# Integration of machine learning- and dictionary-based approach for identification of adverse drug reactions in drug labels

**Junguk Hur**

**University of North Dakota**
**School of Medicine and Health Sciences**
**hurlab.med.und.edu**

UND UNIVERSITY OF NORTH DAKOTA
SCHOOL OF MEDICINE AND HEALTH SCIENCES

# Team: CONDL

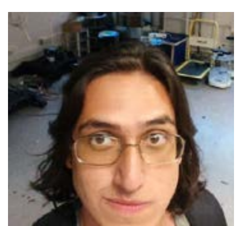
- **C**entrality- and **O**ntology-based **N**etwork **D**iscovery using **L**iterature data


- Mert Tiftikci[1], Arzucan Özgür[1], Yongqun (Oliver) He[2], and Junguk Hur[3]

  [1]Bogazici University, Istanbul, Turkey

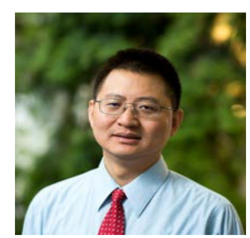
  [2]University of Michigan, Ann Arbor, MI, USA

  [3]University of North Dakota, Grand Forks, ND, USA

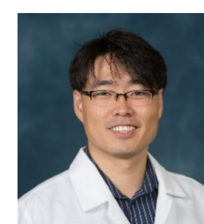



**Mert**

**Arzucan**



**Oliver**



**Junguk**

UND UNIVERSITY OF NORTH DAKOTA
SCHOOL OF MEDICINE AND HEALTH SCIENCES

# Outline

- Background
  - Adverse drug reactions
- Our approach & results
  - Mention Extraction from drug label (Deep learning / SciMiner)
  - ADR normalization (SciMiner)
- Summary & discussion

# Adverse Drug Reaction (ADR)

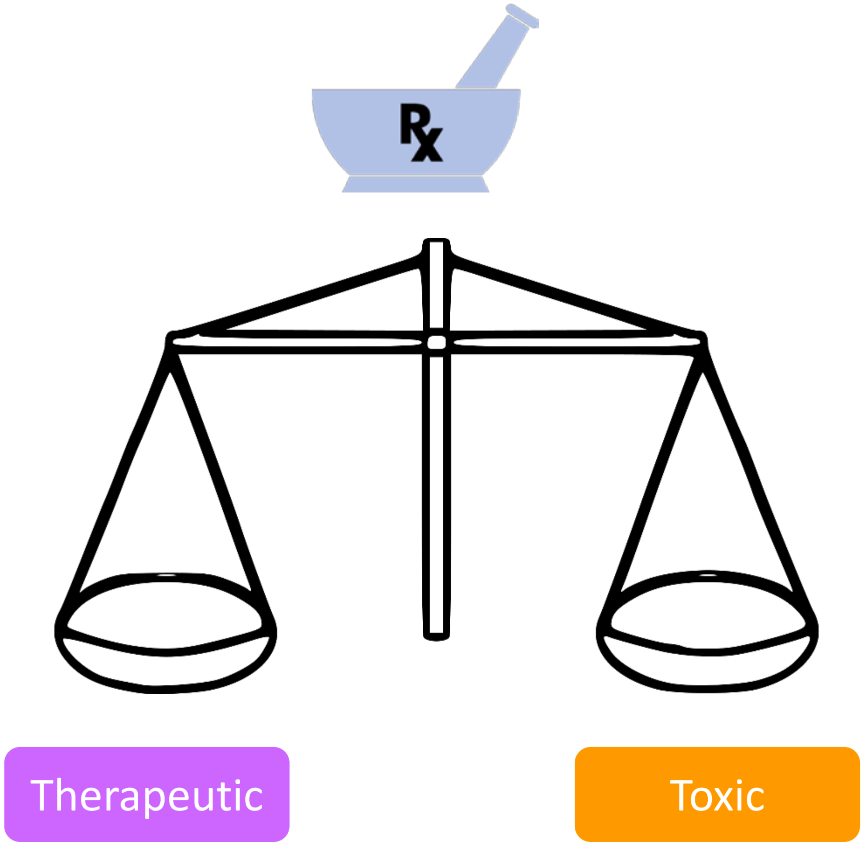




Image from BioJobBlog.com

# Resources for ADR

- Drug labels (prescribing information or package inserts)
  - Drugs@FDA database
  - SIDER4.1 database
- Post-marketing
  - FDA's Adverse Event Reporting System (FAERS)
  - Database of Suspected Adverse Drug Reaction (EDSADR)

**HIGHLIGHTS OF PRESCRIBING INFORMATION**
**These highlights do not include all the information needed to use VELCADE safely and effectively. See full prescribing information for VELCADE.**

