

# Zenith NLTK Methods Manifest To Textual Data Cleaning And Pre-Processing Destined For Obstetrics And Gynec Free-Text Radiology Reports

Ms.Vaishali M. Kumbhakarna<sup>1</sup>,Dr.Sonali B. Kulkarni(Asst.Prof)<sup>2</sup>,Ms.Apurva D.Dhawale<sup>3</sup>

Dr.Babasaheb Ambedkar Marathwada University, Department of Computer Science, Aurangabad

## ABSTRACT

*One of the customary obstacles in developing Machine Learning Algorithm in healthcare is a lack of annotated data. Even though an enormous, trendy research is going on in the field of language technology, its adaption into the medical domain is not fully sightseen .To discourse this research gap, we develop a machine learning models, which is based on NLP. This model accomplish textual data cleaning and clinical text pre-processing It will be further used for advanced search and annotation corpus of obstetrics and gynaecology radiology reports by means of well-structured dataset. It is further used for development of NLP based application employed for automation of clinical audit, for smooth functioning and regulation of PCPNDT Act.*

**Keywords:** NLP, PCPNDT, Obstetrics, Gynae, Porter Stemmer ,Lancaster Stemmer

## 1. INTRODUCTION

Natural language processing (NLP)[1] is one of the most emerging technologies that address complex and knowledge-dependent tasks such as automated search, machine translation, automated question answering ,and textual data mining[18] etc. Huge amount of electronic medical records[22] available across healthcare networks, as the increased use of different machines and modalities for diagnostic purpose.Such huge data availability inspiring researcher to develop the innovative clinical NLP[5],[15],[16] solutions. Thus NLP based application helps to optimize clinical care and improve patient aftermaths across the healthcare continuum.

In this paper, we put forward a text mining[20] system which will extract[25] the information from Obstetrics and Gynec radiology reports[2] which are unstructured free-text data[23] .

This system consists of three main modules : 1)Reading PDF text

2)Extracting medical findings

3)Cleaning and pre-processing text data.

Report reading module read the pdf file using python .The medical finding extraction[3] module automatically extracts only medical findings related text from pdf report and pre-processing[4] used to structure the clinical text data. The structuring of the free text reports[19] bridges the gap between users and report database .The first module serves as raw data for next component of system that is for cleaning and pre-processing module.The cleaning and pre-processing [28]module convert the raw data into clean and structured data which is used for further processing . Thus the output of one module is the input for second and output of second is the input for third module.Our proposed system will improves the performance of analysing the clinical text of obes/gynec radiology reports. The pre- processed data is future used for development of machine learning algorithm ,which is used in application for automation of clinical audit conducted as per rule of Parliament law of India ,for smooth functioning and regulation of PCPNDT Act

