

# AI in Research: Will AI Write Your Manuscripts?

**Tanmay Basu**

**Head, Department of Data Science and Engineering  
IISER Bhopal**

**[tanmay@iiserb.ac.in](mailto:tanmay@iiserb.ac.in)**



# Cat or Dog?



Source: <https://www.theguardian.com/science/2017/jul/26/cats-vs-dogs-in-terms-of-evolution-are-we-barking-up-the-wrong-tree> Last accessed on July 16, 2024

# Cat or Tiger?



Source: <https://www.iflscience.com/new-species-of-super-cute-tiger-cat-discovered-with-an-uncertain-future-73812> Last accessed on July 16, 2024

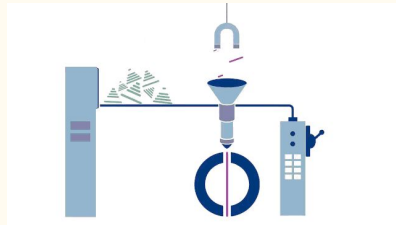
# AI Based Surveillance for Road Safety



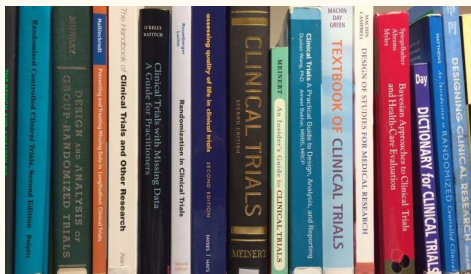
Source: <https://m.rediff.com/news/report/minor-girl-bitten-dragged-by-stray-dogs-in-mp/20220102.htm>

Last accessed on July 16, 2024

# NLP for Automatic Biomedical Text Mining



# Fuels: Clinical and Biomedical Data



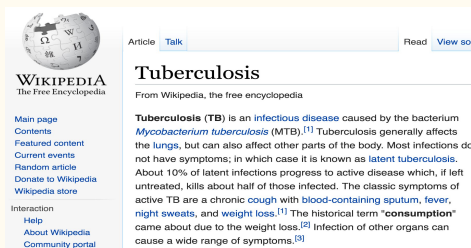
Publications



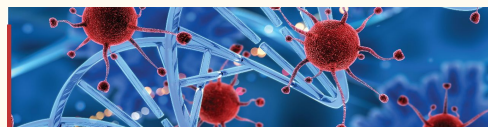
Medical Records



Social Media



Wikipedia



Symptoms of Blood Cancer and Why Early Detection is Important

By Dr. Renu Sharma in Cancer Care / Oncology

May 29, 2023 | 1 min read



Blood cancer, also known as haematological cancer, refers to cancer that affects the blood, bone marrow, and lymphatic system. This type of cancer is a result of abnormal growth and multiplication of blood cells, which can lead

HealthCare Blogs



Radiology Images

Image Sources: 1) <https://blogs.lshtm.ac.uk/library/2018/05/28/books-on-clinical-trials> 2) <https://www.cresceremed.com/overlook-of-medical-record-indexing-service-and-its-benefits>  
3) <https://www.reviewsxp.com/blog/best-social-media-sites> 4) <https://en.wikipedia.org/wiki/Tuberculosis>  
5) <https://www.maxhealthcare.in/blogs/blood-cancer-symptoms-early-detection-important> 6) <https://skeletalrad.org/>

Last accessed on March 07, 2024

# Fuels: Biomedical Ontologies

## Mondo Ontology

Semi-automatically constructed ontology that merges in multiple disease resources

## Human Phenotype Ontology

It provides a standardized vocabulary of phenotypic abnormalities encountered in human disease

## PhenomeNET Ontology

It can identify gene disease associations through the comparison of mutant phenotypes

## Unified Medical Language System (UMLS)

It is a large collection of controlled vocabularies used in biomedical sciences



# How Different Ontologies Work?

12 January 2009

HCR/JA/261024

Dear Mr

Clinic attendance: 5 January 2009 Clinic number: 1

Diagnoses: Type II diabetes mellitus  
Slowly progressive kidney impairment  
High blood pressure  
Anaemia

Medication: Gliclazide 80mg at lunchtime  
Atorvastatin 40mg at night  
Amlodipine 10mg once daily  
Lisinopril – **stopped completely**  
Doxazosin 2mg once daily  
Omeprazole 20mg once daily  
Rosiglitazone - **changed to Pioglitazone 30mg once daily**  
Furosemide – **restarted at 40mg once daily as needed**

It was good to see you in the clinic again today. Your kidney function today was 18% (creatinine = 306, urea = 22.4 and potassium = 7.5). Your haemoglobin anaemia test was 8.5g/dL.

We agreed that you should start retaking one Furosemide tablet per day to control the swelling and improve your breathing and once the swelling has cleared, you should take the water tablet when needed. You should also stop taking Lisinopril completely.

Some of the breathlessness may be due to anaemia, which we will be able to treat in due course.

On discussion today with Professor Barnett, we agreed that you should switch from Rosiglitazone to Pioglitazone (30mg once daily) because of current concerns about the long term safety of this drug. We will see you next in two months' time with blood tests the week before.

Yours sincerely



## Mondo Ontology

congenital heart defect  
cardiomyopathy

- .
- .
- .



## PhenomeNET Ontology

cardiac arrest  
abnormality of cardiac ventricle

- .
- .
- .



## Human Phenotype Ontology

cardiac amyloidosis  
muscular dystrophy



# Illustration: Human Phenotypes

Phenomebrowser Home Browse SPARQL Downloads About Related Databases Help

HP Search ?  
e.g. Abnormal cerebral morphology | dementia | ACAT1 | acipimox | Variola virus

**Abdominal distention** HP:0003270 [← Back](#)

Distention of the abdomen.

**Disease**  
Drug  
Gene  
Metabolite  
Pathogen  
Similar Entities

1 to 20 of 291 Disease Associations

Filter By:  Include Subclass Phenotypes datasets evidences

Disease	Evidence	Phenotype	Dataset	Source	Created on
Toxic epidermal necrolysis ⓘ	author statement supported by traceable reference	Abdominal distention	hpo disease-phenotypes	<a href="#">Link</a>	2021-10-10
obsolete gastrointestinal mucormycosis ⓘ	inferred from electronic annotation	Abdominal distention	textmined disease-phenotypes	<a href="#">Link</a>	
Mitochondrial neurogastrointestinal encephalomyopathy ⓘ	author statement supported by traceable reference	Abdominal distention	hpo disease-phenotypes	<a href="#">Link</a>	2021-10-10
diverticulitis of colon ⓘ	inferred from electronic annotation	Abdominal distention	textmined disease-phenotypes	<a href="#">Link</a>	
heterophyiasis ⓘ	inferred from electronic annotation	Abdominal distention	textmined disease-phenotypes	<a href="#">Link</a>	
gastrointestinal anthrax ⓘ	inferred from electronic annotation	Abdominal distention	textmined disease-phenotypes	<a href="#">Link</a>	
Congenital hypothyroidism due to maternal intake of antithyroid drugs ⓘ	author statement supported by traceable reference	Abdominal distention	hpo disease-phenotypes	<a href="#">Link</a>	2021-10-10
HYPOTHYROIDISM, CONGENITAL, NONGOITROUS, 2; CHNG2 ⓘ	evidence used in automatic assertion	Abdominal distention	hpo disease-phenotypes	<a href="#">Link</a>	2009-02-17

