



# Fall Detection in EHR using Word Embeddings and Deep Learning

**Henrique Santos** - PUCRS - Brazil

19th IEEE International Conference on Bioinformatics and Bioengineering

Brief context ...

---



# Multidisciplinary Group and ...

---

## **PUCRS**

Pontifical Catholic University - Rio Grande do Sul

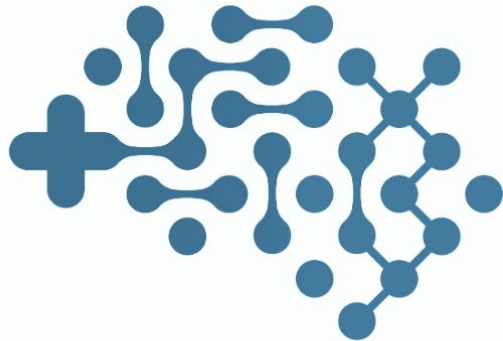
## **PPGCC and PPGGB**

Graduate program in Computer Science

Graduate program in Biomedical Gerontology



## ... and Multicentric Group



Research Group on  
**Artificial Intelligence**  
in Healthcare



<http://www.inf.pucrs.br/ia-saude/>



# Scope of this Research

---

## **Fall**

This adverse event is very common in the hospital environment. The Morse Fall Scale measure the risk of patient fall.

There is a underreport number of fall event notifications.



# Report System

io S/A - Terminal: 131.158.10.91/ Job: 988

Prescrição Acompanhamento Bloco Cirúrgico Linhas de Cuidado Consultorias Núcleo de Epidemiologia Agravos Notificação Rede Sentinela Arquivo Médico Cartão do Usuário



**Sistemas de Informações Médicas**  
**Prontuário Eletrônico do Paciente**

### Registro de Evento - Farmacovigilância

Selecione o Conselho CRM Senha

CRM

Registro do Paciente Medicamento

Pesquisa Paciente

Evento

Erro de medicação

Tipo de erro de medicação

- ☐ Atraso na administração(superior a 30 min)
- ☐ Diluição incorreta
- ☐ Dose administrada incorretamente
- ☐ Medicamento dispensado incorretamente ou não dispensado
- ☐ Medicamento não administrado
- ☐ Paciente com alergia não considerada
- ☐ Paciente recebeu o medicamento incorreto
- ☐ Velocidade de infusão incorreta
- ☐ Via de administração incorreta
- ☐ Outros

Resumo do Evento:

OK Cancela



# Underreported Events

---

Events are notified voluntarily in the system

Source of useful information to improve care and subsidize continuing education!

Due to lack of knowledge, forgetfulness or lack of time, it is underreported. Only 10-20% of these events are reported.



# IHI Global Trigger Tool

---



“This tool includes a list of known AE triggers as well as **instructions for selecting records**, training information, and appendices with references and common questions.”