## 2. LITERATURE REVIEW

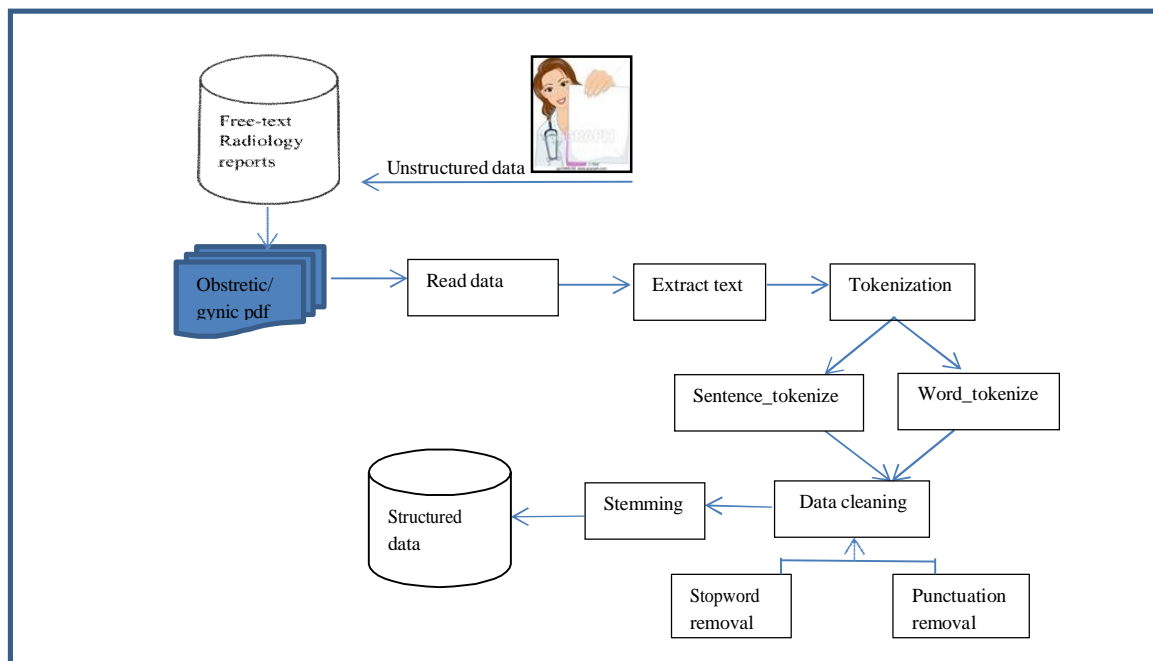
Radiology is the branch of medical discipline that uses imaging technology to diagnose and treat diseases[21] within the human body .In various bio-medical literature reviews ,it is found that Radiology is a specialty of medicine that employs various imaging modalities for diagnostic[17] of different body parts are given bellow.

- Ultrasound sonography
- Computed Tomography.(CT)
- Magnetic Resonance Imaging.(MRI)
- Positron Emission Tomography(PET) and Single Photon Emission Computed Tomography(PECT)
- Fluoroscopy.
- Mammography .
- X ray

The modern practice of radiology involves several different professionals working together as a team. The radiologist- interprets medical images, physicians- communicates the radiology report findings, The nurse involved in report writing against observations of patient and application engineer-the non-medico professional, install software on machine .The radiographer, also known as a "radiologic technologist is a healthcare professional that uses sophisticated technology to produce medical images.The impulsive growth in use of all maintained modalities, gives birth to large volumes of clinical free text documents .It results into the need of new technology-computer expert,which will work as a helping hand for the team of medical professionals’

## 3. METHOD

The proposed method for structuring the unstructured ,free text report of Obstetrics /Gynec radiology clinic generated from MINDRAY Z5 basic colour doppler ultrasound[7] or sonography machine. Accordingly to hide the hospital detail we have to de-identify the data .So we use replace() of python for security and privacy measure.



**Figure 1** Machine Learning Algorithm

### 3.1 DATABASE

The row unstructured data is collected from radiology clinic .It is PDF data ,which contain 727 sonography test reports of pregnancy. These PDF files has unstructured free text data regarding radiology diagnosis .Each file has several ultrasound images of fetus and one text report which is in pdf format .Data pre-processing was applied on pdf text report .It consist of Clinical Indications & findings information regarding patient ,detail of current gynec status of mother and 2D obstetrics measurements of fetus.

## 4. EXPERIMENTAL WORK

**4.1 READING DATA:** The radiology reports[6] are unstructured,free text data as it is generated from customised software incorporated into sonography machine by manufacturer .so the format of radiology report varies from machine to machine.The data we are using is in pdf format ,so first we have to read the pdf in python using PyPDF2 package.

```
import PyPDF2 as p2
```

```
pdfFileObj=open("D:\Anaconda\envs\sample.pdf","rb")
```

Once the pdf has been read,the file object is created which is in read binary mode,then next we can extract the text [24]from file object by using extractText() function in python.

```
pdfReader=p2.PdfFileReader(pdfFileObj) pdfobj=pdfReader.getPage(0) print(pdfobj.extractText())
```

```
op: XYZ HOSPITALName:AABDAR,ANJALI ID:20180328-134931-F572
```

```
Age:23Years Gender:Female
```

```
BEHIND BUS STAND, DP ROAD,BEED
```

```
Telephone: 02442-229972Fax:
```

```
Website:
```

```
OB2/3
```

```
Exam Date: 28/03/2018 Equipment Used:
```

```
MINDRAY Z5
```

```
Clinical Indications & DataLMP: 10/01/2018 GA: 11w0d EDD(LMP): 17/10/2018 AUA: 8w2d EDD(AUA): 05/11/2018
```

```
Gestations: 1 Gravida: 3 Para: 2
```

```
Obstetrics2D MeasurementsGSHansmann3.53cm8w2d~
```

```
CRLHadlock1.85cm8w3d~
```

```
8w0d-8w6dDoppler MeasurementsFHR: 175BpmComments:
```

```
Signature(seal):
```

```
Date Signed:
```