**VELCADE (bortezomib) for Injection**
**Initial US Approval: 2003**

**-----RECENT MAJOR CHANGES-----**

Patients with Renal Impairment (8.6) 10/2007
Indications and Usage, Multiple Myeloma (1.1) 06/2008

**-----INDICATIONS AND USAGE-----**

VELCADE is a proteasome inhibitor indicated for:
- treatment of patients with multiple myeloma (1.1)
- treatment of patients with mantle cell lymphoma who have received at least 1 prior therapy (1.2)

**-----DOSAGE AND ADMINISTRATION-----**

The recommended dose of VELCADE is 1.3 mg/m$^2$ administered as a 3 to 5 second bolus intravenous injection. (2.1, 2.3)
Dose adjustment may be used to manage adverse events that occur during treatment (2.2, 2.4)

**-----DOSAGE FORMS AND STRENGTHS-----**

- 1 single use vial contains 3.5 mg of bortezomib. Dose must be individualized to prevent overdose. (3)

**-----CONTRAINDICATIONS-----**

- VELCADE is contraindicated in patients with hypersensitivity to bortezomib, boron, or mannitol. (4)

**-----WARNINGS AND PRECAUTIONS-----**

- Women should avoid becoming pregnant while being treated with VELCADE. Pregnant women should be apprised of the potential harm to the fetus. (5.1, 8.1)
- Peripheral neuropathy, including severe cases, may occur - manage with dose modification or discontinuation. (2.2, 2.4) Patients with preexisting severe neuropathy should be treated with VELCADE only after careful risk-benefit assessment. (2.2, 2.4, 5.2)
- Hypotension can occur. Caution should be used when treating patients receiving antihypertensives, those with a history of syncope, and those who are dehydrated. (5.3)
- Patients with risk factors for, or existing heart disease, should be closely monitored. (5.4)
- Acute diffuse infiltrative pulmonary disease has been reported. (5.5)
- Nausea, diarrhea, constipation, and vomiting have occurred and may require use of antiemetic and antidiarrheal medications or fluid replacement. (5.7)
- Thrombocytopenia or neutropenia can occur; complete blood counts should be regularly monitored throughout treatment. (5.8)
- Tumor Lysis Syndrome (5.9), Reversible Posterior Leukoencephalopathy Syndrome (5.6), and acute hepatic failure (5.10) have been reported.

**-----ADVERSE REACTIONS-----**

Most commonly reported adverse reactions (incidence ≥30%) in clinical studies include asthenic conditions, diarrhea, nausea, constipation, peripheral neuropathy, vomiting, pyrexia, thrombocytopenia, psychiatric disorders, anorexia and decreased appetite, neutropenia, neuralgia, leukopenia and anemia. Other adverse reactions, including serious adverse reactions, have been reported. (6.1)
**To report SUSPECTED ADVERSE REACTIONS, contact Millennium Pharmaceuticals at (1-866 VELCADE or FDA at 1-800-FDA-1088 or *www.fda.gov/medwatch*.**

**-----USE IN SPECIFIC POPULATIONS-----**

- Women should be advised against breast feeding or becoming pregnant while being treated with VELCADE. (5.1, 8.1, 8.3)
- Patients with diabetes may require close monitoring of blood glucose and adjustment of anti-diabetic medication. (8.8)

**See 17 for PATIENT COUNSELING INFORMATION.**

**Revised: [06/2008]**

**Table 5-Most Commonly Reported Adverse Events (≥ 10% in VELCADE, Melphalan and Prednisone arm) with Grades 3 and ≥4 Intensity in the Previously Untreated Multiple Myeloma Study**

