

Machine Learning Algorithm Employed for Automated Clinical Audit Destined to Regulate the Inevitable PC-PNDT Act

Ms. Vaishali M. Kumbhakarna
Dept of CS and IT, Dr BAMU
Ch. Sambhajinagar, India

Dr. Sonali B. Kulkarni
Dept of CS and IT, Dr BAMU
Ch. Sambhajinagar, India

Ms. Apurva D. Dhawale
Dept of CS and IT, Dr BAMU
Ch. Sambhajinagar, India

Abstract—Massive volume of valuable clinical information is still hidden in a clinical narrative format. Consequently, the development of clinical NLP techniques is required to unlock the full potential of clinical data. In this study we developed the machine learning algorithm, which converts an unstructured clinical narrative automatically into structured clinical data. These clinical NLP-based Machine Learning applications retains transparency in clinical practice. It also helps for smooth conduction of clinical audit destined to regulate the law of PC-PNDT act of Parliament of India. It effectively reduce the legislative pressure of medical professional's. This paper discusses the machine learning methodologies used to process radiology text reports generated from sonography machine, which is in pdf format. It also improve the understanding of the patient's clinical records and the prediction of chronic diseases, risk and extract clinical information. This machine learning algorithm lays down basic safety standards for unstructured text of obstetric radiology reports and establish quality assurance through automated clinical audit.

Keywords—PC-PNDT, foeticide, infanticide, ML, USG, EDA, EHR, Pandas Profiling.

I. INTRODUCTION (HEADING 1)

Recognizing the key role of the medical community, the Government of India had made clinical audit mandatory for all radiology clinics in India. The objectives of this clinical audit is to improve the quality of patient care, promote the effective use of resources, and smooth functioning of law. Modern medical diagnostics have made sex determination and selection much easier. Even though law had passed long back in 1994, there are many perspectives and challenges arise while implementation of the PC-PNDT act in India. The Pre-Conception and Pre-Natal Diagnostic Techniques (Prohibition of Sex Selection) (PC-PNDT) Act is a legal tool aimed to regulate and prevent the misuse of diagnostic techniques to detect the sex of the foetus. Socio-cultural and religious biases of community customs such as the practice of dowry are perceived as a financial burden on the bride's family during and after marriage. It has direct influence on unwelcomes of female foetus. As part of an effort to understand the reasons

for such lax implementation of a crucial social legislation, the National Human Rights Commission (NHRC) and the United Nations Population Fund (UNFPA) jointly requested the Public Health Foundation of India (PHFI) to undertake a study to assess the status of implementation of the PC-PNDT Act. Across 18 states in the country where skewed sex ratios are a major problem. The birth rate of female babies were continuously declining. Abortion of a girl child in the mother's womb refers to female foeticide and killing a girl child after her birth is referred to as female infanticide. However, the persistent imperative is that the law must work and become a major tool in the hands of those who want to bring change. If the current reality of an available but grossly under-used legal framework can be altered. The reason of falling sex ratios is misuse of prenatal diagnostic technologies. The influence of widespread availability and use of sex selection and determination techniques on sex ratios is well-documented and acknowledged. Accordingly, the Indian government's response to this trend in the mid-nineties was to introduce Act to control sex selective abortion. Under the Act, all medical professionals performing pre-natal diagnostic techniques have been vested with the responsibility of ensuring adherence to the Act. The Ministry of Health and Family Welfare (MoHFW) works with a variety of partners, including different stakeholders. According to the law, clinics sealed for incomplete filing of Form F. Form F is mandatory document for radiology diagnosis. Deficiency or inaccuracy in filling Form F prescribed under Rule 9 made under the PNDT Act, is offence amounting to contravention of the provisions of sections 5 and 6. The Court would be justified, while imposing punishment on not maintaining the record of the conducting ultrasonography on a pregnant woman. Filling up of incorrect particulars may be taken in all seriousness in analysis of clinic audit, as if the provisions of sections 5 or 6 were violated. However filling up of the full name and address of the pregnant woman may be treated leniently if her identity and address were mentioned in a manner sufficient to identify and trace her.