**4.2 TOKENIZATION :** Is the first step of data preprocessing It is used to safeguard sensitive data like, medical records[9].Tokenization is also used to divide a large text is into smaller parts called tokens. These tokens are very useful for finding patterns in text data.Understanding the pattern in the text helps for building applications based on Natural language processing ,such as Text classification, sentimental analysis, language translation, etc.Natural Language toolkit has very important module NLTK tokenize which further comprises of sub- modules: a)Word tokenize b)Sentence tokenize

```
import nltk nltk.download()
```

```
from nltk.tokenize import sent_tokenize, word_tokenize
```

```
egtext="XYZ HOSPITALName:AABDAR,ANJALI ID:20180328-134931-F572 Age:23Years
```

```
Gender:Female BEHIND BUS STAND, DP ROAD,BEED Telephone: 02442-229972 Fax: Website:"
```

**4.2.1 Tokenization of words:** Word tokenization is used for splitting a large sample of text into single words[8]. This is a vital tasks for medical data where each word is important and needs to be captured,which is subjected to further analysis

like classifying or counting for a particular sentiment etc. In python the Natural Language Tool kit(NLTK) provides very library to achieve this. We use the method `word_tokenize()` to split a sentence into words. The output of word tokenization can be converted to Data Frame for better text understanding in machine learning applications. It can also be provided as input for further text cleaning steps Machine learning models need numeric data to be trained and make a prediction. Word tokenization becomes a crucial part of converting the text data into numeric data .

```
print(word_tokenize(egtext)) for i in
word_tokenize(egtext):
print(i)
```

```
op: ['XYZ', 'HOSPITALName', ':', 'AABDAR', ',', 'ANJALI', 'ID:20180328-134931-F572', 'Age:23Ye ars', 'Gender', ':',
'Female', 'BEHIND', 'BUS', 'STAND', ',', 'DP', 'ROAD', ',', 'BEED', 'Telephone', ':', '0 2442-229972', 'Fax', ':', 'Website',
':']
```

**4.2.2 Tokenization of Sentences:** `sent_tokenize()` is needed to count average words per sentence, to calculate the ratio of word per sentence etc . As output of such tasks are numeric values, it is served as an important feature for training dataset in machine learning applications.

```
print(sent_tokenize(egtext)) for i in
sent_tokenize(egtext):
print(i)
```

```
op: ['XYZ HOSPITALName:AABDAR,ANJALI ID:20180328-134931-F572 Age:23Years Gender:F
emale BEHIND BUS STAND, DP ROAD,BEED Telephone: 02442-229972 Fax: Website:']
```

- Tokenization in NLP is the process by which a large quantity of text is divided into smaller parts called tokens.
- Natural language processing is used for building applications such as Text classification, intelligent chatbot, sentimental analysis, language translation, etc.
- Natural Language toolkit has very important module NLTK tokenize sentence which further comprises of sub-modules
- We use the method `word_tokenize()` to split a sentence into words. The output of word tokenizer in NLTK can be converted to Data Frame for better text understanding in machine learning applications.
- The `sent_tokenize()`, Sentence tokenizer in Python NLTK is an important feature for machine training.

**4.3 REMOVAL OF STOPWORDS:** Before coming towards stopword removal ,we will discuss Text pre- processing .stopword removal is the part of text pre-processing .It is the process of preparing text data for machine learning application. There are many different steps in text pre-processing .

