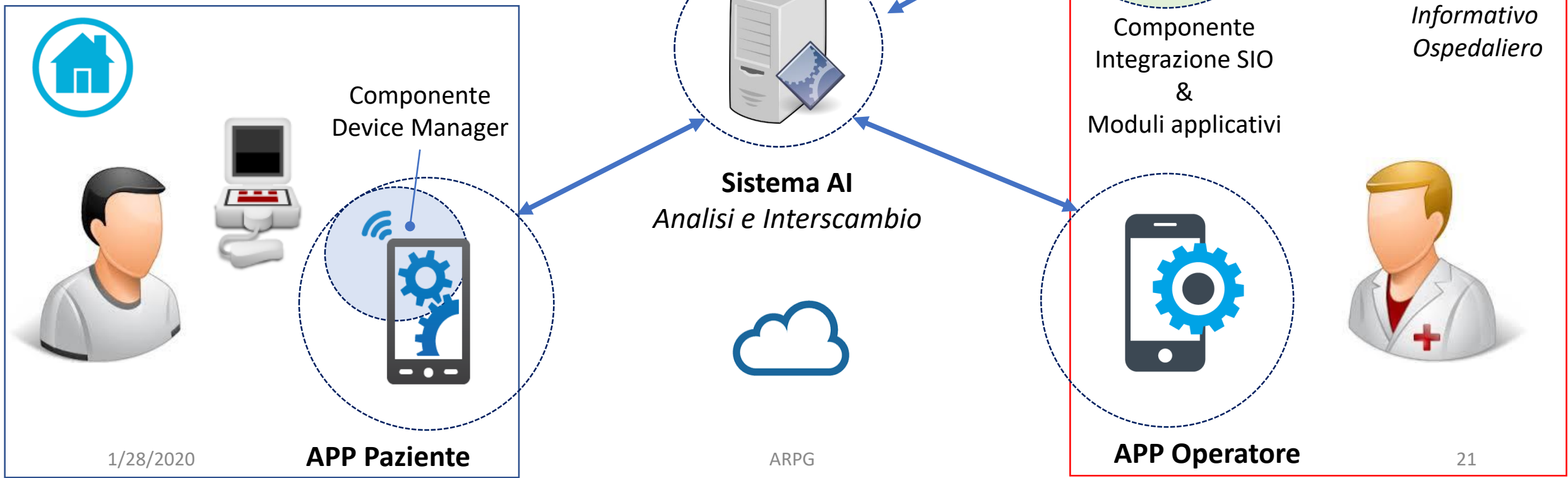Feedback on Priority Score can be given with a 'too much '/ 'too little' interface.