| | VELCADE, Melphalan and Prednisone (N=340) | | | Melphalan and Prednisone (N=337) | | |
|---|---|---|---|---|---|---|
| MedDRA System Organ Class | Total | Toxicity Grade, n (%) | | Total | Toxicity Grade, n (%) | |
| Preferred Term | n (%) | 3 | ≥4 | n (%) | 3 | ≥4 |
| **Blood and Lymphatic System Disorders** | | | | | | |
| Thrombocytopenia | 178 (52) | 68 (20) | 59 (17) | 159 (47) | 55 (16) | 47 (14) |
| Neutropenia | 165 (49) | 102 (30) | 35 (10) | 155 (46) | 79 (23) | 49 (15) |
| Anemia | 147 (43) | 53 (16) | 9 (3) | 187 (55) | 66 (20) | 26 (8) |
| Leukopenia | 113 (33) | 67 (20) | 10 (3) | 100 (30) | 55 (16) | 13 (4) |
| Lymphopenia | 83 (24) | 49 (14) | 18 (5) | 58 (17) | 30 (9) | 7 (2) |
| **Gastrointestinal Disorders** | | | | | | |
| Nausea | 164 (48) | 14 (4) | 0 | 94 (28) | 1 (<1) | 0 |
| Diarrhea | 157 (46) | 23 (7) | 2 (1) | 58 (17) | 2 (1) | 0 |
| Constipation | 125 (37) | 2 (1) | 0 | 54 (16) | 0 | 0 |
| Vomiting | 112 (33) | 14 (4) | 0 | 35 (10) | 2 (1) | 0 |
| Abdominal Pain | 49 (14) | 7 (2) | 0 | 22 (7) | 1 (<1) | 0 |
| Abdominal Pain Upper | 40 (12) | 1 (<1) | 0 | 29 (9) | 0 | 0 |
| Dyspepsia | 39 (11) | 0 | 0 | 23 (7) | 0 | 0 |
| **Nervous System Disorders** | | | | | | |
| Peripheral Neuropathy | 159 (47) | 43 (13) | 2 (1) | 18 (5) | 0 | 0 |
| Neuralgia | 121 (36) | 28 (8) | 2 (1) | 5 (1) | 1 (<1) | 0 |
| Dizziness | 56 (16) | 7 (2) | 0 | 37 (11) | 1 (<1) | 0 |
| Headache | 49 (14) | 2 (1) | 0 | 35 (10) | 4 (1) | 0 |
| Paresthesia | 45 (13) | 6 (2) | 0 | 15 (4) | 0 | 0 |
| **General Disorders and Administration Site Conditions** | | | | | | |
| Pyrexia | 99 (29) | 8 (2) | 2 (1) | 64 (19) | 6 (2) | 2 (1) |
| Fatigue | 98 (29) | 23 (7) | 2 (1) | 86 (26) | 7 (2) | 0 |
| Asthenia | 73 (21) | 20 (6) | 1 (<1) | 60 (18) | 9 (3) | 0 |
| Edema Peripheral | 68 (20) | 2 (1) | 0 | 34 (10) | 0 | 0 |
| **Infections and Infestations** | | | | | | |
| Pneumonia | 56 (16) | 16 (5) | 13 (4) | 36 (11) | 13 (4) | 9 (3) |
| Herpes Zoster | 45 (13) | 11 (3) | 0 | 14 (4) | 6 (2) | 0 |
| Bronchitis | 44 (13) | 4 (1) | 0 | 27 (8) | 4 (1) | 0 |
| Nasopharyngitis | 39 (11) | 1 (<1) | 0 | 27 (8) | 0 | 0 |


**Table 5-Most Commonly Reported Adverse Events (≥ 10% in VELCADE, Melphalan and Prednisone arm) with Grades 3 and ≥4 Intensity in the Previously Untreated Multiple Myeloma Study**

| | VELCADE, Melphalan and Prednisone (N=340) | | | Melphalan and Prednisone (N=337) | | |
|---|---|---|---|---|---|---|
| MedDRA System Organ Class | Total | Toxicity Grade, n (%) | | Total | Toxicity Grade, n (%) | |
| Preferred Term | n (%) | 3 | ≥4 | n (%) | 3 | ≥4 |
| **Blood and Lymphatic System Disorders** | | | | | | |
| Thrombocytopenia | 178 (52) | 68 (20) | 59 (17) | 159 (47) | 55 (16) | 47 (14) |
| Neutropenia | 165 (49) | 102 (30) | 35 (10) | 155 (46) | 79 (23) | 49 (15) |
| Anemia | 147 (43) | 53 (16) | 9 (3) | 187 (55) | 66 (20) | 26 (8) |
| Leukopenia | 113 (33) | 67 (20) | 10 (3) | 100 (30) | 55 (16) | 13 (4) |
| Lymphopenia | 83 (24) | 49 (14) | 18 (5) | 58 (17) | 30 (9) | 7 (2) |
| **Gastrointestinal Disorders** | | | | | | |
| Nausea | 164 (48) | 14 (4) | 0 | 94 (28) | 1 (<1) | 0 |
| Diarrhea | 157 (46) | 23 (7) | 2 (1) | 58 (17) | 2 (1) | 0 |
| Constipation | 125 (37) | 2 (1) | 0 | 54 (16) | 0 | 0 |
| Vomiting | 112 (33) | 14 (4) | 0 | 35 (10) | 2 (1) | 0 |
| Abdominal Pain | 49 (14) | 7 (2) | 0 | 22 (7) | 1 (<1) | 0 |
| Abdominal Pain Upper | 40 (12) | 1 (<1) | 0 | 29 (9) | 0 | 0 |
| Dyspepsia | 39 (11) | 0 | 0 | 23 (7) | 0 | 0 |
| **Nervous System Disorders** | | | | | | |
| Peripheral Neuropathy | 159 (47) | 43 (13) | 2 (1) | 18 (5) | 0 | 0 |
| Neuralgia | 121 (36) | 28 (8) | 2 (1) | 5 (1) | 1 (<1) | 0 |
| Dizziness | 56 (16) | 7 (2) | 0 | 37 (11) | 1 (<1) | 0 |
| Headache | 49 (14) | 2 (1) | 0 | 35 (10) | 4 (1) | 0 |
| Paresthesia | 45 (13) | 6 (2) | 0 | 15 (4) | 0 | 0 |
| **General Disorders and Administration Site Conditions** | | | | | | |
| Pyrexia | 99 (29) | 8 (2) | 2 (1) | 64 (19) | 6 (2) | 2 (1) |
| Fatigue | 98 (29) | 23 (7) | 2 (1) | 86 (26) | 7 (2) | 0 |
| Asthenia | 73 (21) | 20 (6) | 1 (<1) | 60 (18) | 9 (3) | 0 |
| Edema Peripheral | 68 (20) | 2 (1) | 0 | 34 (10) | 0 | 0 |
| **Infections and Infestations** | | | | | | |
| Pneumonia | 56 (16) | 16 (5) | 13 (4) | 36 (11) | 13 (4) | 9 (3) |
| Herpes Zoster | 45 (13) | 11 (3) | 0 | 14 (4) | 6 (2) | 0 |
| Bronchitis | 44 (13) | 4 (1) | 0 | 27 (8) | 4 (1) | 0 |
| Nasopharyngitis | 39 (11) | 1 (<1) | 0 | 27 (8) | 0 | 0 |