**4.3.1 Stopwords-** These are actually the most common words in any language (like articles, prepositions, pronouns, conjunctions, etc) and does not add much information to the text. In this paper stopwords are the common English words which does not add much meaning to a text data. They can safely be removed without disturbing the meaning of the sentence. the removal of such words does not show any negative consequences on the development of model. It will definitely reduces the dataset size and thus reduces the training time due to the fewer number of tokens.

```
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
egtxt = "XYZ HOSPITAL Name is :AABDAR,ANJALI ID:20180328-134931-F572 Age:23Years
Gender:Female BEHIND BUS STAND, DP ROAD,BEED Telephone: 02442-229972 Fax: Website: OB2/3 Exam Date:
28/03/2018 Equipment Used: MINDRAY Z5 Clinical Indications & DataLMP: 10/01/2018 GA: 11w0d
EDD(LMP): 17/10/2018 AUA: 8w2d EDD(AUA): 05/11/2018
Gestations: 1 Gravida: 3 Para: 2 Obstetrics2D MeasurementsGSHansmann3.53cm8w2d` CRLHadlock1.85cm8w3d`
8w0d-8w6dDoppler MeasurementsFHR: 175Bpm Comments: Signature(seal): Date Signed: "
stop_words = set(stopwords.words("english")) words =
word_tokenize(egtxt)
filter_sent = [w for w in words if not w in stop_words]
```

```
print(filter_sent)
```

```
op: ['XYZ', 'HOSPITAL', 'Name', ':', 'AABDAR', ',', 'ANJALI', 'ID:20180328-134931-F572', 'Age:23  
Years', 'Gender', ':', 'Female', 'BEHIND', 'BUS', 'STAND', ',', 'DP', 'ROAD', ',', 'BEED', 'Telephone', ':',  
'02442-229972', 'Fax', ':', 'Website', ':', 'OB2/3', 'Exam', 'Date', ':', '28/03/2018', 'Equipment', 'Used', ':',  
MINDRAY', 'Z5', 'Clinical', 'Indications', '&', 'DataLMP', ':', '10/01/2018', 'GA', ':', '11w0d', 'EDD', '(',  
'LMP', ')', ':', '17/10/2018', 'AUA', ':', '8w2d', 'EDD', '(', 'AUA', ')', ':', '05/11/2018', 'Gestations', ':', '1',  
Gravida', ':', '3', 'Para', ':', '2', 'Obstetrics2D', 'MeasurementsGSHansmann3.53cm8w2d', 'CRLHadlock 1.85cm8w3d',  
'8w0d-8w6dDoppler', 'MeasurementsFHR', ':', '175Bpm', 'Comments', ':', 'Signature', '(',  
'seal', ')', ':', 'Date', 'Signed', ':']
```

**4.4 REMOVAL OF PUNCTUATION :** We cannot go directly from raw text to fitting a machine learning or deep learning model. We must clean [27] the text first, and prepare the text ready for modeling with machine learning application. We can use following steps to remove all punctuation marks with NLTK in Python.

```
# define punctuation  
punctuations = ""!()-[]{};:'",<>./?@$%^&*~_'"  
my_str = ['XYZ', 'HOSPITAL', 'Name', ':', 'AABDAR', ',', 'ANJALI', 'ID:20180328-134931-F572', 'A  
ge:23Years', 'Gender', ':', 'Female', 'BEHIND', 'BUS', 'STAND', ',', 'DP', 'ROAD', ',', 'BEED', 'Telepho  
ne', ':', '02442-229972', 'Fax', ':', 'Website', ':', 'OB2/3', 'Exam', 'Date', ':', '28/03/2018', 'Equipment', 'Us  
ed', ':', 'MINDRAY', 'Z5', 'Clinical', 'Indications', '&', 'DataLMP', ':', '10/01/2018', 'GA', ':', '11w0d', 'E  
DD', '(', 'LMP', ')', ':', '17/10/2018', 'AUA', ':', '8w2d', 'EDD', '(', 'AUA', ')', ':', '05/11/2018', 'Gestations',  
'1', 'Gravida', ':', '3', 'Para', ':', '2', 'Obstetrics2D', 'MeasurementsGSHansmann3.53cm8w2d', 'CRLH  
adlock1.85cm8w3d', '8w0d-8w6dDoppler', 'MeasurementsFHR', ':', '175Bpm', 'Comments', ':', 'Signat ure', '(', 'seal', ')',  
'Date', 'Signed', ':']  
# remove punctuation from the string no_punct = ""  
for char in my_str:  
if char not in punctuations: no_punct =  
no_punct + char  
  
# display the unpunctuated string  
print(no_punct)
```

```
op: XYZHOSPITALNameAABDARANJALIID:20180328-134931-F572Age:23YearsGenderFemale  
BEHINDBUSSTANDDDPROADBEEDTelephone02442-229972FaxWebsiteOB2/3ExamDate28/03/2018  
EquipmentUsedMINDRAYZ5ClinicalIndicationsDataLMP10/01/2018GA11w0dEDDLMP17/10/ 2018  
AUA8w2dEDDAUA05/11/2018Gestations1Gravida3Para2Obstetrics2DMeasurementsGSHans  
mann3.53cm8w2d"CRLHadlock1.85cm8w3d"8w0d-8w6dDopplerMeasurementsFHR175BpmComm  
entsSignaturesealDateSigned
```