# Problems and Objectives

---



Detect falls incidents in  
Progress Notes



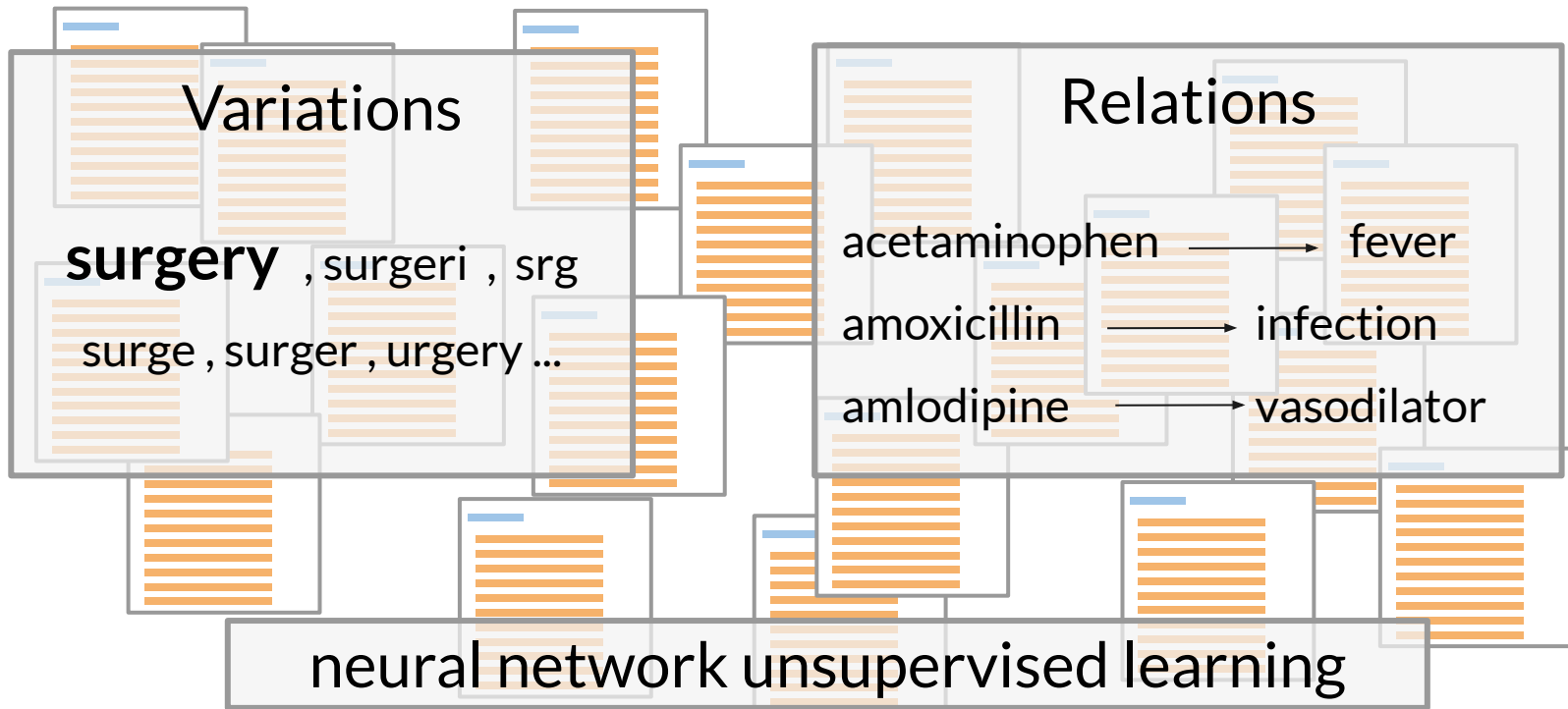
Evaluate language models  
performance using LSTM in this task