# Illustration: MedCAT

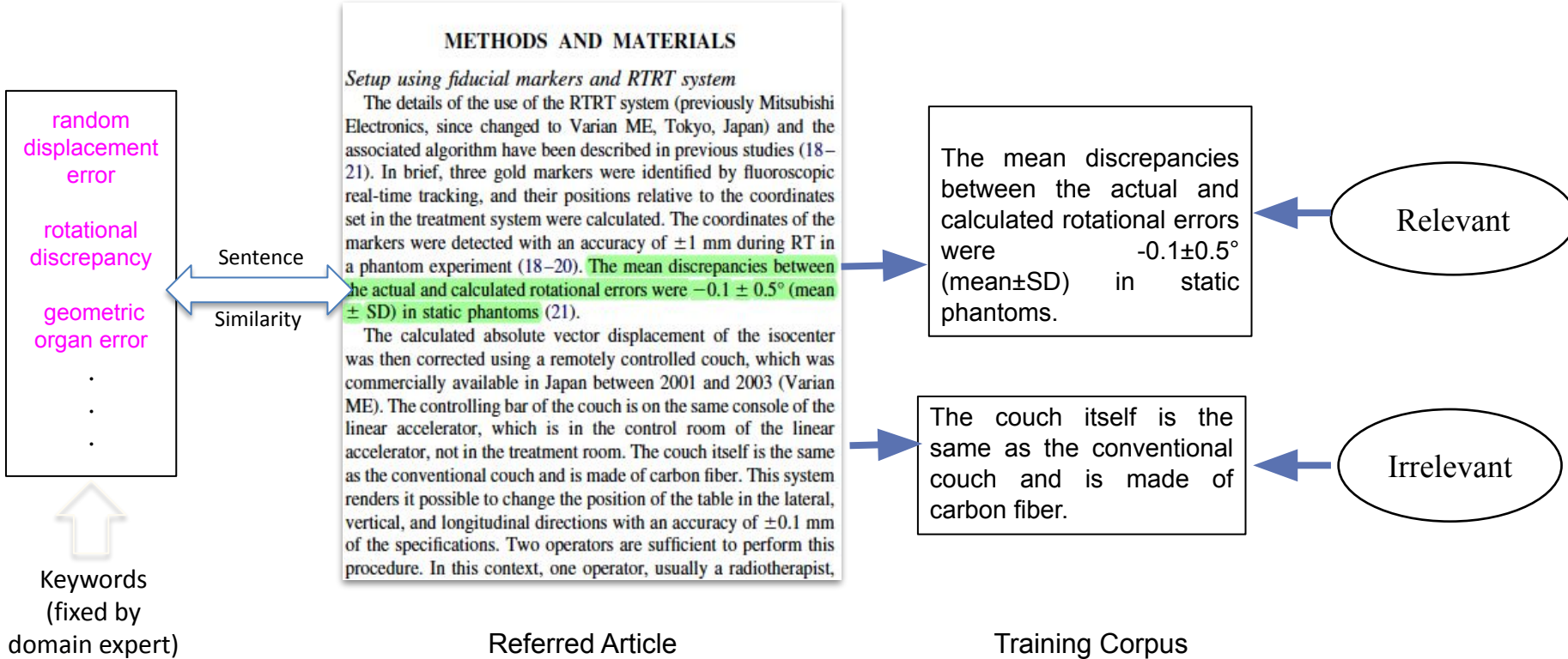
The screenshot shows the MedCAT interface. At the top, there is a navigation bar with a back arrow, a dropdown arrow, a search icon, and the text 'CAT Trainer medcat.rosalind.kcl.ac.uk'. On the right side of the navigation bar, there are share and menu icons. Below the navigation bar, the title 'MedCAT' is displayed on the left, and a hamburger menu icon is on the right. The main content area is split into two panels. The left panel contains a text-based relationship: 'Myocardial Infarction' is a type of 'heart disease'. The right panel is a table with the following data:

Pretty Name	Myocardial Infarction
Identifier	C0027051
Type	Disease or Syndrome
Confidence Score	0.99
Start Index	0
End Index	21
ICD-10 Code	{ "chapter": "I21", "name": "Acute myocardial infarction" }
id	0
Status	Other

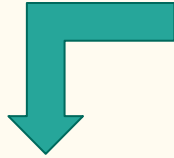
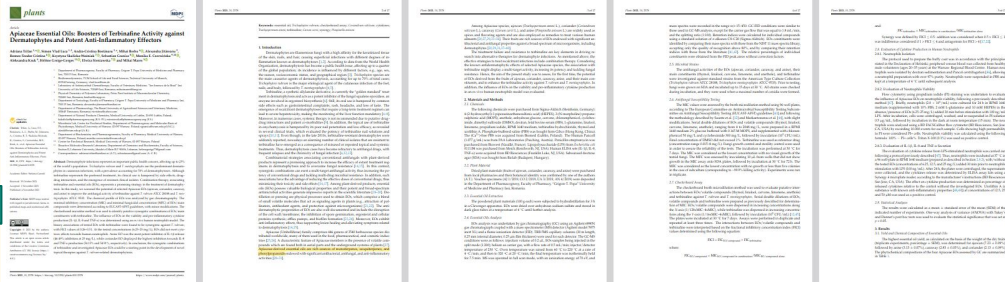
MedCAT Github: <https://github.com/CogStack/MedCAT>

MedCAT Interface: <https://medcat.rosalind.kcl.ac.uk/>

# NLP to Identify Geometric Errors from Literature



# NLP to Identify Nutraceuticals in Literature



Research Article

## Antibacterial Activity and Phytochemical Screening of Essential Oil of *Foeniculum vulgare*

Azadeh Foroughi<sup>1</sup>, Pouya Pournaghi<sup>2</sup>, Mohsen Zhaleh<sup>3</sup>, Akram Zangeneh<sup>1,3</sup>, Mohammad Mahdi Zangeneh<sup>1,3\*</sup>, Rohallah Moradi<sup>4</sup>

<sup>1</sup>Microbiology section, Pathobiology & Basic sciences department, Veterinary faculty, Bazi University, Kermanshah, Iran.  
<sup>2</sup>Department of Biology, Kermanshah Center of Payame Noor University, Kermanshah, Iran.  
<sup>3</sup>Department of Laboratory Science, School of Paramedicine, Kermanshah University of Medical Science, Kermanshah, Iran.  
<sup>4</sup>Department of Chemistry, Kermanshah Center of Payame Noor University, Kermanshah, Iran.

Available Online: 15<sup>th</sup> November, 2016

**ABSTRACT**  
Certainly, using herbal plants is the oldest way of mankind to treat the diseases. Considering the drug resistance and the side effects of chemical antibacterial drugs, the research approach is increasingly going toward using natural resources. The aim of the study was to evaluate the chemical composition and antibacterial activity of essential oil of *Foeniculum vulgare* against *Escherichia coli* O157:H7 and *Staphylococcus aureus*. Gas chromatography mass spectrometry was done to specify chemical composition. As a screen test to detect antibacterial properties of the essential oil, agar disk and agar well diffusion methods were employed. Macrobroth tube test was performed to determine MIC. The results indicated that the most substance found in *F. vulgare* essential oil was Trans-anethole (47.41 %), also the essential oil of *F. vulgare* with 0.007 g/ml concentration has prevented *E. coli* and with 0.003 g/ml concentration has prevented *S. aureus*, from the growth. Thus, the research represents the antibacterial effects of the medical herb on *E. coli* and *S. aureus*. We believe that the article provide support to the antibacterial properties of the essential oil. The results indicate the fact that the essential oil of *F. vulgare* can be useful as medicinal or preservatives composition. Fractionation and characterization of active molecules will be the future work to investigate.

**Keywords:** *Foeniculum vulgare*, Essential oil, Chemical composition, Antibacterial activity.

## An Article

Paper Source: Foroughi, Azadeh et al. "Antibacterial activity and phytochemical screening of essential oil of *Foeniculum vulgare*." International Journal of Pharmaceutical and Clinical Research 8, no. 11 (2016): 1505-1509. accessed on July 25, 2024



The results indicated that the most substance found in *F. vulgare* essential oil was Trans-anethole (47.41 %), also the essential oil of *F. vulgare* with 0.007 g/ml concentration has prevented *E. coli* and with 0.003 g/ml concentration has prevented *S. aureus*, from the growth. Thus, the research represents the antibacterial effects of the medical herb on *E. coli* and *S. aureus*.