**4.5 STEMMING AND LEMMATIZATION:** Both are the text normalizing and text mining [13][14][18] procedures in the field of Natural Language Processing that are applied to adjust text, words, documents for more processing.

**4.5.1 Stemming :** Stemming is a technique used to extract the original form of the words. It is faster as it chops [10] words into root word without knowing the context of the word [14] in given sentences.

**4.5.2 Lemmatization:** Lemmatization is a similar technique like stemming but it gives valid root word which is dictionary-based and that means the same thing. It is slower as compared to stemming but it gives proper context of the word [12] before proceeding.

In this study, we are going to explore the concept of stemming and how it helps doctors and radiologists as well as clinical auditors to deal with unstructured, free-text information [29] of Obs/Gynec radiology reports [26][30] and support decisions making. In this paper we are exploring some of the most common stemmers available in the NLTK library, for example PorterStemmer, SnowballStemmer and LancasterStemmer. Within these three stemmers for this study we are using porter and Lancaster stemmer only because the database we are using is an Indian and the snowballstemmer is exactly similar to the porter stemmer. It is an updated version of Porter's Stemmer with new rules that were already

existing in Porter’s Stemmer. relevant difference between Snowball and Porter is that Snowball applies details rules for other languages like Portuguese but as we are using database of Indian English language so we have excluded snowball stemmer as logic and process is exactly the same as Porter’s Stemmer.

```
from nltk.stem import PorterStemmer from nltk.stem
import LancasterStemmer ps=PorterStemmer()
ls=LancasterStemmer()
word_list = ['XYZ', 'HOSPITAL', 'Name', 'Gender', 'Female', 'Age:23Years', 'BEHIND', 'BUS',
'STAND', 'DP', 'ROAD', 'BEED', 'Telephone', 'Website',]
print("{0:20}{1:20}{2:20}".format("Word","Porter Stemmer","Lancaster Stemmer")) for word in
word_list:
print("{0:20}{1:20}{2:20}".format(word,ps.stem(word),ls.stem(word)))
```

**Table 1: Output**

Word	Porter Stemmer	Lancaster Stemmer
XYZ	Xyz	xyz
HOSPITAL	Hospit	hospit
Name	Name	Nam
Gender	Gender	Gend
Female	Femal	Fem
Age:23Years	age:23year	age:23years
BEHIND	Behind	Behind
BUS	Bu	Bus
STAND	Stand	Stand
DP	DP	Dp
ROAD	Road	Road
BEED	Beed	Bee
Telephone	Telephon	Telephon
Website	Websit	Websit

Evaluating Stemmers’ impact ,our original pdf report contained 519 characters, and after removal of stopwords we get 87 characters obtained this by using word tokenize().

```
words=word_tokenize(egtxt)
print(len(words))
```

### 5. RESULT OF STEMMED VERSION

Our stemmed version contains 78 characters, excluding spaces. We can obtain this using:

```
p=(word)
sw = sw+len(p)
```