# Methods and Materials

---

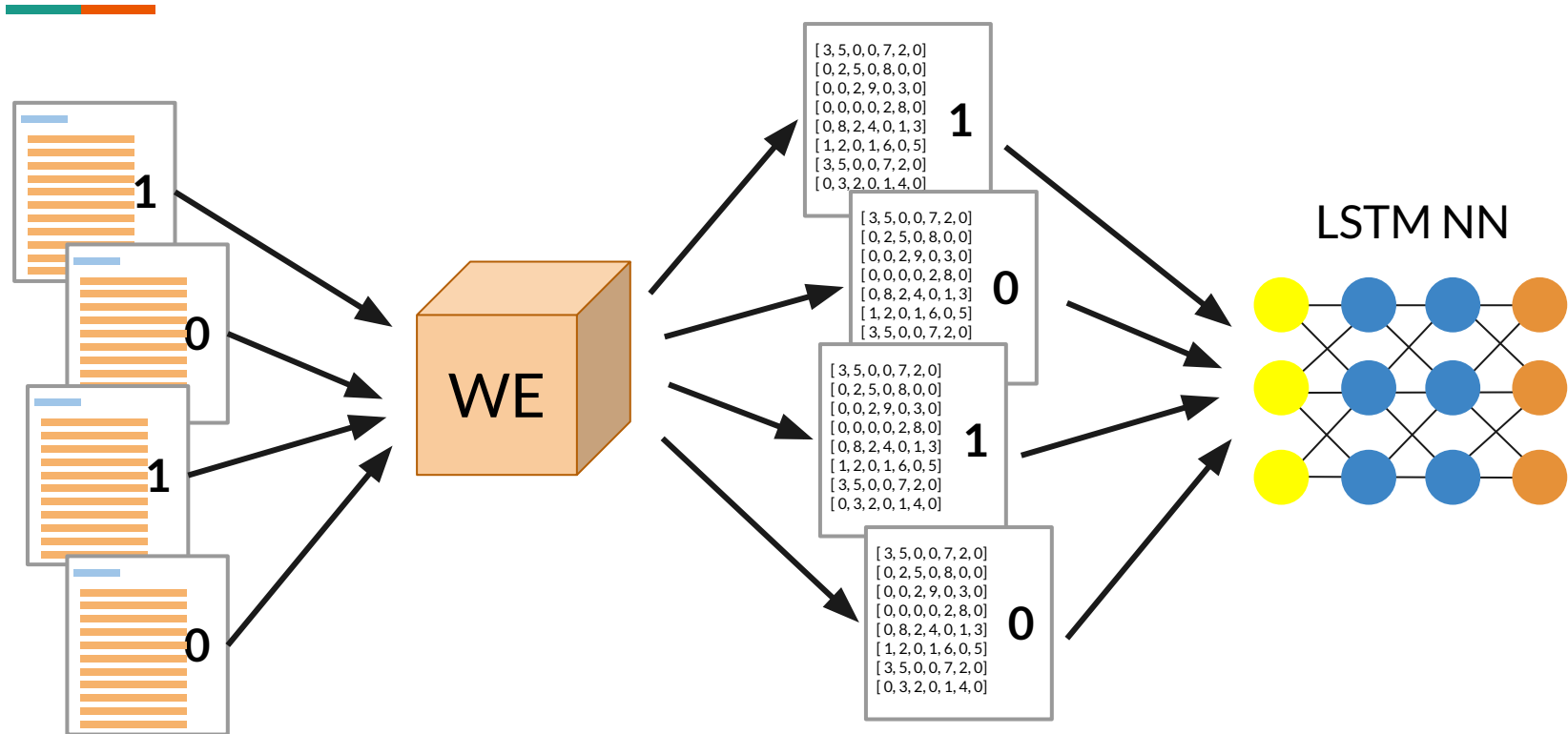


# Word Embeddings



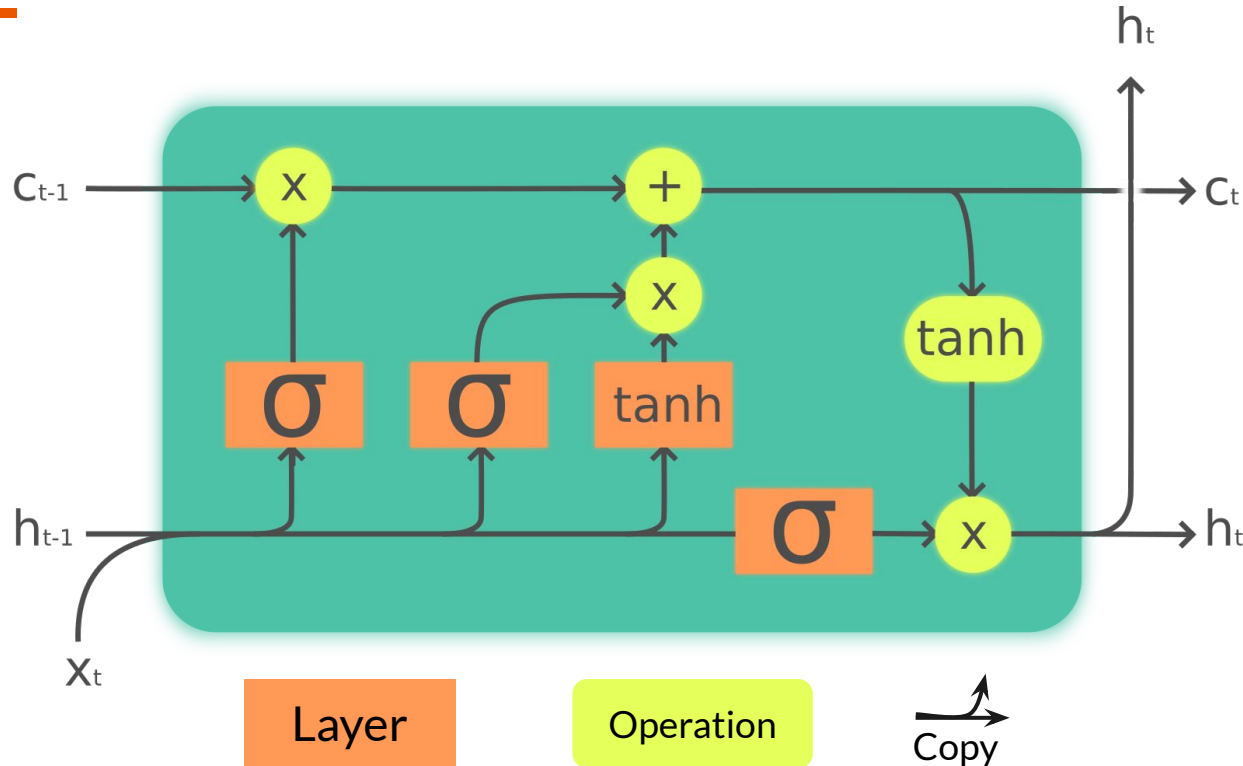


# Word Embeddings Encoding





# LSTM Neuron (Long Short-Term Memory)





# Annotation Process



441

Incident Reports



# Annotation Process



441

Incident Reports



723  
(68%)

335  
(32%)