Keeping the records for audit and maintaining the clinical data for 3 years ,facing the auditor is time consuming and sort of headache for doctors . The automated machine learning tools provide helping hand to doctor so that they can use their quality time more for patient care and clinical services. The computer professionals has a significant role in handling clinical data which is now in digital form .in this study we developed the machine learning algorithm .This algorithm extracts clinical text information from radiology reports. The structured data is used for clinical audit .It involves evaluation of data, documents and resources to check performance against standards .The purpose of a multidisciplinary clinical audit can be generally summarized as:

To regulate PC-PNDT Act

To control the female infanticide and foeticide

To improve the quality of patient care;

To promote the effective use of resources;

To enhance the provision and organization of clinical services;

To further professional education and training.

This paper focuses on clinical NLP system developed for evaluation of word, sentence, and document level annotations .The anticipated model specify attributes and features, extracted form obstetric radiology report .the document contains , named entities , diagnoses semantic attributes, negation, severity etc.

II. LITERATURE VIEW

In the paper entitled, “Audit of radiology request forms – “Are they adequately filled?” author cited that, the radiologist provide information and by referring it as an input, clinician[1] get clinical data. If the data of the request forms is not filled properly then the the analysis of the audits will be improper.

In the paper entitled“ A systematic review of natural language processing applied to radiology reports”,author cited that Automated tool[2] of clinical narratives from the radiology reports has the prospective to improve the healthcare practice. It lessens variety in reporting and provides uniformity in format. This study provided a systematic synthesis of clinical data, identify gaps, opportunities for collaboration and avoid repetition.

In this paper entitled “A Comprehensive Survey of Deep Learning in the field of Medical Imaging and Medical Natural Language Processing: Challenges and research directions”, author studed The literature included in the present work is collected from the Scopus database from 2017 to 2020. A total of 211 published manuscripts and cited that Deep learning has being popular in the last 5 years in medical NLP[3]. This study outlined how deep learning [4] deals with laborious manual feature engineering task, segmentation challenges, ambiguity in medical terms, the small volume of data, blurred boundaries of segments, the expensive computational overhead of processing pipeline and task, low-resolution images, reconstruction overhead of images and annotation issues of medical text.challenges in ambiguity in medical terms, ambiguity in segment boundaries in images, labelling issues in both medical images and medical text, and requirement of an expert in both sub domains.

In the paper entitled “The Need for Clinical Audits in Diagnostic Radiology” ,author cited that Clinical audit is a systematic inspection of medical radiological procedures. It is a tool strive for improving the quality and the outcome of patient care through structured review. In the clinical audit ,radiological[5][6],procedures, and results are scrutinized against standards for good medical radiological practices. It is carried out by auditors with extensive knowledge and experience of the radiological practices. It includes chapters on: quality management procedures and infrastructure; patientrelated procedures; technical procedures, and education, training and research.The objectives of a clinical audit are to improve the quality of patient care, promote the effective use of resources, enhance the provision and organisation of clinical services.the objectives should be agreed between the auditing organisation and the healthcare unit to be audited. The objectives should be based on the audit programmes by national coordinating organisations

In the paper entitled “Healthcare Techniques Through Deep Learning: Issues, Challenges and Opportunities ” ,author cited that The supervised and unsupervised learning techniques has major importance, and medical devices should be built over time based on these models, such as the microscope, phonendoscope, and electrocardiogram. Many researchers used deep learning with EHR data as it offers a firm prognostic results .It helps to compare different diseases based on different factors and parameters. Different DL[7] applications and algorithms, tools and data sets are also compared and laid down research opportunities and challenges in healthcare and is extensive, and still, many challenges are laid forward.

In the paper entitled, “Machine Learning Techniques for Biomedical Natural Language Processing:

A Comprehensive Review” ,author discusses existing challenges currently faced and open issues associated with the processing of the biomedical and clinical text and providing the NLP domain with sufficient resources and opportunities to extract new methodologies An important area of research related to the understanding of the challenges involved in processing the clinical text is the development of methodologies for processing the diverse format of clinical texts. Each format, on its own, is a challenge for NLP[8] researchers and can be explored using traditional and hybrid methodologies. biomedical NLP methods need to be modified and updated beyond the extraction of clinical terms to concentrate more on the interpretation of concepts.