Parts of drug label for Velcade (bortezomib)

# Importance of label mining

- All about safety
- From unpredictable to predictable events
- Personalized medicine

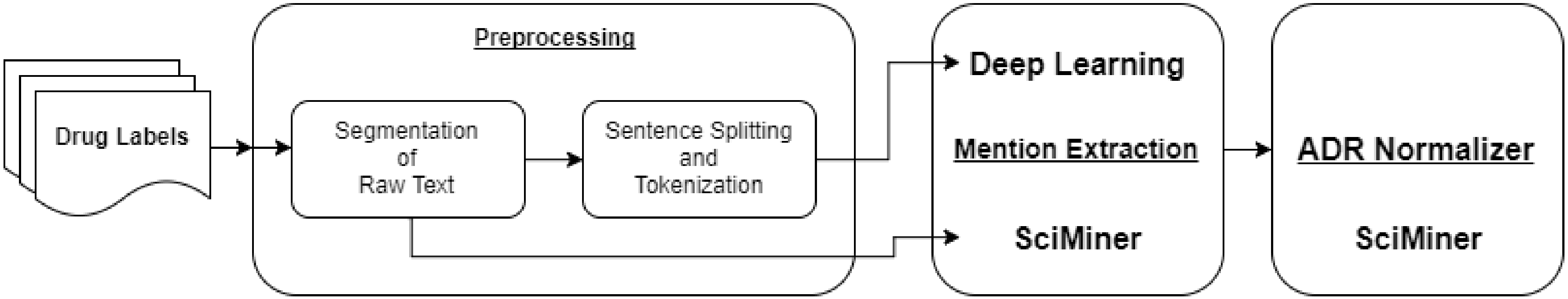
- Automatic extraction of ADRs from drug labels
  - comparing the ADRs present in labels from different manufacturers for the same drug
  - performing post-marketing safety analysis (pharmacovigilance) by identifying new ADRs not currently present in the labels
  - to improve the efficiency of this process, the extraction of the ADRs from the drug labels needs to be automated

# Goals

(1) To develop text mining system of mentions (ADR, drug class, animal, severity, factor, and negation) from drug labels (Task#1)

(2) To normalize extracted ADRs onto MedDRA Preferred Terms (PTs) (Task#4)

# Our Workflow




- **Deep Learning (DL) model** works on vector representation of tokens of sentences
  - Rule-base text segmentation applied on raw text
  - Text segments split to sentences & Sentences tokenized[1]
- **Dictionary- and Rule-based SciMiner** for mention extraction and normalizing detected ADRs

[1)]NLTK package for sentence splitting and tokenization

# DL - Preprocessing

**Raw Text from label APTIOM**

* Suicidal Behavior and Ideation [see Warnings and Precautions ( 5.1 )]