## Desired Information



## Foeniculum Vulgare

# Overview of the Proposed Framework

**Nutraceuticals** (Nutrition+Pharmaceuticals) that are present in **cinnamon, jeera, garlic, ginger** etc. are useful for therapeutic treatment and drug discovery

Keywords (e.g., essential oil, thymine, tannin etc.)

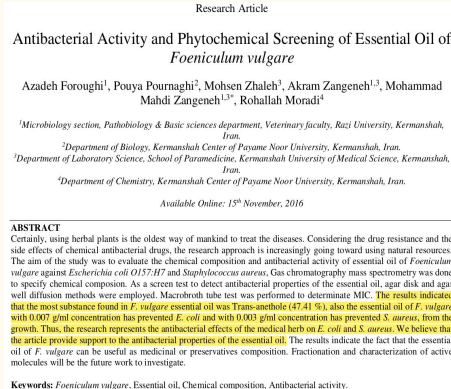


NLP Framework

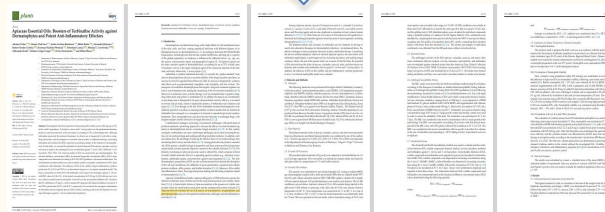


The results indicated that the most substance found in *F. vulgare* essential oil was **Trans-anethole (47.41 %)**, also the essential oil of *F. vulgare* with **0.007 g/ml concentration** has prevented *E. coli* and with **0.003 g/ml concentration** has prevented *S. aureus*, from the growth. Thus, the research represents the antibacterial effects of the medical herb on *E. coli* and *S. aureus*.

Desired Information

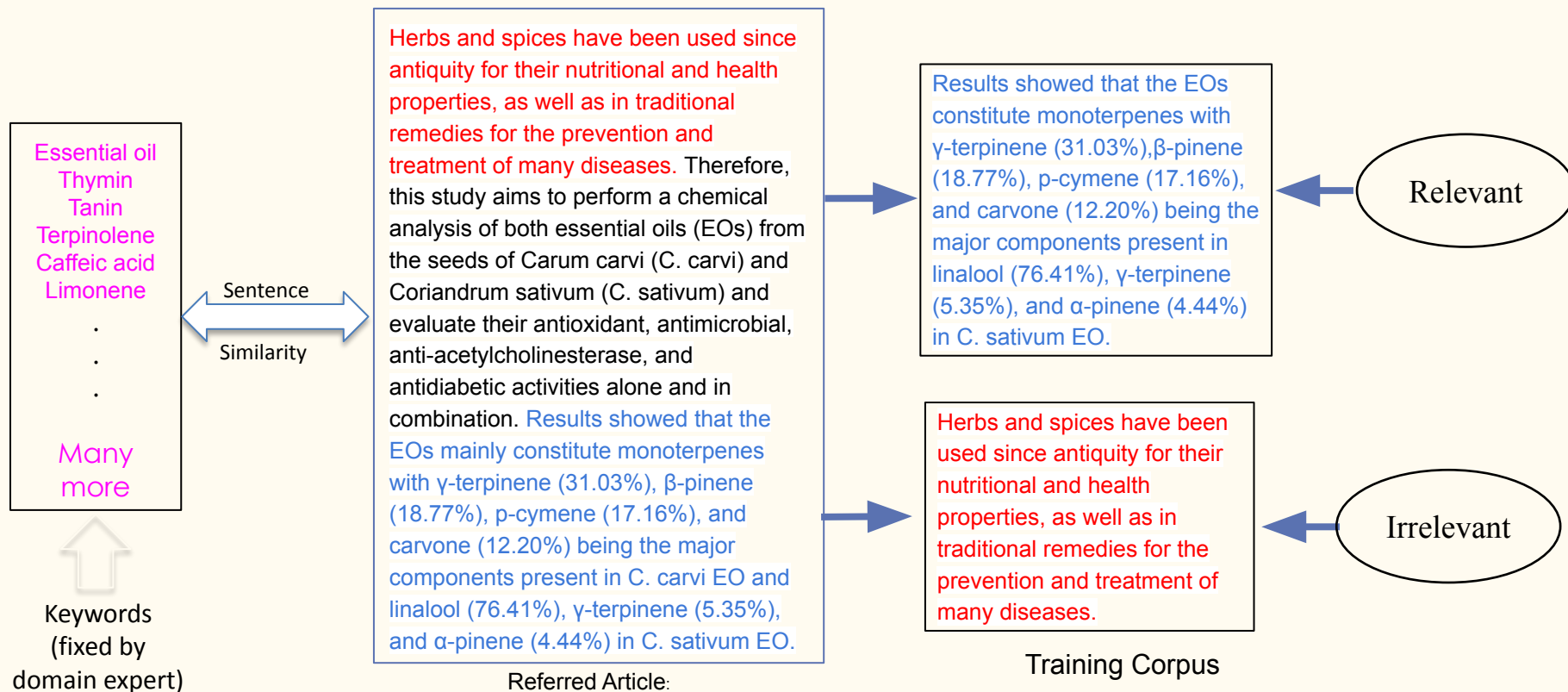


An Article



Training Corpus (41 Articles)

# Create Training Corpus and Ground Truths

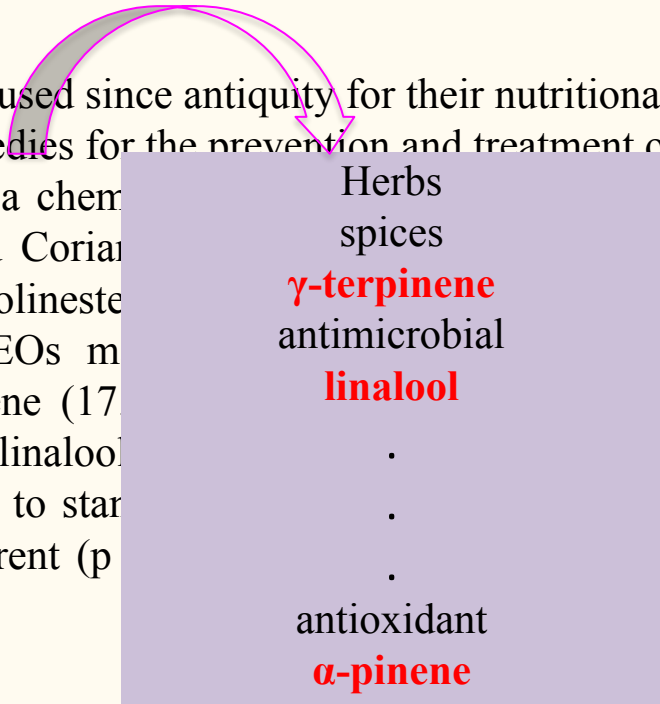


Paper Source:

Hajlaoui, Hafedh, et al. "Antimicrobial, antioxidant, anti-acetylcholinesterase, antidiabetic, and pharmacokinetic properties of *Carum carvi* L. and *Coriandrum sativum* L. essential oils alone and in combination." *Molecules* 26.12 (2021): 3625.