In the paper entitled “Key Technology Considerations in Developing and Deploying Machine Learning Models in Clinical Radiology Practice ” ,author discuss artificial intelligence (AI) and machine learning models [9] have shown remarkable performance in the automated evaluation of medical images and text. insufficient training data, decentralized data sets, high cost of annotations, ambiguous ground truth, imbalance in class representation, asymmetric misclassification costs, relevant performance metrics, generalization of models to unseen data sets, model decay, adversarial attacks, explainability, fairness and bias, and clinical validation form of a laundry list. Author also have hopefulness to make it more accessible to researchers, software developers, radiologists, and other stakeholders .

In the paper entitled ,“Radiological safety status and quality assurance audit of medical X-ray diagnostic installations in India ” ,author discusses importance of periodic radiological safety and QA audits to ensure optimal radiological protection for patients, occupational workers in diagnostic radiology, and the public. This study makes significant contributions for improving the radiological safety status of medical X-ray installations in India [10]. It also pretence provision of regulatory documents pertaining to medical diagnostic X-ray practice.

In the paper entitled “Clinical audit in nuclear medicine ” ,author cited that Clinical Audit is the “Quality Management Audits In Nuclear Medicine Practices .The Clinical Audit [11] process which should be written protocols traceable to National or International standards for the given situation .It is a powerful tool to improve patient care and also connected with the ethical aspects of using Ionizing radiation guidelines by the National Society.

In the paper entitled “The current status of radiological clinical audit - an ESR Survey of European National Radiology Societies” ,author cited that ,Effective clinical audit is a key priority for all radiology departments across Europe.It has potential to improve patient care and outcomes. clinical audit “in accordance with national procedures” is mandatory and a legal requirement within the European Union. Clinical audit[12] is a valuable tool in modern healthcare systems, it is of particular relevance to radiologists due to its incorporation into the Basic Safety Standards Directive.

In the paper entitled “Deep Learning-Based Natural Language Processing in Radiology: The Impact of Report Complexity, Disease Prevalence, Dataset Size, and Algorithm Type on Model Performance” ,author cited that Automated text processing is the area of natural language processing and has an growing role in healthcare . Clinical NLP[13] has been applied in various applications in radiology to annotate texts or extract information .Natural language processing has evolved from hand crafted rule-based algorithms to machine learning-based approaches.

III. CLINICAL NLP METHOD

In this study ,the Machine Learning tool which we used for data cleaning and analysis is Pandas. Pandas word was derived from the term “Panel-data-s” an econometrics term for data sets include observations over multiple time periods for the same individuals. It has features which we are used for exploring, cleaning, transforming and visualizing data. The Python library called Pandas Profiling are used for Automated Exploratory Data Analysis (EDA) . It is a great tool to create reports in the interactive HTML format which is quite easy to understand and analyze the data in a very short time and with just a single line code using Python .

Exploratory Data Analysis is an approach for exploring and analysing datasets to generate insights in visual form. EDA is used to understand the main features of the dataset.It helps to identify missing values, count, mean, median, quintiles, distribution of data, correlation of variables with each other, data type, the shape of data, etc. Pandas deals with three types of data structures :Series,DataFrame and Panel.In this study we used dataframes.



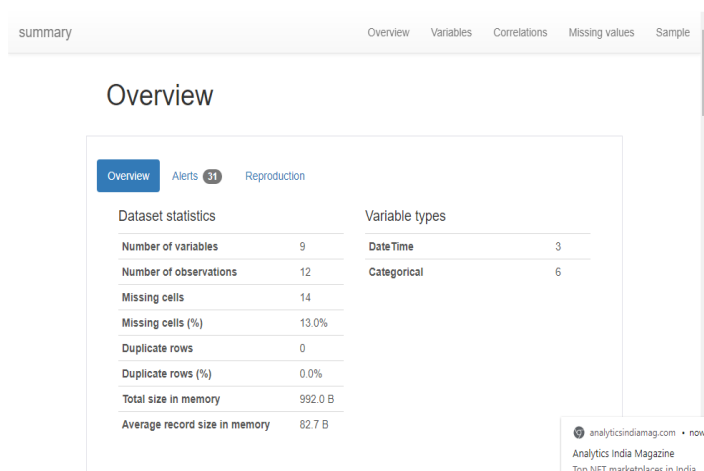
Fig 1: Clinical NLP Dataprocessing

IV. RESULT

Pandas Profiling Report Consists of the Following Sections:

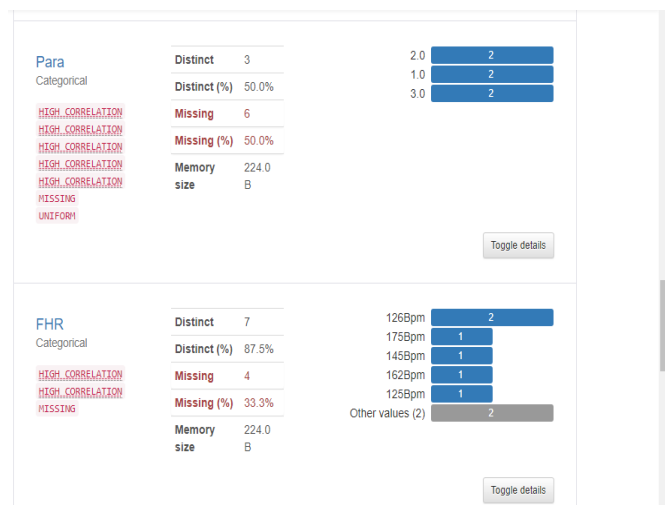
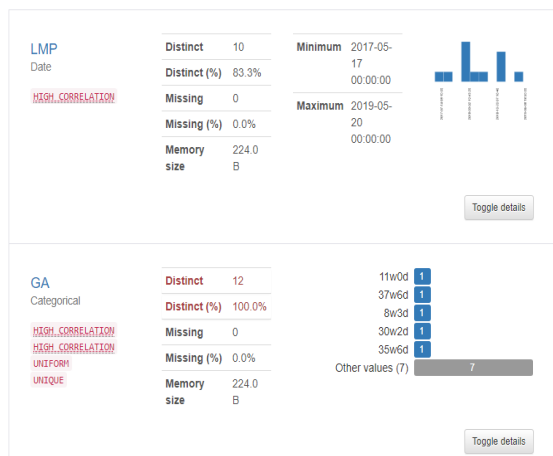
- Overview
- Variables
- Interactions
- Co-relations
- Missing Values
- Sample

a) Overview Section : This section provides overall data set information. Dataset statistics and Variable types.Dataset statistics display columns, rows, missing values, etc. Variable Types shows data types of the attributes of the data set. It also shows “Warnings”, where it shows which feature(s) are highly correlated to others



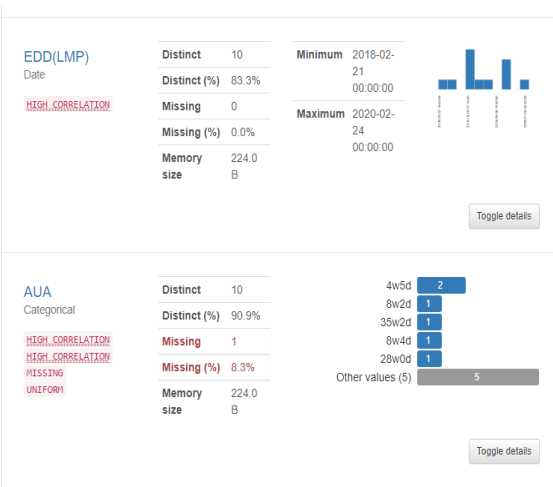
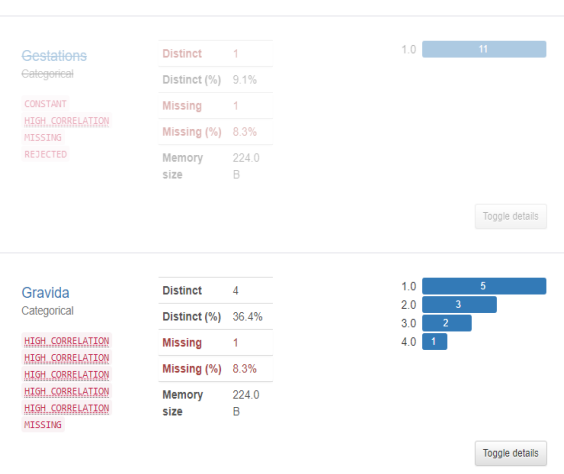
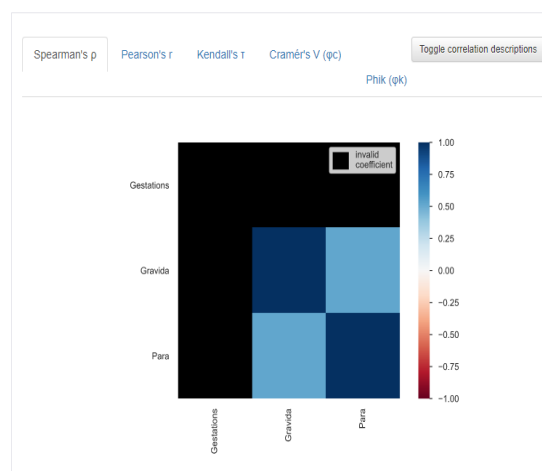
b) Variables : This section provides information about every feature individually in detail.