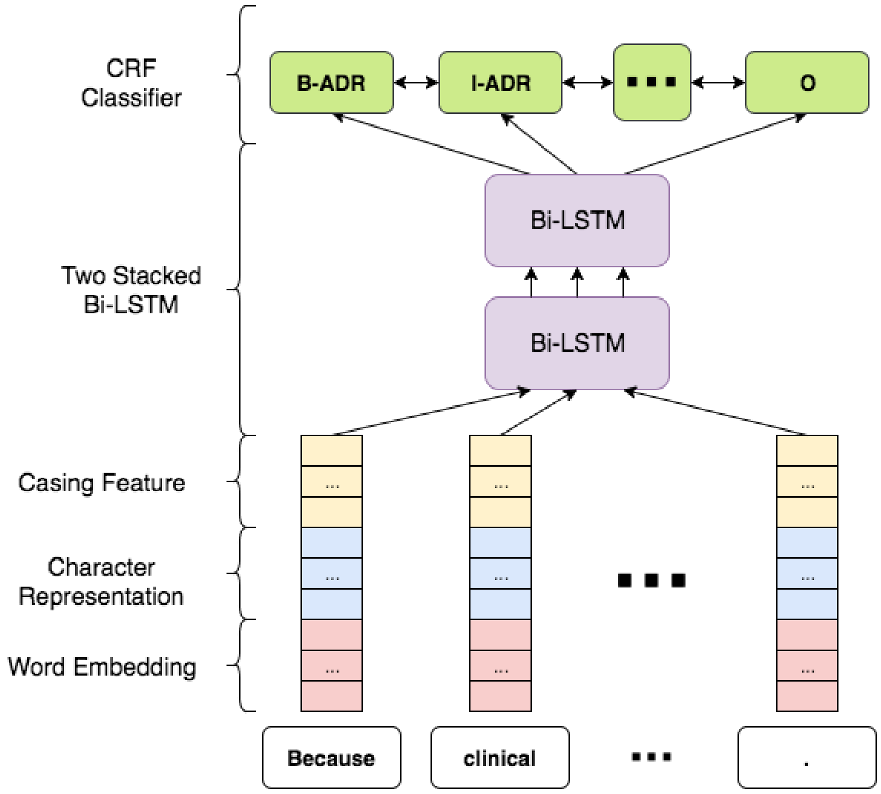
**Mentions (Overlapping and non-contiguous example)**

<**Mention id**="M1" **section**="S1" **type**="AdverseReaction"
**start**="151" **len**="17" **str**="Suicidal Behavior" />
<**Mention id**="M2" **section**="S1" **type**="AdverseReaction”
**start**="151,173" **len**="8,8" **str**="Suicidal Ideation" />

**CoNLL Format**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| * | O | NN | S1 | 148 | 1 | Warnings | O | NNP | S1 | 187 | 8 |
| Suicidal | B-ADR | NNP | S1 | 151 | 17 | and | O | CCP | S1 | 196 | 3 |
| Behavior | I-ADR | NNP | S1 | 160 | 8 | Precautions | O | NNP | S1 | 200 | 11 |
| and | O | CC | S1 | 169 | 3 | ( | O | ( | S1 | 212 | 1 |
| Ideation | I-ADR | NNP | S1 | 173 | 8 | 5.1 | O | CD | S1 | 215 | 3 |
| [ | O | NNP | S1 | 182 | 1 | ) | O | ) | S1 | 220 | 1 |
| see | O | VBP | S1 | 183 | 3 | ] | O | NN | S1 | 221 | 1 |