# Create Training Corpus and Ground Truths

Herbs and spices have been used since antiquity for their nutritional and health properties, as well as in traditional remedies for the prevention and treatment of many diseases. Therefore, this study aims to perform a chemical analysis of essential oils (EOs) from the seeds of Carum carvi (C. carvi) and Coriandrum sativum (C. sativum) and evaluate their antioxidant, antimicrobial, and anti-acetylcholinesterase activities alone and in combination. Results showed that the EOs mainly contain  $\beta$ -pinene (18.77%), p-cymene (17.11%), and linalool (17.11%) in C. carvi EO and linalool (31.03%),  $\alpha$ -pinene (4.44%), and p-cymene (17.11%) in C. sativum EO. In comparison to standard compounds, C. carvi EO showed high and significantly different (p < 0.05) antioxidant activity than the mixture.



Herbs  
spices  
 **$\gamma$ -terpinene**  
antimicrobial  
**linalool**  
.  
.  
.  
antioxidant  
 **$\alpha$ -pinene**

oils (EOs) from the seeds of  
and evaluate their antioxidant,  
s alone and in combination.  
with  $\gamma$ -terpinene (31.03%),  
being the major components  
, and  $\alpha$ -pinene (4.44%) in C.  
aled that C. carvi EO showed  
an C. sativum EO, but lower

# UMLS Concepts

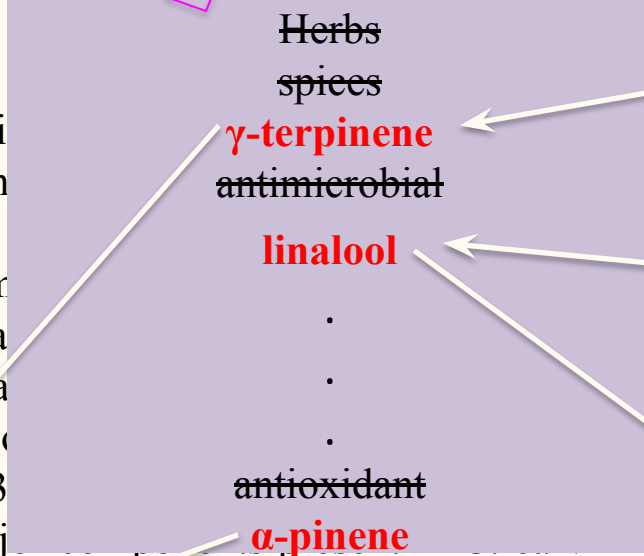
- **UMLS** (Unified Medical Language Systems) is a comprehensive list of biomedical terms for developing automated systems capable of understanding the specialized vocabulary used in biomedicine and health care.
- **MetaMap** is a tool used to identify the semantic type of a biomedical concept and assigns a concept ID to a particular term belonging to a particular semantic type.

Semantic Category	Semantic Type	Example Terms	UMLS Concept ID
Chemicals & Drugs	Organic Chemical	$\alpha$ -pinene	C0301253
	Chemical Viewed Functionally	Flavouring	C0682897
	Pharmacologic Substance	Linalool	C0064997
Living Beings	Plants	Coriandrum sativum	C0946611
Disorders	Disease or Syndrome	Thrombocytopenia	C0040028



# UMLS Features

Herbs and spices as well as in Therefore, this study aim of Carum ca antioxidant, a in combinatio C0439084 (3 being the maj (5.35%), and C0301253 (4.44%) in C. sativum EO. In comparison to standards, statistical analysis revealed that C. carvi EO showed high and significantly different ( $p < 0.05$ ) antioxidant activity than C. sativum EO, but lower than the mixture.



their tion both essential in the seeds vum (*C. sativum*) and evaluate their e, and antidiabetic activities alone and mainly constitute monoterpenes with ene (17.16%), and carvone (12.20%) and C0064997 (76.41%), gamma-terpinene (5.35%), and C0301253 (4.44%) in *C. sativum* EO. In comparison to standards, statistical analysis revealed that *C. carvi* EO showed high and significantly different ( $p < 0.05$ ) antioxidant activity than *C. sativum* EO, but lower than the mixture.

# Experimental Settings

	Number of sentences	
	Training corpus (1843)	Test corpus (6317)
Nutraceuticals	259	56
Non-nutraceuticals	1584	6260

- **Synthetic Minority Over-sampling Technique (SMOTE):** oversampling approach used to address class imbalance
- The performance of the AdaBoost model is also compared with **BioBERT, RoBERTa and GPT2**.

# Results

Feature Types	Classifier	Precision	Recall	F1-Score
Given Keywords	Keyword Matching	0.167	0.044	0.07
BioBERT	BioBERT	0.242	<b>0.981</b>	0.388
BioBERT + UMLS	BioBERT	0.243	<b>0.981</b>	0.389
RoBERTa	RoBERTa	0.242	<b>0.981</b>	0.388
RoBERTa + UMLS	RoBERTa	0.243	<b>0.981</b>	0.389
GPT2	GPT2	0.242	<b>0.981</b>	0.388
GPT2 + UMLS	GPT2	0.243	<b>0.981</b>	0.389
BoW + TF-IDF	SVM	0.242	<b>0.981</b>	0.388
	LR	0.242	<b>0.981</b>	0.388
	RF	0.276	0.971	0.43
	AdaBoost	0.264	0.971	0.415
BoW + Entropy	SVM	0.274	<b>0.981</b>	0.428
	LR	0.277	0.971	0.431
	RF	0.279	0.961	0.432
	AdaBoost	0.281	0.961	0.435
BoW + UMLS + TF-IDF	AdaBoost	0.278	0.971	0.432
BoW + UMLS+ Entropy	AdaBoost	<b>0.291</b>	<b>0.981</b>	<b>0.447</b>

# Results

<b>Test Articles</b>	<b>True Positive</b>	<b>False Positive</b>	<b>False Negative</b>	<b>Precision</b>	<b>Recall</b>
<b>Article 1</b>	7	6	0	0.538	1
<b>Article 2</b>	9	4	2	0.692	0.818
<b>Article 3</b>	8	45	0	0.151	1
<b>Article 4</b>	10	25	0	0.286	1
<b>Article 5</b>	5	52	1	0.088	0.833
<b>Article 6</b>	6	14	0	0.300	1
<b>Article 7</b>	3	14	0	0.176	1
<b>Article 8</b>	4	11	0	0.267	1
<b>Article 9</b>	1	7	0	0.125	1

# Analysis of Results

## *False -ve*

*The main compound **estragole** was ranging from 34 to 89%.*

## *False +ve*

*In the tarragon water-ethanolic extract, the amount of analysed compounds was also high and equal to 60.81 mg CQA·g<sup>-1</sup>, while **caffeic acid** derivatives **were not found** in the water extract of tarragon.*

## *False +ve corrected after using UMLS concepts*


***Triton X-100 (0.1%)** caused a very strong reduction in the cell viability, with only 1.63% viable cells*

Using **UMLS features** it was correctly classified



***C0146951** caused a very strong reduction in the cell viability, with only 1.63% viable cells*


# Large Language Models

See the latest updates to the [Gemini Apps Privacy Hub](#)

 Trans-anethole (47.41 %), also the essential oil of *F. vulgare* with 0.007 g/ml concentration has prevented *E. coli*

Show drafts  

 You are right, there is research suggesting that trans-anethole, a component of fennel (*F. vulgare*) essential oil, can inhibit *E. coli* growth [1, 4]. Studies have shown that trans-anethole exhibits antimicrobial activity against various bacteria, including *E. coli* [4]. The exact mechanism for this inhibition is not fully understood, but it may involve damaging the bacterial cell membrane or interfering with essential cellular processes [5]. 

However, it is important to note that the concentration of trans-anethole in fennel essential oil (0.007 g/ml) used in the research you mentioned might be lower than the concentration needed for complete eradication of *E. coli*.  Additionally, the effectiveness of essential oils can vary depending on factors such as the specific strain of bacteria and the environmental conditions.

If you are considering using fennel essential oil to combat *E. coli*, it is important to consult with a healthcare professional. They can advise you on the appropriate use of essential oils and potential risks or interactions with other medications.