0

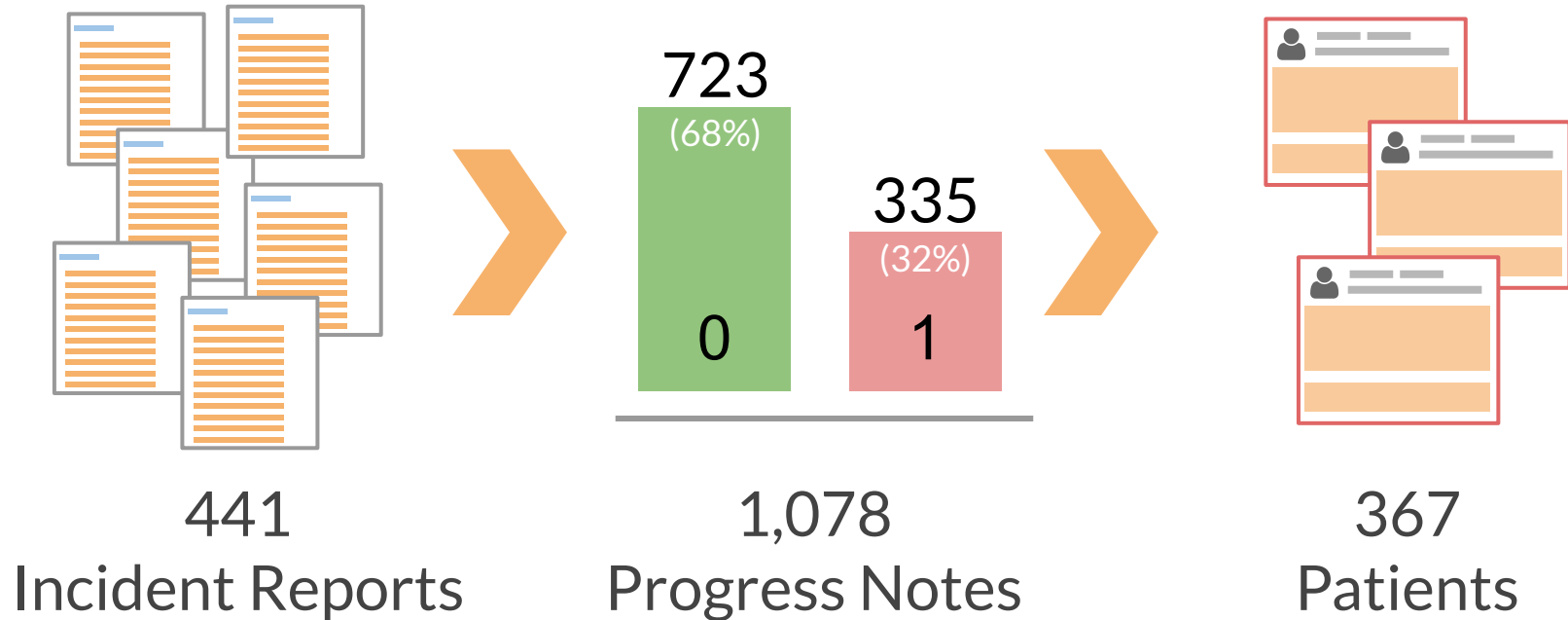
1

1,078

Progress Notes



# Annotation Process







# Annotation Process

## Fall per Patient in Annotated Dataset

# of Patients	% of Total	# of Falls
316	87.0 %	1 fall
36	10.0%	2 falls
11	3.0 %	3 falls
1	0.3 %	4 falls
...	...	...



# Language Models

---

## Wikipedia-PT

- Wikipedia dump of May 2019
- 250 million tokens

## NILC (broad-domain)

- Wikipedia + 20 sources (News and Researches)
- 1.3 billion tokens

## EHR-Notes (biomedical-domain)

- 24 million sentences from Progress Notes
- 603 million tokens



# Fall Detection - Experiments

---

Annotation of 1,078 Progress Notes

Evaluation of three Word Embedding Models

- Wikipedia-PT
- NILC (Wikipedia + 20 sources)
- EHR-Notes (Progress Notes)

State-of-the-art NLP Neural Network (LSTM)

**Baseline:** SVM and Random Forest

# Results ...

---



# Fall Detection - Results

Embeddings 5-Fold Cross Validation: F-Measure

	Word2Vec	FastText
Wikipedia	$0.88 \pm 0.14$	$0.87 \pm 0.11$
NILC (broad-domain)	$0.77 \pm 0.06$	$0.89 \pm 0.13$
EHR (biomedical-domain)	$0.88 \pm 0.14$	<b><math>0.90 \pm 0.13</math></b>



# Fall Detection - Results

Embeddings 5-Fold Cross Validation: F-Measure

	Word2Vec	FastText
Wikipedia	$0.88 \pm 0.14$	$0.87 \pm 0.11$
NILC (broad-domain)	$0.77 \pm 0.06$	$0.89 \pm 0.13$
EHR (biomedical-domain)	$0.88 \pm 0.14$	<b><math>0.90 \pm 0.13</math></b>
Random Forest	$0.73 \pm 0.03$	
SVM	$0.60 \pm 0.05$	



# Conclusion



Better with biomedical-domain  
Language Model

LSTM prove to be best than  
Random Forest



# Research Limitations

---

- Progress Notes Sampling (natural distribution)
- Self-Attention Neural Network
- BERT, GPT-2, XLNet Embeddings
- Quality Evaluation (why right, why wrong)





# Source Code

- Experiments scripts
- Annotated Dataset (1,078 records)
- Pre-trained Word Embeddings