# Deep Learning Architecture
# Bi-directional LSTM-CNNs-CRF



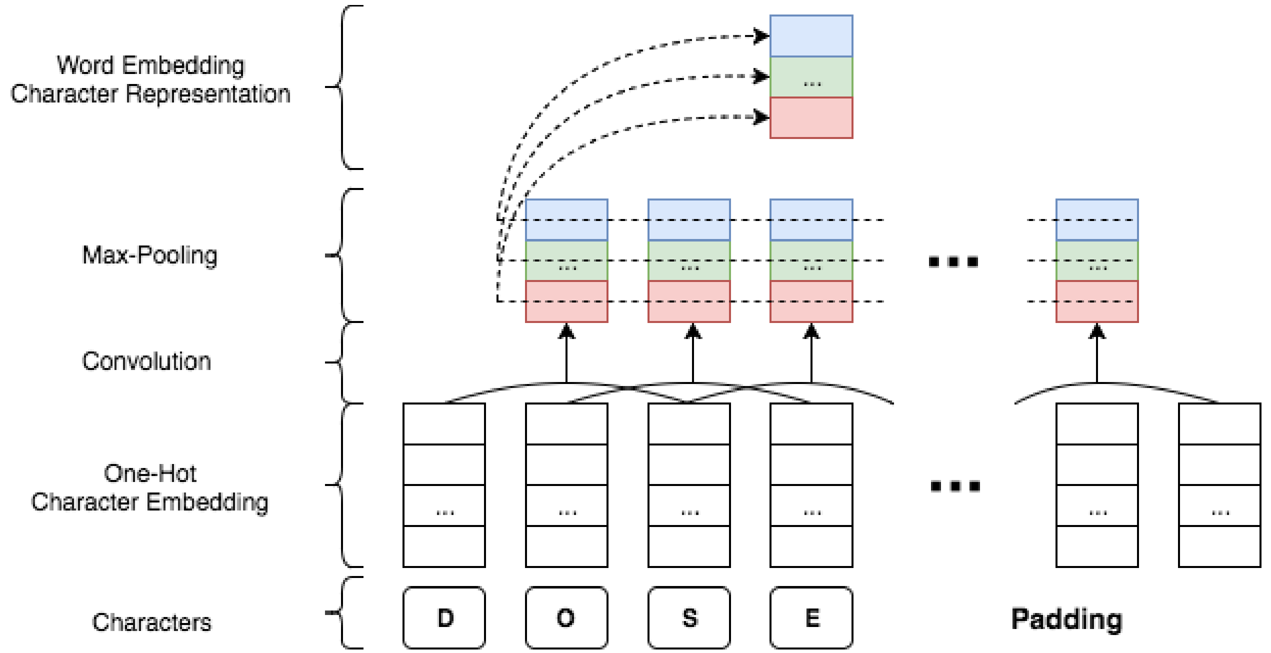
Neural Network Architecture

- Combined Word Embeddings (CWE) are generated for each token of a given sentence
- First Bi-directional long short-term memory LSTM runs on CWEs and second LSTM runs on the output of the first one.
- Conditional Random Fields (CRF) classifier jointly decodes as mention predictions for each token.
- **Keras2 library** was used in our work. **No early stopping** was used in our work.

UND UNIVERSITY OF NORTH DAKOTA
SCHOOL OF MEDICINE AND HEALTH SCIENCES

- This model is an adaptation of implementation for paper [Nils Reimers, and Iryna Gurevych. "Reporting score distributions makes a difference: Performance study of lstm-networks for sequence tagging." *arXiv preprint arXiv:1707.09861* (2017)]

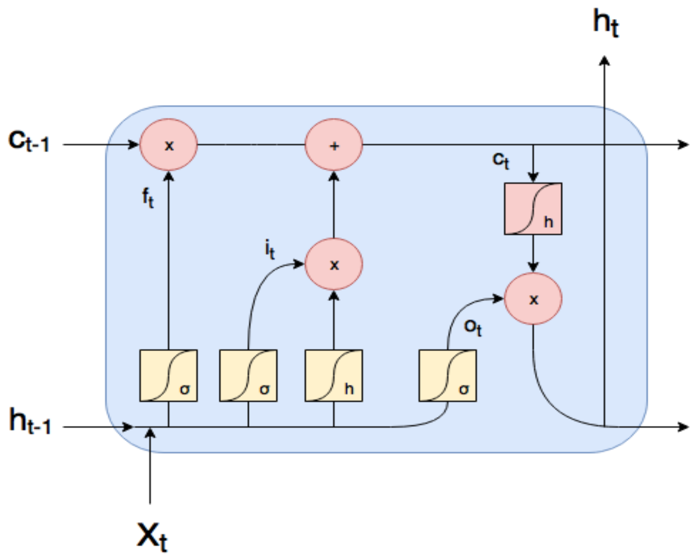
# Combined Word Embeddings




- CWEs are created from the concatenation
  - Character Embedding (Generated by CNN)
  - Word Embedding (Generated by Word2Vec) – based on **PubMed (200D)**
  - Casing Embedding (one-hot encoded)