Check completed • [Understand the results](#)



 You

Trans-anethole (47.41 %), also the essential oil of *F. vulgare* with 0.007 g/ml concentration has prevented *E. coli*

 ChatGPT

Trans-anethole is a compound found in essential oils, including those of *Foeniculum vulgare*, commonly known as fennel. This compound has been reported to exhibit antibacterial properties. Your statement suggests that trans-anethole, present in the essential oil of *F. vulgare* at a concentration of 0.007 g/ml, has shown efficacy in preventing the growth or activity of *Escherichia coli* (*E. coli*), a common bacterium.

It's important to note that while such findings are promising, they should be interpreted with caution. The effectiveness of trans-anethole or any other compound against bacterial growth can depend on various factors, including the specific strain of bacteria, the concentration and purity of the compound, and the conditions of the experiment.

Further research, including controlled studies and clinical trials, would be necessary to confirm and better understand the potential antibacterial properties of trans-anethole and its applicability in preventing bacterial infections caused by *E. coli*.



# Another Approach: Sequence Classification

- Assigning labels to each word of the sentence
- It is performed to identify the essential named entities: **nutraceuticals** and **concentrations** in each sentence.
- The sentences where both entities are present are relevant for the study.

# Creation of Sense Tagged Training Corpus

We have carried out in two ways:

➤ **Three-tags:**

- Nutraceutical keywords marked as ‘**Nutraceutical**’
- Concentration values marked as ‘**Concentration**’
- Other tokens marked as ‘**Non-nutraceutical**’

Trans-anethole	(47.41%)	found	in	F.vulgare	with	0.007 g/ml	concentration	has	prevented	E.coli
[Nutra]	[Conc]	[Non]	[Non]	[Non]	[Non]	[Conc]	[Non]	[Non]	[Non]	[Non]



# Creation of Sense Tagged Training Corpus

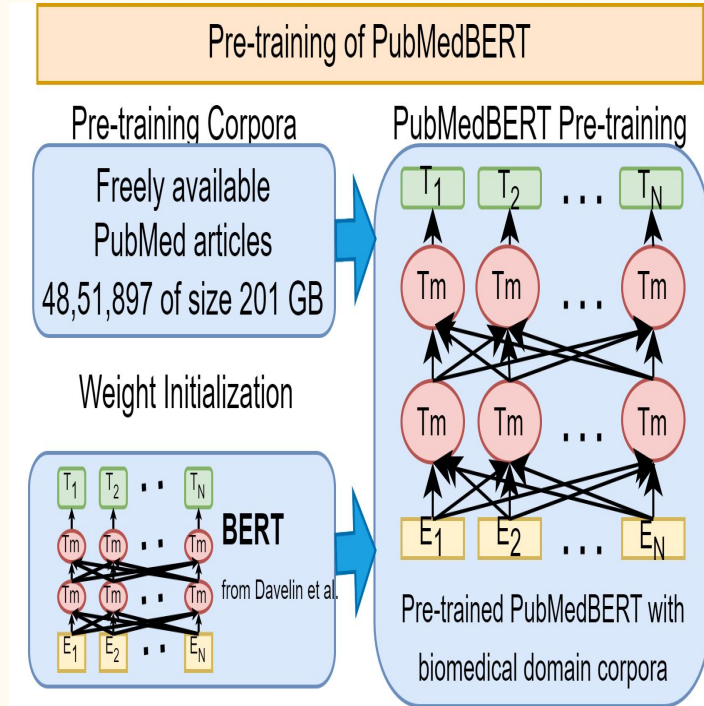
## ➤ Multi-tags:

- Nutraceutical keywords marked as ‘Nutraceutical’
- Concentration values marked as ‘Concentration’
- Other tokens marked as Part-of-Speech Tags

Trans-anethole	(47.41%)	found	in	F.vulgare	with	0.007 g/ml	concentration	has	prevented	E.coli
[Nutra]	[Conc]	[VBD]	[IN]	[NNP]	[IN]	[Conc]	[NN]	[VB]	[VBD]	[NNP]

# Overview of PubMedBERT

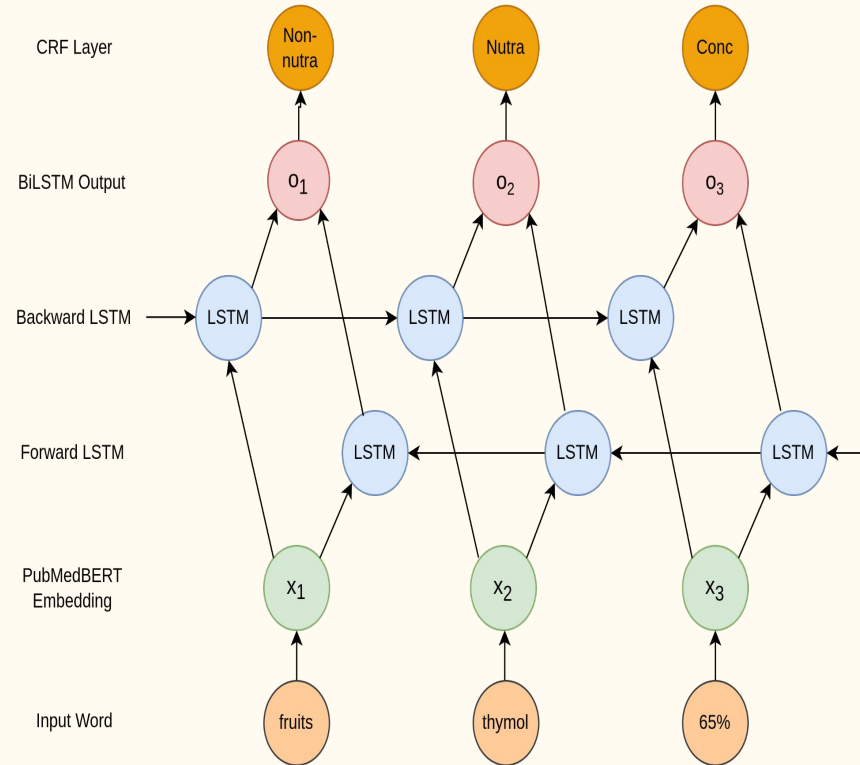
- ❖ Downloaded freely available PubMed articles from [ftp.ncbi.nlm.nih.gov/pub/pmc/oa\\_bulk](ftp.ncbi.nlm.nih.gov/pub/pmc/oa_bulk)
- ❖ Extracted **48,51,897** articles in text format, whose size is **201 GB**.
- ❖ These articles are peer reviewed publications and contain useful information.
- ❖ Objectives:
  - **Pre-Training** of BERT using only full text articles over PubMed
  - **Fine-Tune** it to address different biomedical NLP problems



# Overview of Bi-LSTM + CRF + PubMedBERT

- **First layer:** Map each word to a word vector based on pre-trained PubMedBERT embedding
- **Second layer:** the Bidirectional Long Short-Term Memory (BiLSTM) layer, to extract the sentence features
- **Third layer:** the Conditional Random Field (CRF) layer to perform sequence labelling

Pass on the each word into the model until all the data have been labelled



# Results (3-class)

Nutraceutical, Concentration, Non-nutraceutical

	CRF	Bi-LSTM	Bi-LSTM +CRF	BERT + Bi-LSTM + CRF	BioBERT + Bi-LSTM + CRF	PubMedBERT + Bi-LSTM + CRF
<b>Accuracy</b>	0.971	0.913	0.981	0.990	0.992	<b>0.994</b>
<b>Precision</b>	0.935	0.740	0.899	0.917	0.938	<b>0.940</b>
<b>Recall</b>	0.941	0.782	0.775	0.928	0.942	<b>0.942</b>
<b>F1-Score</b>	0.938	0.761	0.829	0.922	0.940	<b>0.941</b>