The feedback on Topics Classification consist on "A Forward To" button

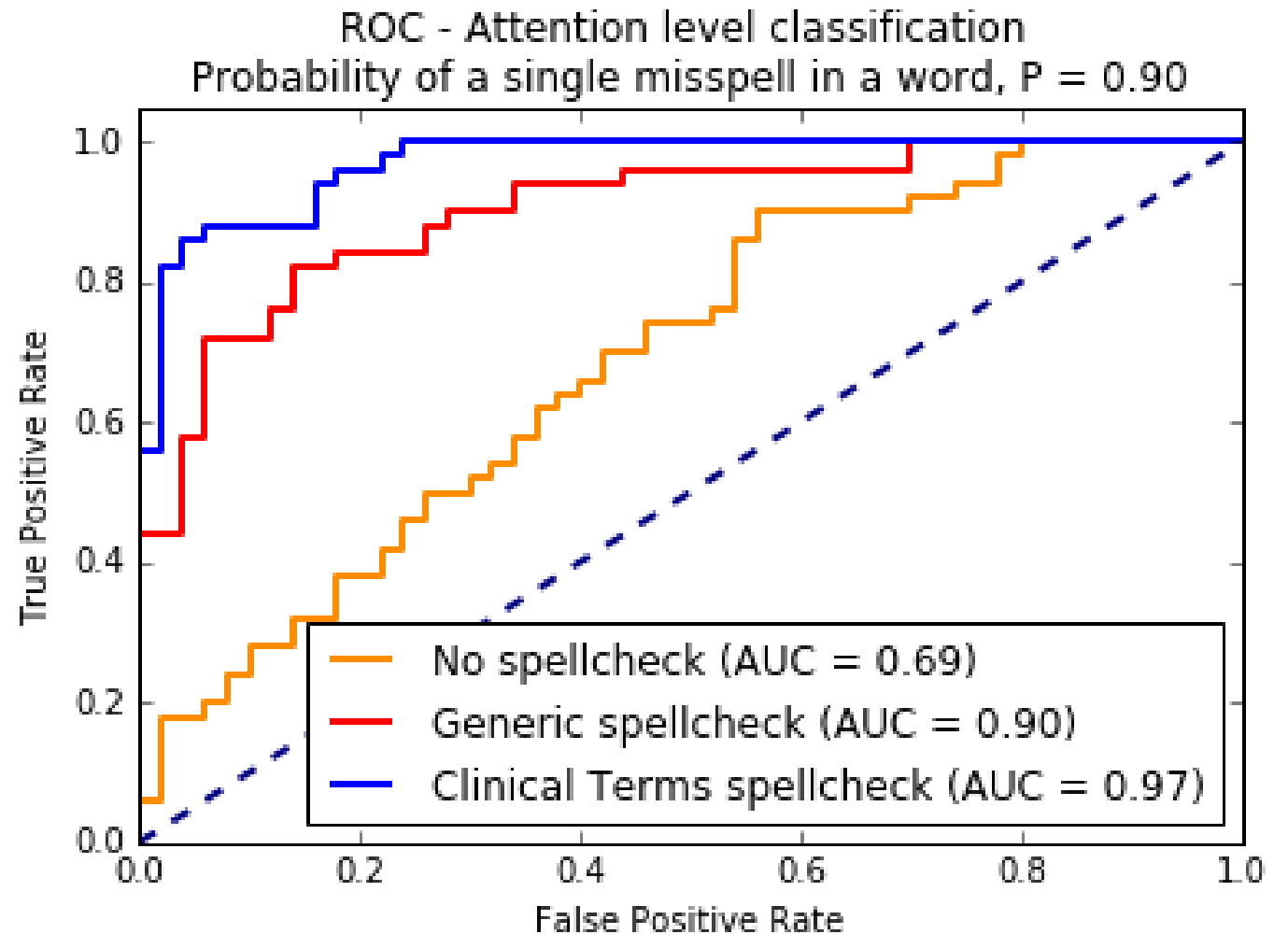Sentiment Analysis

INFN

Exprivia

Bimind

Filippetti

Componente
Integrazione SIO
&
Moduli applicativi

*SIO - Sistema
Informativo
Ospedaliero*

Componente
Device Manager

Sistema AI
Analisi e Interscambio

APP Paziente

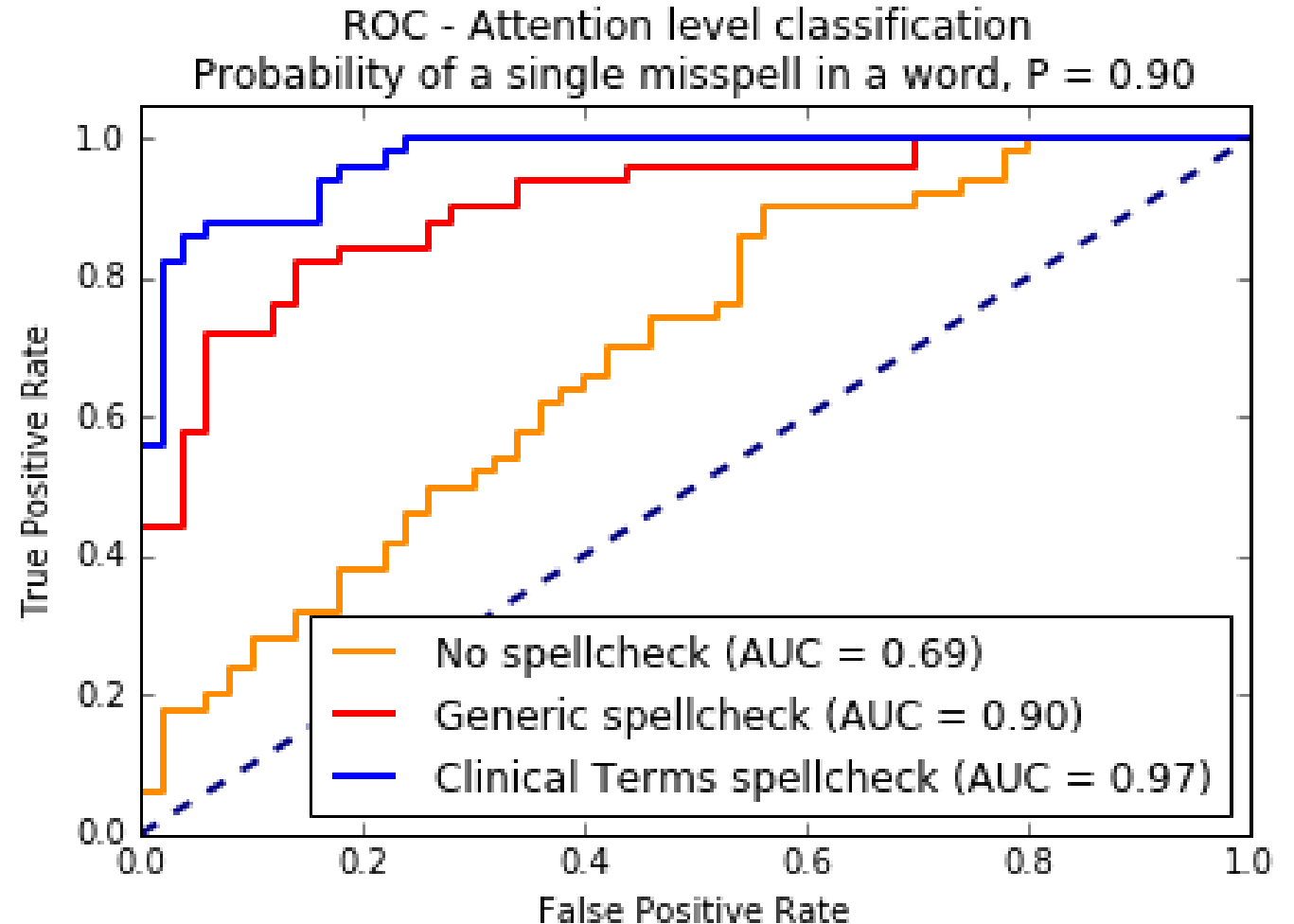**APP Operatore**

Spell Check

Spelling errors can be destructive. When a word is not recognized it will appear as OOD (out of dictionary), effectively removing information from the message.

ROC - Attention level classification
Probability of a single misspell in a word, P = 0.90

Legend:
- No spellcheck (AUC = 0.69)
- Generic spellcheck (AUC = 0.90)
- Clinical Terms spellcheck (AUC = 0.97)

- Introduzione di un semplice correttore automatico

- Spell checker generico :

- <mark>fegate</mark> → <mark>negate</mark>

- la correzione dipende dalla frequenza relativa delle parole

- Spell checker a tema clinico

- <mark>fegate</mark> → <mark>fegato</mark>

ROC - Attention level classification
Probability of a single misspell in a word, P = 0.90

No spellcheck (AUC = 0.69)
Generic spellcheck (AUC = 0.90)
Clinical Terms spellcheck (AUC = 0.97)

- Anche per costruire un spell checker semplicissimo ma che sia contesto dipendente abbiamo bisogno di un corpus adatto