In this paper we compared two stemmers:

**5.1 PORTER’S ALGORITHM-** stems common word endings that are applied according to different combinations of preceding consonants and vowels.PorterStemmer contains 71 characters .We can obtain this using:

```
p=(ps.stem(word))
sp = sp+len(p)
```

**5.2 LANCASTER’S ALGORITHM-** stems try to reduce the word to the shortest stem possible.LancasterStemmer contains 67 characters .We can obtain this using:

```
p=(ls.stem(word))
sl = sl+len(p)
```

The percentage of information retained is the ratio around 91.02%. If we apply the same logic for the same text with Lancaster, we have the following values: 85.89%.

$$\text{Ratio of information retained} = \frac{\text{length of stemmed character}}{\text{Total length of character}} * 100$$

## 6. CONCLUSION

This study explains the stemmer that retains less information is the Lancaster Stemmer. Information retained by Porter is around 91.02%, where as the information retained by Lancaster is around 85.89%. We conclude that even if Porter stemmer chop the stem of word less, it retain more information vice versa of, Lancaster reduce the word length and chop the stem more near to the root word, but retain the information less. In this paper after comparing both stemmers, we had proven that, Lancaster tends to over stem some words and this is the main reason why we say that Lancaster Stemming is one of the most aggressive stemmers. Porter's Stemmer is actually one of the oldest stemmer applications applied in computer science. The Porter stemmer pre-process the free text radiology report with more accuracy and rarer loss of information and we get structured data from unstructured text without loss of data.

## References

- [1] Sohrab Towfighi, "Labelling chest x-ray reports using an open-source NLP and ML tool for text data binary classification", Technical Notes, medRxiv, 2019.
- [2] S. Hassanpour, C.P. Langlotz, Information extraction from multi-institutional radiology reports, *Artif. Intell. Med.* 66 (2016) 29–39.
- [3] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer. Neural architectures for named entity recognition, 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference, 2016, pp. 260–270.
- [4] J. Devlin, M.-W. Chang, K. Lee, K.T. Google, A.I. Language, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: North American Association for Computational Linguistics (NAACL), 2019, pp. 4171–4186.
- [5] A. N'ev'eol, H. Dalianis, S. Velupillai, G. Savova, P. Zweigenbaum, Clinical Natural Language Processing in languages other than English: opportunities and challenges, *J. Biomed. Semant.* 9 (2018) 12.
- [6] S. Cornegruta, R. Bakewell, S. Withey, G. Montana, Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks, *Proceedings of the Seventh International Workshop on Health Text Mining and Information Analysis (LOUHI)*, 2016, p. 17–27.
- [7] S. Miao, T. Xu, Y. Wu, H. Xie, J. Wang, S. Jing, et al., Extraction of BI-RADS findings from breast ultrasound reports in Chinese using deep learning approaches, *Int. J. Med. Inf.* 119 (2018) 17–21.
- [8] M.E. Peters, M. Neumann, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: North American Association for Computational Linguistics (NAACL), 2018.
- [9] Y. Si, J. Wang, H. Xu, K. Roberts, Enhancing Clinical Concept Extraction with Contextual Embeddings, *J. Am. Med. Inform. Assoc.* 26 (2019) 1297–1304.
- [10] A. Savkov, J. Carroll, R. Koeling, J. Cassell, Annotating patient clinical records with syntactic chunks and named entities: the Harvey Corpus, *Lang Resour. Evaluat.* 50 (2016) 523–548.
- [11] S. Mayhew, D. Roth, TALEN: Tool for Annotation of Low-resource Entities, *ACL System Demonstrations*, 2018.
- [12] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, Neural Architectures for Named Entity Recognition, *Proc. NAACL-HLT 2016* (2016) 260–270.
- [13] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, A Next-generation Hyperparameter Optimization Framework, *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.
- [14] I. Jauregi Unanue, E. Zare Borzeshi, M. Piccardi, Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition, *J. Biomed. Inform.* 76 (2017) 102–109.
- [15] E. Alsentzer, J.R. Murphy, W. Boag, W.-H. Weng, D. Jin, T. Naumann, et al., Publicly Available Clinical BERT Embeddings, *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 2019, pp. 72–78.
- [16] A.E.W. Johnson, T.J. Pollard, L. Shen, L.-W.H. Lehman, M. Feng, M. Ghassemi, et al., MIMIC-III, a freely accessible critical care database, *Sci. Data* 3 (2016). K. Sugimoto et al.