CRF: <https://pypi.org/project/sklearn-crfsuite/>

Bi-LSTM: <https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html>

Bi-LSTM+CRF: <https://github.com/goxdve/BiLSTM-CRF/tree/master>

Bi-LSTM+CRF+BERT: <https://github.com/hertz-pj/BERT-BiLSTM-CRF-NER-pytorch>

## Results (More than 3 Class)

	CRF	Bi-LSTM	Bi-LSTM + CRF	BERT + Bi-LSTM + CRF	BioBERT + Bi-LSTM + CRF	PubMedBERT + Bi-LSTM + CRF
<b>Accuracy</b>	<b>0.917</b>	0.841	0.843	0.886	0.888	0.890
<b>Precision</b>	0.883	0.838	0.859	0.877	0.884	<b>0.885</b>
<b>Recall</b>	0.821	0.762	0.720	0.802	0.823	<b>0.825</b>
<b>F1-Score</b>	0.841	0.788	0.763	0.838	0.852	<b>0.855</b>

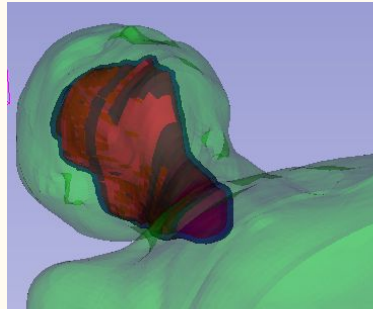
CRF: <https://pypi.org/project/sklearn-crfsuite/>

Bi-LSTM: <https://pytorch.org/docs/stable/generated/torch.nn.LSTM.html>

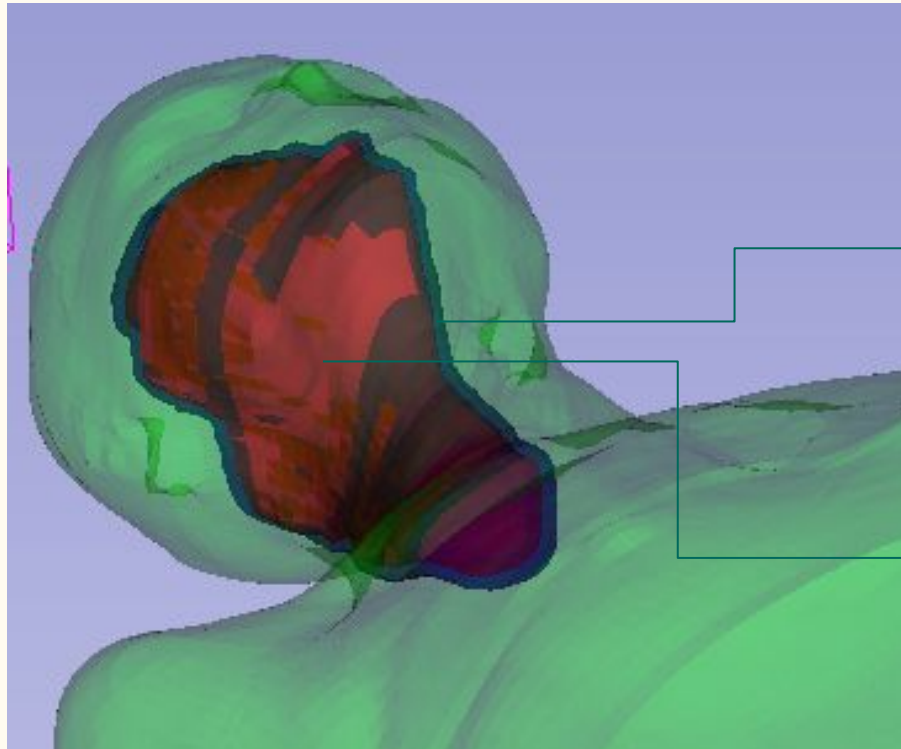
Bi-LSTM+CRF: <https://github.com/goxdve/BiLSTM-CRF/tree/master>

Bi-LSTM+CRF+BERT: <https://github.com/hertz-pj/BERT-BiLSTM-CRF-NER-pytorch>

# Deep Learning for Tumor Contouring



# AI for Automatic Tumor Contouring

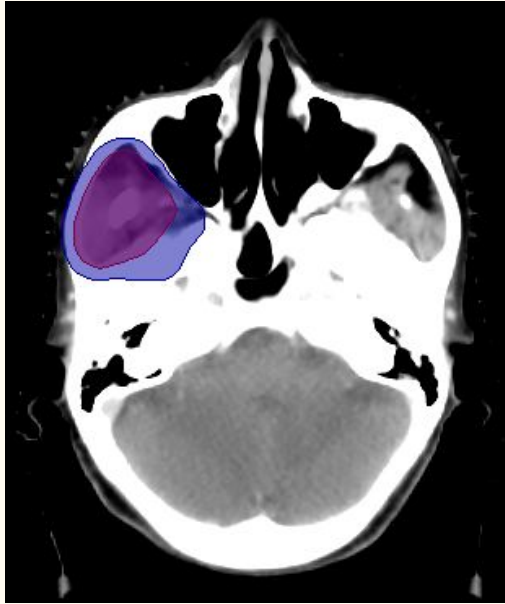


Blue region - Planned Tumor Volume (**PTV**)

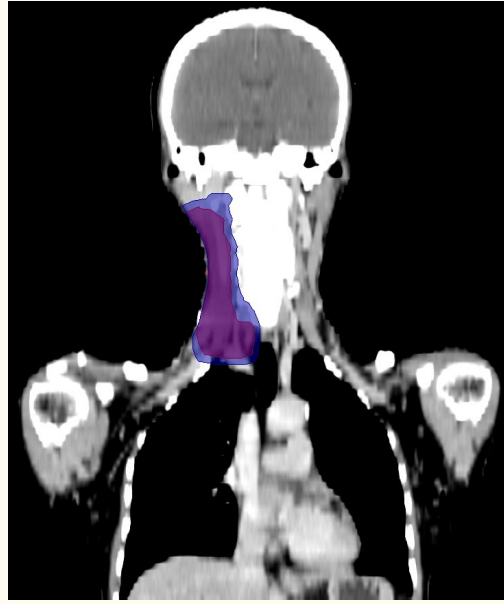
Red region - Gross Tumor Volume (**GTV**)