UNIVERSITY OF NORTH DAKOTA
SCHOOL OF MEDICINE AND HEALTH SCIENCES

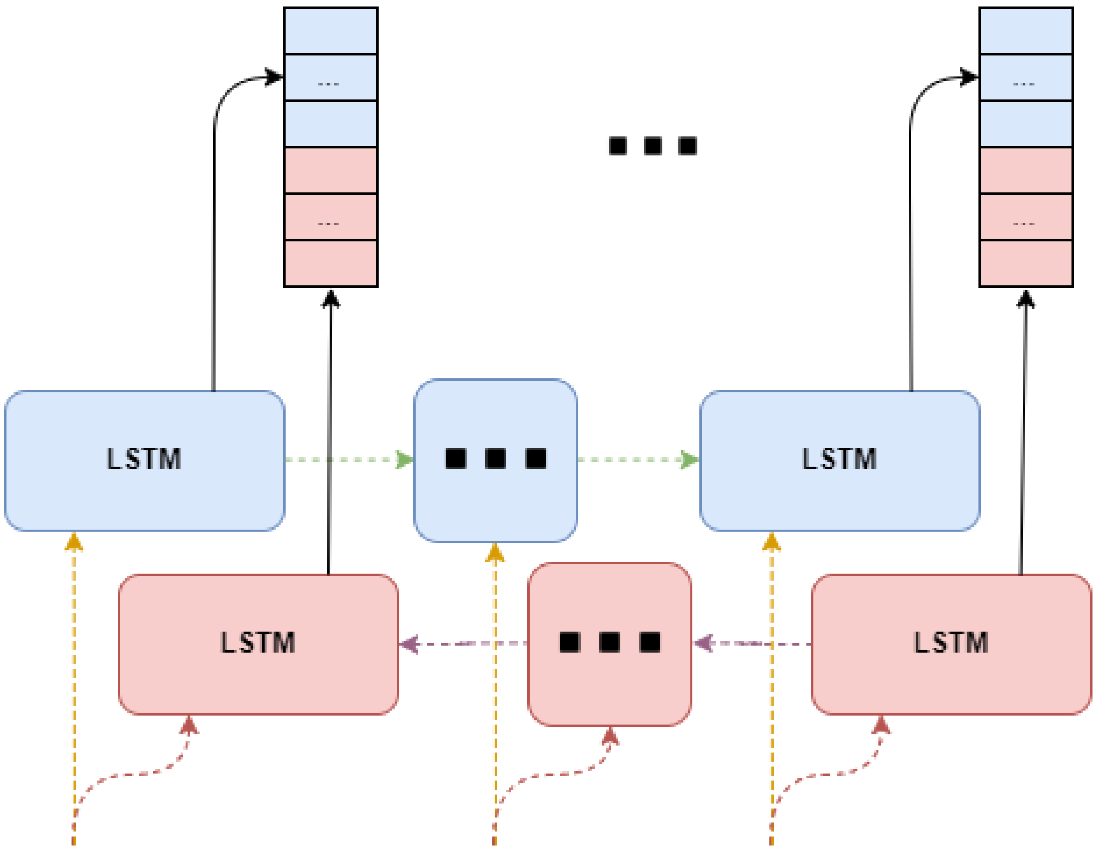
# LSTM component




$$i_t = \alpha(W_i[h_{t-1}, x_t] + b_i)$$
$$f_t = \alpha\left(W_f[h_{t-1}, x_t] + b_f\right)$$
$$\tilde{c}_t = tanh(W_c[h_{t-1}, x_t] + b_c)$$
$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t$$
$$o_t = \alpha(W_o[h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(c_t)$$

S. Hochreiter and J. Schmidhuber

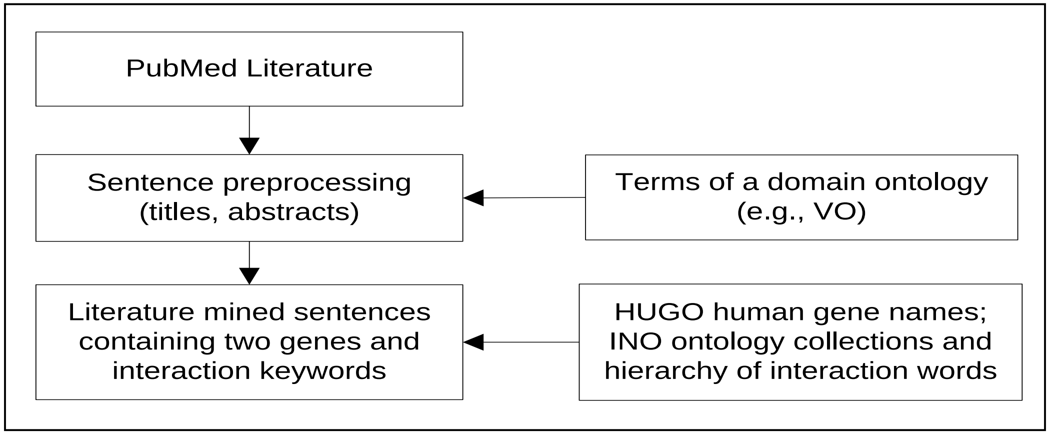
# Bi-LSTM component with Variational Dropout




*Variational dropout (0.25) depicted by colored & dashed lines*

# SciMiner

- SciMiner: A web-based literature mining tool for (http://hurlab.med.und.edu/SciMiner/)
- Dictionary- and Rule-based mining
- Optimized for identifying genes/proteins and VO/INO/EColi ontology terms