<https://github.com/nlp-pucrs/fall-detection>

# Further Work ...

---



# Further Work

Sequence Tagging Task : F-Measure

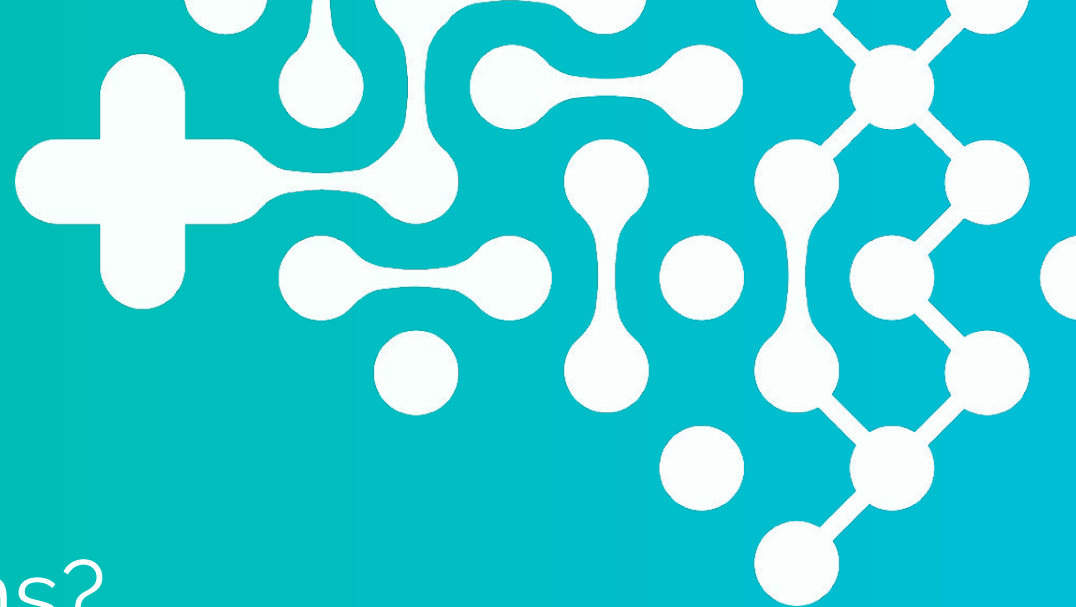
	LSTM + CRF
Wikipedia	0.55
NILC (broad-domain)	0.67
EHR (specific-domain)	<b>0.80</b>



# Further Work

## Quality Embeddings - Analogies Task

	Specific	Broad
Wikipedia	1.38 %	79.00 %
NILC (broad-domain)	2.61 %	<b>82.38 %</b>
EHR (specific-domain)	<b>2.85 %</b>	0.00 %



# Thanks! Questions?

<http://www.inf.pucrs.br/ia-saude/>

**Henrique Dias**

henrique.santos.003@acad.pucrs.br



# Annotation Process - WebAnno

The screenshot displays the WebAnno web-based annotation interface. At the top, there are five main toolbars: Document (Open, Prev, Next, Export, Settings), Page (First, Prev, Go to, Next, Last), Script (LTR/RTL), Help (Guidelines), and Workflow (Reset, Finish). Below these, the document path is shown as 'to Adverse - Evolução/21286370\_2017-09-24.txt' and the status 'Showing 1-20 of 40 sentences [document 599 of 1,971]'.

The main text area contains a document with several paragraphs of medical text. Annotations are visible as colored boxes and labels above the text. For example, 'Tipo de Queda II - Própria Altura' and 'Condições do paciente antes da queda' are highlighted in green. Other labels include 'Fator de Risco - Intrínseco', 'Refeita unidade de internação. Qual local?', 'Tipo de Queda II - Outro', 'Fator de Risco - Comportamental', and 'Orientações para prevenção ANTES da queda?'. The text itself includes details like 'Registro Paciente: 21286370.0', 'Data da Queda: 2017-09-24', 'Hora da Queda: 18:30:00', 'Gênero: Masculino', 'Idade: 33.0', 'UH: HCO', and a summary of the patient's condition and fall.

On the right side, there is a sidebar with controls. The 'Layer' dropdown is set to 'Fatores de Queda'. Below it, there is a 'Forward annotation' checkbox. The 'Annotation' section has a red 'Delete' button and a 'Clear' button. The 'Text' field contains the text 'Destaco que os pacientes estão sem a proteção das grades devido suas'. The 'value' dropdown is set to 'Fator de Risco - Extrínseco', and a list of other values is visible below it, including 'Condições do paciente antes da queda', 'Dano não classificado adequadamente', 'Dúvida para Janete', 'Fator de Risco - Comportamental', and 'Fator de Risco - Extrínseco'.