AI system is needed to precisely predict the PTV

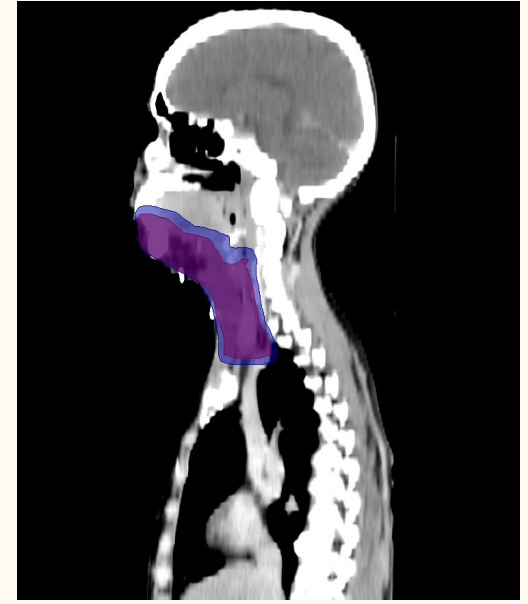
# Tumor Diagnosis: illustration



Top view



Front view

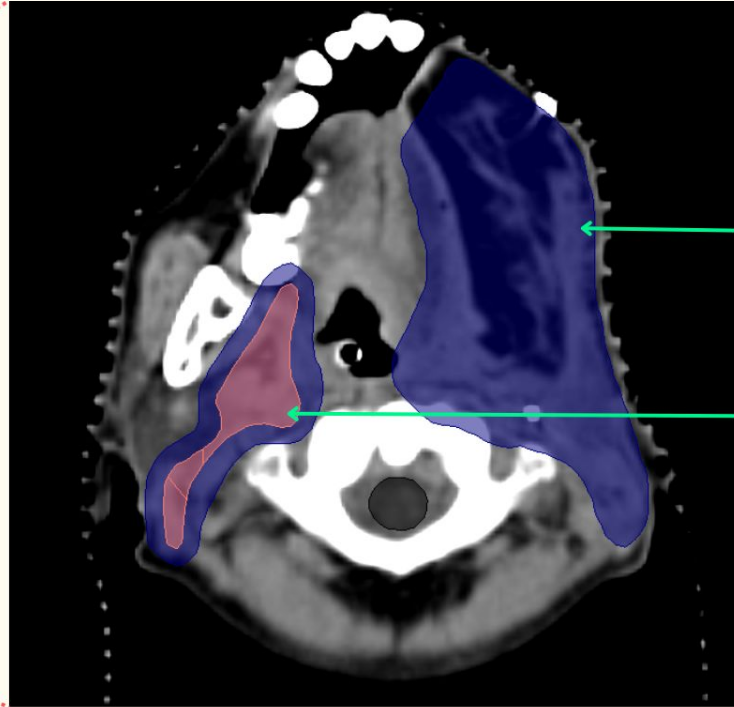
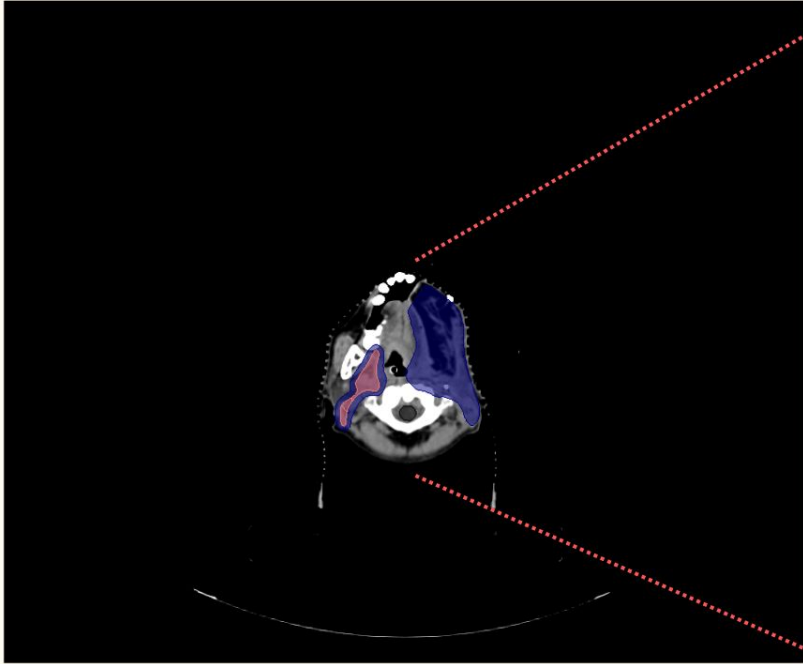


Side view

Red Region - GTV

Blue Region - PTV



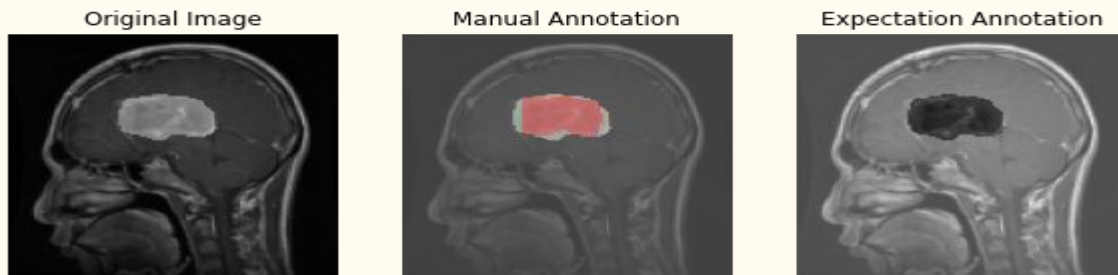


PTV\_54/30 -  
Dark Blue Color

CTV\_54/30 -  
Light red

# TCPNet: Framework for Brain Tumor Detection

- The radiology images e.g., CT Scan, MRI etc. are used to manually identify the size and characteristics of the tumor by the experts to deliver radiotherapy.
- Consequently many times it leads to error due to deviation between actual and identified regions of tumors.
- The image segmentation techniques can do the same for safe and accurate radiotherapy, if trained well.

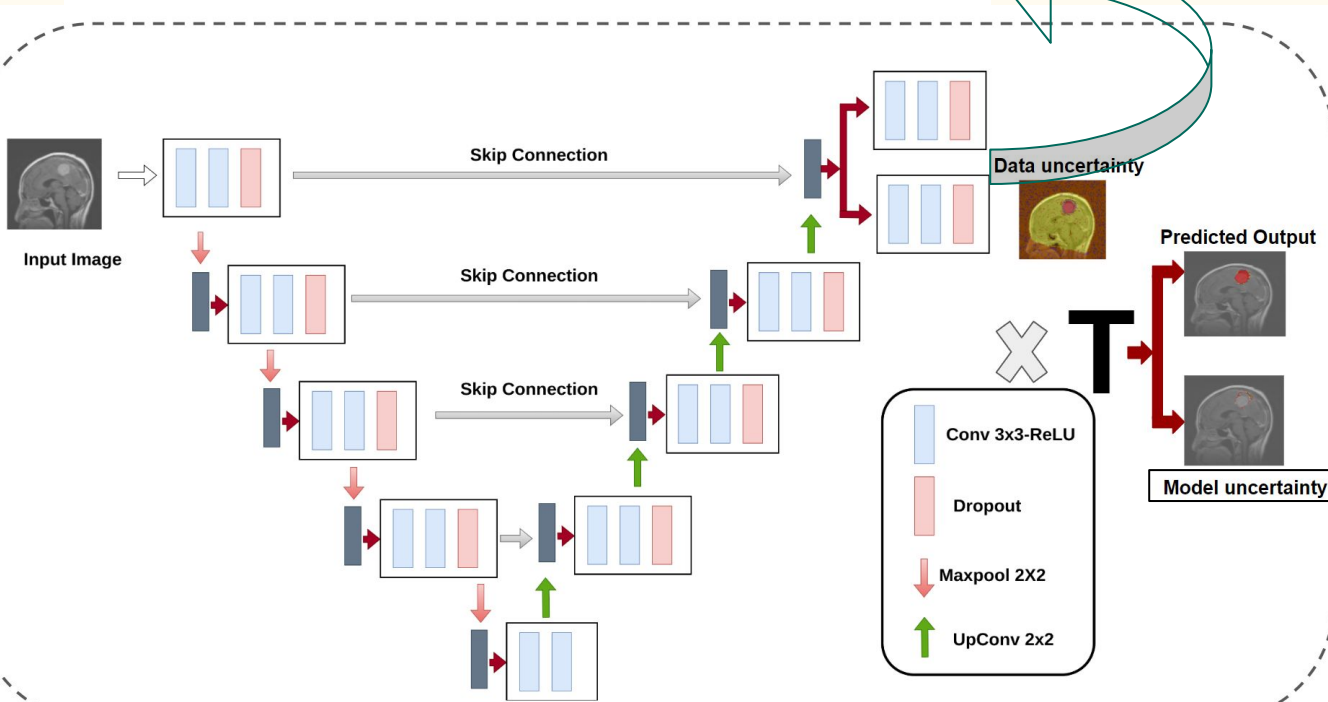


# TCPNet Architecture

$$\mathcal{L}_{NN} = \sum_{t=1}^T \left( \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{2\sigma_i^2} \|y_i - \hat{y}_i^t\|^2 + \frac{1}{2} \log(\sigma_i^2)^t \right) \right)$$

$$\mathcal{L}_{combined} = \frac{1}{T} (\mathcal{L}_{CE} + \mathcal{L}_{Dice} + \lambda * \mathcal{L}_{NN}).$$