Variables



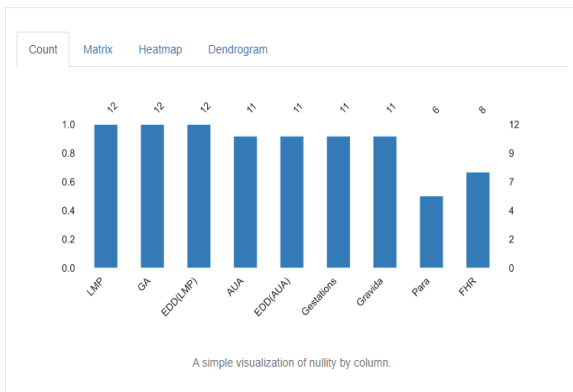
c) Correlation Section : This Section shows how features are co-related with each other

Correlations



d) Missing Value Section: We can see the Missing values in the Graph shown above. The report generated also contains the visualizations for the missing values present in the dataset in 4 types of plot: Count, matrix, Heatmap and dendrogram. The count plot is a basic bar plot with an x-axis as column names and the length of the bar represents the number of values present (without null values).

Missing values



Sample Section: This section displays the first and last 10 rows of the dataset.

Sample

First rows

	LMP	GA	EDD(LMP)	AUA	EDD(AUA)	Gestations	Gravida	Para	FHR
0	2018-01-10	11w0d	2018-10-17	8w2d	2018-11-05	1.0	3.0	2.0	175Bpm
1	2017-05-17	37w6d	2018-02-21	35w2d	2018-03-11	1.0	1.0	NaN	145Bpm
2	2018-11-25	8w3d	2019-09-01	8w4d	2019-08-31	1.0	2.0	1.0	162Bpm
3	2018-01-24	30w2d	2018-10-31	28w0d	2018-11-16	1.0	1.0	NaN	126Bpm
4	2018-01-24	35w6d	2018-10-31	31w4d	2018-11-30	1.0	1.0	NaN	125Bpm
5	2018-01-24	40w3d	2018-10-31	33w5d	2018-12-17	1.0	1.0	NaN	135Bpm
6	2018-11-10	6w6d	2019-08-17	5w4d	2019-08-26	1.0	4.0	3.0	NaN
7	2018-03-07	7w3d	2018-12-12	4w5d	2018-12-31	1.0	3.0	2.0	NaN
8	2018-05-22	30w6d	2019-02-26	30w6d	2019-02-26	1.0	2.0	3.0	126Bpm
9	2017-08-20	29w3d	2018-05-27	29w6d	2018-05-24	1.0	2.0	1.0	163Bpm