**References:**

- Hur J, Schuyler AD, States DJ, Feldman EL: SciMiner: web-based literature mining tool for target identification and functional enrichment analysis. *Bioinformatics* 2009, 25(6):838-840.
- Hur J, Xiang Z, Feldman EL, He Y. Ontology-based *Brucella* vaccine literature indexing and systematic analysis of gene-vaccine association network. *BMC Immunology*. 12(1):49 2011 Aug 26. PMID: 21871085.
- Hur J, Ozgur A, and He Y: *Ontology-based literature mining of E. coli vaccine-associated gene interaction networks.* J Biomed Semantics, vol. 8, p. 12,


UNIVERSITY OF NORTH DAKOTA
SCHOOL OF MEDICINE AND HEALTH SCIENCES

# ADR-SciMiner

- Expanded SciMiner for ADRs identification
- Dictionaries compiled from MedDRA (v20.0 English)
- Term expansion rules for improved coverage
  - Lingua::EN Perl library
  - Token order
  - Casing information (eg. all vs ALL - leukaemia)
  - Alternative terms: (eg. increase -> elevation)
- Some exclusions criteria
  - Disease/syndrome names and etc
  - Section titles
- Currently, only for **ADR** terms

# Our submissions

| Set | Mentions (Task 1) | ADR Normalization (Task 4) |
|---|---|---|
| **CONDL1** | DL | ADR-SciMiner |
| **CONDL2** | ADR-SciMiner (ADR) | ADR-SciMiner |
| **CONDL3** | ADR-SciMiner (ADR) + non-ADRs from DL | ADR-SciMiner |

UND UNIVERSITY OF NORTH DAKOTA
SCHOOL OF MEDICINE AND HEALTH SCIENCES

# Results

| | | | CONDL1 | CONDL2 | CONDL3 |
|---|---|---|---|---|---|
| Task 1 | | | Deep Learning | SciMiner | SciMiner + non-ADRs from DL |
| | +type | Precision | 76.5 | 65.5 | 65.2 |
| | | Recall | 77.5 | 61.4 | 69.8 |
| | | F1 | 77.0 | 63.4 | 67.4 |
| | -type | Precision | 76.5 | 65.5 | 65.2 |
| | | Recall | 77.5 | 61.4 | 69.8 |
| | | F1 | 77.0 | 63.4 | 67.4 |
| Task 4 | | | SciMiner | SciMiner | SciMiner |
| | micro | Precision | **88.8** | 74.6 | 74.6 |
| | | Recall | 77.2 | **81.0** | **81.0** |
| | | F1 | **82.6** | 77.6 | 77.6 |
| | macro | Precision | **88.2** | 73.1 | 73.1 |
| | | Recall | 75.8 | **79.9** | **79.9** |
| | | F1 | **80.5** | 75.6 | 75.6 |

<u>Our results on the TAC ADR testing data (99 drug labels)</u>

CONDL1 (DL+SciMiner): Precision (88.8 / 88.2) – 1st place among 12 submissions in Task#4
F1 (82.6 / 80.5) – 4th place

UNIVERSITY OF NORTH DAKOTA
SCHOOL OF MEDICINE AND HEALTH SCIENCES

# Summary

- Deep learning adaptation (Bi-directional LSTM-CNNs-CRF)
- Dictionary- and Rule-based ADR-SciMiner for ADR extraction and normalization
- Combined system
- Still, much room for improvement

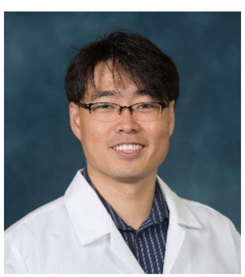
# Future Work

- Performance improvement of DL
  - Better representation for overlapping & non-contiguous chunks
- Performance improvement of ADR-SciMiner
  - Severity of ADR
  - Improved rules
  - Additional dictionary including SNOMED CT
- Better integration

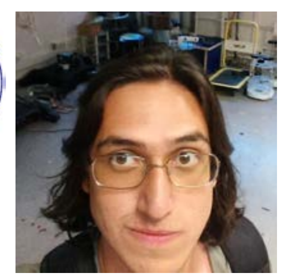
# Acknowledgements


## Funding:

- University of North Dakota, Epigenomics COBRE (NIGMS P20GM104360) (to JH).
- Marie Curie FP7-Reintegration-Grants within the 7th European Community Framework Programme (to AO)
- R01AI081062 from the US NIH NIAID (to YH)

hurlab.med.und.edu



www.cmpe.boun.edu.tr/~ozgur/

www.hegroup.org

# Thank you

UNIVERSITY OF NORTH DAKOTA
SCHOOL OF MEDICINE AND HEALTH SCIENCES