**Final Loss**



$$\mathcal{L}_{CE} = - \sum_{t=1}^T \sum_{i=1}^N \left( y_i \log(\hat{y}_i^t) \right)$$

$$+ \mathcal{L}_{Dice} = \sum_{t=1}^T \frac{2 * \sum_{i=1}^N y_i \hat{y}_i^t}{\sum_{i=1}^N y_i^2 + \sum_{i=1}^N (\hat{y}_i^t)^2}$$

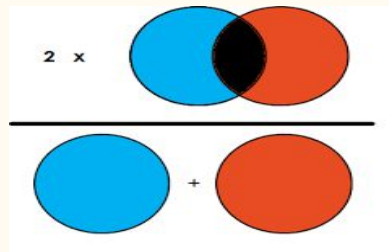
# TCPNet: DataSet

Dataset Name	Brain Tumor Dataset	LGG Segmentation Dataset
Types of Tumor	Meningioma (708 images) Glioma (1426 images) Pituitary (930 images)	Glioma (4165 images)
Total number of subjects	233 subjects	110 subjects
No. of subjects in training set	2335 images of 143 subjects	2322 images of 66 subjects
No. of subjects in validation set	488 images of 45 subjects	950 images of 22 subjects
No. of subjects in test set	241 images of 45 subjects	893 images of 22 patients

# TCPNet: Evaluation Criteria

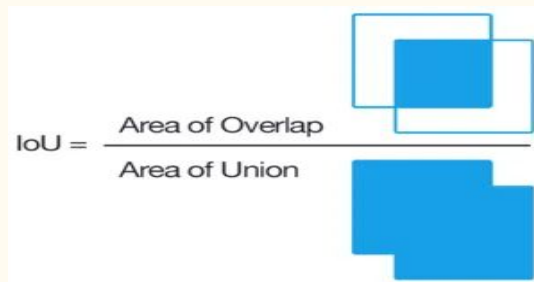
- **Dice Coefficient:** a spatial overlap index. It is same as F1 score.

$$\text{Dice} = (2 \times \text{TP}) / (2 * \text{TP} + \text{FP} + \text{FN})$$



- **Intersection Over Union (Jaccard Coefficient):** The area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth

$$\text{IoU} = (\text{TP}) / (\text{TP} + \text{FP} + \text{FN})$$



**TP: True Positive**    **FP: False Positive**

**FN: False Negative**    **TN: True Negative**

# TCPNet: Experimental Results

**T=20;  $\lambda = 0$** : Model uncertainty **T=20;  $\lambda = 0.01$** : Model + Data uncertainty

Dataset	Metrics	UNet	UNet++	LinkNet	TCPNet (Proposed)	
					<b>T=20 <math>\lambda = 0</math></b>	<b>T=20 <math>\lambda = 0.01</math></b>
Brain Tumor	IoU	0.767 $\pm$ 0.001	0.600 $\pm$ 0.018	0.259 $\pm$ 0.011	<b>0.783 <math>\pm</math> 0.001</b>	0.699 $\pm$ 0.006
	Dice	0.813 $\pm$ 0.005	0.750 $\pm$ 0.014	0.411 $\pm$ 0.021	<b>0.835 <math>\pm</math> 0.006</b>	0.756 $\pm$ 0.007
	Precision	<b>0.826 <math>\pm</math> 0.017</b>	0.699 $\pm$ 0.009	0.291 $\pm$ 0.025	<b>0.826 <math>\pm</math> 0.005</b>	0.743 $\pm$ 0.007
	Recall	0.840 $\pm$ 0.028	0.809 $\pm$ 0.018	0.698 $\pm$ 0.026	<b>0.862 <math>\pm</math> 0.009</b>	0.825 $\pm$ 0.006
	Specificity	0.840 $\pm$ 0.028	0.809 $\pm$ 0.018	0.698 $\pm$ 0.026	<b>0.862 <math>\pm</math> 0.009</b>	0.825 $\pm$ 0.006
LGG	IoU	0.624 $\pm$ 0.004	0.209 $\pm$ 0.015	0.554 $\pm$ 0.049	<b>0.652 <math>\pm</math> 0.015</b>	0.612 $\pm$ 0.006
	Dice	0.636 $\pm$ 0.005	0.309 $\pm$ 0.015	0.538 $\pm$ 0.023	<b>0.656 <math>\pm</math> 0.021</b>	0.639 $\pm$ 0.006
	Precision	0.646 $\pm$ 0.003	0.222 $\pm$ 0.023	0.576 $\pm$ 0.025	<b>0.667 <math>\pm</math> 0.015</b>	0.660 $\pm$ 0.002
	Recall	0.632 $\pm$ 0.005	0.226 $\pm$ 0.026	0.586 $\pm$ 0.033	<b>0.653 <math>\pm</math> 0.022</b>	0.634 $\pm$ 0.007
	Specificity	0.632 $\pm$ 0.005	0.226 $\pm$ 0.026	0.586 $\pm$ 0.033	<b>0.653 <math>\pm</math> 0.022</b>	0.634 $\pm$ 0.007

# TCPNet: Experimental Results @AIIMS Bhopal

Number of Subjects: 49

Resolution: 512 \* 512

Number of images in training set: 406

Number of images in test set: 114

Metrics	UNET	UNET++	Link Net	TCPNet
Precision	0.696	0.717	0.657	0.702
Recall	0.755	0.787	0.6166	0.848
Specificity	0.755	0.787	0.6166	0.848
IoU	0.648	0.661	0.577	0.669

---

# TCPNet: Experimental Results @JNCH Bhopal

Number of Subjects: 93

Resolution: 512 \* 512

Number of images in training set: 2620

Number of images in test set: 781

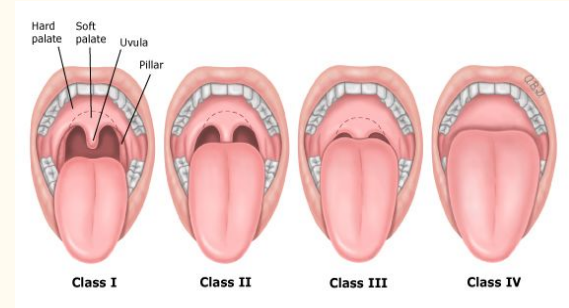
Metrics	UNET	UNET++	Link Net	TCPNet
Precision	0.866	0.833	0.799	0.870
Recall	0.868	0.847	0.869	0.888
Specificity	0.868	0.847	0.869	0.888
IoU	0.791	0.760	0.738	0.802

---



# Early Diagnosis of Sleep Disorder

- ❑ **Obstructive Sleep Apnea (OSA)** is a **sleep disorder** where a person's breathing is interrupted during sleep.
- ❑ The gold standard for diagnosing sleep apnea is full-night **Polysomnography (PSG)** test, carried out in a sleep lab, which is costly.
- ❑ Mallampati scale,  $SPO_2$ , diabetes, smoking habits etc. are salient indicators of OSA.
- ❑ Machine learning and NLP can indicate the severity of OSA by analysing the health profiles like BMI, Oxygen Level, **Mallampati Scale**



# Conclusions

- ❑ NLP and machine learning is useful for semantic interpretation of medical records.
- ❑ **Automatic annotation** of clinical notes is possible through NLP.
- ❑ **NLP** has potential to be **combined with** other techniques (e.g., **image processing**) to derive knowledge from data.
- ❑ **Image segmentation techniques** will be precious for **automatic tumor delineation**.
- ❑ **Multimodal information extraction** framework using radiology data combined with other clinical notes will be useful for early prognosis of many diseases.

# Acknowledgement



Thank You!