Last rows

	LMP	GA	EDD(LMP)	AUA	EDD(AUA)	Gestations	Gravida	Para	FHR
2	2018-11-25	8w3d	2019-09-01	8w4d	2019-08-31	1.0	2.0	1.0	162Bpm
3	2018-01-24	30w2d	2018-10-31	28w0d	2018-11-16	1.0	1.0	NaN	126Bpm
4	2018-01-24	35w6d	2018-10-31	31w4d	2018-11-30	1.0	1.0	NaN	125Bpm
5	2018-01-24	40w3d	2018-10-31	33w5d	2018-12-17	1.0	1.0	NaN	135Bpm
6	2018-11-10	6w6d	2019-08-17	5w4d	2019-08-26	1.0	4.0	3.0	NaN
7	2018-03-07	7w3d	2018-12-12	4w5d	2018-12-31	1.0	3.0	2.0	NaN
8	2018-05-22	30w6d	2019-02-26	30w6d	2019-02-26	1.0	2.0	3.0	126Bpm
9	2017-08-20	29w3d	2018-05-27	29w6d	2018-05-24	1.0	2.0	1.0	163Bpm
10	2019-05-20	7w0d	2020-02-24	4w5d	2020-03-11	1.0	1.0	NaN	NaN
11	2018-11-19	0w5d	2019-08-26	NaN	NaN	NaN	NaN	NaN	NaN

V. CONCLUSION

The volume of clinical data is growing faster than the number of radiologists. The high workload on medical professionals leads to errors in diagnosis because of human fatigue, unacceptable delays in reporting, and stress and burnout in radiologists. NLP provides a powerful methods for unlocking information from unstructured clinical narratives. Despite of developing new standards and better encoding EHR with clinical terminology standards, there is still a narrative aspect, which makes the biomedical/clinical NLP methods essential for clinical research informatics. In this study we have used random sampling of 50 records for EDA out of 727 tuples dataset. We can conclude from this study that Python is the powerful tool for Automated Exploratory Data Analysis with just one click and single line code. There is high correlation between all variables, and missing values of form F easily extracted by using pandas profiling tool.

REFERENCES

- [1] Rajanikanth Rao, "Audit of radiology request forms –“Are they adequately filled?””, Journal of Medical and Scientific Research, 2014.
- [2] Arlene Casey, Emma Davidson, "A systematic review of natural language processing applied to radiology reports", Med Inform Decis Mak, BMC, 2021.
- [3] Sumithra Velupillaia, Hanna Suominen, "Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances, Journal of Biomedical Informatics, Elsevier, 2018.
- [4] Babita Pandey, Devendra Kumar Pandey, "A Comprehensive Survey of Deep Learning in the field of Medical Imaging and Medical Natural Language Processing: Challenges and research directions", Elsevier, 2021.
- [5] Bashiru Babatunde Jimah, "Audit of Radiology Request Form for Completion and Usefulness of Clinical History: Teaching Hospital Experience, Ghana", Hindawi Radiology Research and Practice, 2021.
- [6] Jan Schillebeeckx, "The Need for Clinical Audits in Diagnostic Radiology", HealthManagement.org SSN = 1377-7629, 2017.
- [7] DUR-E-MAKNOON NISAR, RASHID AMIN, "Healthcare Techniques Through Deep Learning: Issues, Challenges and Opportunities", IEEE 2021.
- [8] ESSAM H. HOUSSEIN, REHAB E. MOHAMED, "Machine Learning Techniques for Biomedical Natural Language Processing: A Comprehensive Review", IEEE 2021.
- [9] Viraj Kulkarni, Manish Gawali, "Key Technology Considerations in Developing and Deploying Machine Learning Models in Clinical Radiology Practice JMIR, 2021.
- [10] A. U. Sonawane, Meghraj Singh, "Radiological safety status and quality assurance audit of medical X-ray diagnostic installations in India", jmp, 2010.
- [11] Siroos Mirzaei & Lorenzo Maffioli, "Clinical audit in nuclear medicine", Springer, 2010.

- [14] European Society of Radiology ,“The current status of radiological clinical audit - an ESR Survey of European National Radiology Societies”, Springer, 2019.
- [15] A. W. Olthof, P. M. A. van Ooijen1, “Deep Learning-Based Natural Language Processing in Radiology: The Impact of Report Complexity, Disease Prevalence, Dataset Size,and Algorithm Type on Model Performance”, Springer, 2021.