- [17] Zizhao Zhang, Pingjun Chen, Manish Sapkota, and Lin Yang. 2017a. TandemNet: Distilling knowledge from medical images using diagnostic reports as optional semantic references. In Medical Image Computing and Computer-Assisted Intervention MICCAI 2017, pages 320–328, Quebec City, Quebec, Canada.
- [18] Singh, “Pushing the Limits of Radiology with Joint Modeling of Visual and Textual Information”, Proceedings of ACL 2018.
- [19] Chris L. Siström, “Free Text Versus Structured Format: Information Transfer Efficiency of Radiology Reports”, AJR:185, 2005.
- [20] Kenneth Jung, “Functional evaluation of out-of-the-box text-mining tools for data-mining tasks”, J Am Med Inform Assoc, 2015.
- [21] Anne-Dominique Pham, “Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings”, BMC Bioinformatics 2014.
- [22] Wencheng Sun, “Data Processing and Text Mining Technologies on Electronic Medical Records: A Review” Hindawi, Journal of Healthcare Engineering, 2018.
- [23] Matt O'Connor, “Language processing translates free-text radiology reports for better lung cancer staging”, healthimaging, 2021.
- [24] A.V.Krishna Prasad, “Extraction of Radiology Reports using Text mining” IJCSE, 2010.
- [25] Kento Sugimoto, “Extracting clinical terms from radiology reports with deep learning”, Journal of Biomedical Informatics, Elsevier, 2021.
- [26] Matthew C. Chen, “Deep Learning to Classify Radiology Free-Text Reports”, radiology.rsna.org, 2018.
- [27] Mitali Srivastava, “Analysis of Data Extraction and Data Cleaning in Web Usage Mining”, ResearchGate, 2015.
- [28] Wesam S. Bhaya, “Review of Data Preprocessing Techniques in Data Mining”, Journal of Engineering and Applied Sciences, 2017.
- [29] Imon Banerjee, “Intelligent Word Embeddings of Free-Text Radiology Reports” ResearchGate, 2017.
- [30] Arlene Casey, “A systematic review of natural language processing applied to radiology reports” BMC Med Inform, 2021.

## **AUTHORS**



Ms. Vaishali M. Kumbhakarna Master of Science from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India with First class in year 2013. And M.C.M. from Pune University in 2001. She has also completed M.Phil in Computer Science from Dr. BAM University, Aurangabad in the year 2016 and Currently she is pursuing her Ph.D. in Computer Science from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India. Currently working as Assistant Professor in Department of Computer Science and IT, M.G.M University, Aurangabad. since 2008. She has 14 years of teaching experience. She has published 8 research papers in reputed international journals including Scopus, Elsevier, Springer. Her main research work focuses on, biomedical digital image processing and Natural Language Processing



Dr. Sonali B. Kulkarni Master of Science from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India with First in the order of merit in year 2002. She has also completed Ph.D in Computer Science from Dr. BAM University, Aurangabad and currently working as Assistant Professor in Department of Computer Science and IT, Dr. BAM University, Aurangabad since 2008. She has 18 years of teaching and Research Experience. She is a life member of IETE, ISCA, Grant and IJENG She has published more than 30 research papers in reputed international journals including Scopus, Thomson Reuters, ICI, Web of Science and conferences including IEEE, Elsevier, Springer. Her main research work focuses on Remote Sensing & GIS, Brain Computer Interface, Natural Language Processing and Linguistics.



Ms. Apurva D. Dhawale completed M.Phil in Computer Science in 2015 from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India. Currently she is pursuing her Ph.D. in Computer Science from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India. She has 9 years of teaching experience, and published 9 papers in international & national reputed Journals/Conferences till date. Her research interest areas are Natural Language Processing & Biometric Image Processing.