Chest X-Ray Report Generation Using Machine Learning Techniques

Prof. Pawan Kumar Pal, Anmol Ratan, Sujal Gupta, Vikas Yadav
Department of Computer Science
KIET Group of Institutions, Delhi-NCR

Abstract—Medical Imaging is indispensable for clinical analysis and the visualization of organ functions within the human body, especially chest. Techniques like X-ray play a pivotal role in diagnosing and treating various conditions. The unprecedented demand for x ray reports during the COVID-19 pandemic highlighted the critical need for swift and accurate diagnostics, particularly concerning chest-related ailments. Even in the post pandemic era, increased awareness has led to a surge in medical imaging demands, underscoring the necessity for efficient, accurate and automatic process of generating reports. The traditional method of generating radiological reports relies heavily on the expertise of radiologists. While experienced professionals can provide accurate interpretations, the process is time-intensive, requiring an average of 10 minutes or more per report.

This introduces an intuitive mobile application that streamlines the traditional chest X-ray report generation process for users by incorporating a machine learning (ML) model in an android app. The app seamlessly uses machine learning (ML) model in the form of encoder decoder mechanism, to provide comprehensible reports in layman's language in the android mobile application. The application enables users to capture chest X-ray images using their smartphone cameras, facilitating personalized reports based on the provided x ray image by the user.

The user-friendly process begins with the user capturing the image of the chest X-ray through the app's camera interface or directly selecting the image file from the file, the image is then processed through the ML model, utilizing the encoder decoder mechanism thus generating reports with commonly used words and phrases that are easily understood by individuals having a non-medical background. The generated report not only highlights the presence or absence of chest-related abnormalities but also conveys the information in a clear and accessible manner. By providing actionable insights in layman's language, users can promptly understand the results and make informed decisions about their health. In response to the heightened awareness of chest diseases post-COVID-19, this app serves as a crucial tool for individuals to actively know their chest health. The feature of capturing the image through your own mobile camera of your chest x ray simplifies the process, making chest X-ray analysis accessible to users without specialized medical knowledge.

I. INTRODUCTION

In the medical field there is a term called Medical Imaging, it is the process of creating visual representations of the interior of the body for clinical analysis as well as visual representation of the function of organs or tissues. It is one of most significant and widely used methods for diagnosis and treatment. Some of the significant examples of medical imaging are X- Ray, MRI (Magnetic Resonance Imaging) etc. During the pandemic times (Covid-19) we saw that there was surge in the demands of medical imaging and Nowadays after the post pandemic (Covid19) we have seen that people have become more conscious towards the diseases, especially chest related diseases. Detailed information generated from medical images is necessary for diagnosing illnesses or tracking patients' progress. However, every image requires a radiologist to carefully examine and write a full-text report to describe the findings. Diagnosing medical images requires an appropriate amount of experience from the radiologists to develop more confident and accurate reports. Furthermore, a more glaring issue is the amount of time it takes the radiologist to write a full-text report. It would take on average 10 min or more based on the radiologist's degree of experience, so this would prove very time consuming when considering the number of cases, a radiologist should investigate per day, and in crowded hospitals, regions, and cities, this would be problematic. Also, for less experienced radiologists and pathologists, especially those working in the rural area where the quality of healthcare is relatively low, writing medical-imaging reports is demanding.

II. DATA COLLECTION AND PREPROCESSING

A. Data Sources

The foundation of any machine learning-driven research in the healthcare domain relies on the quality and appropriateness of the data sources. In this study, we accessed a diverse range of data sources to build a comprehensive dataset for chest x ray report generation. The image dataset extracted from the Indiana medical dataset consists of a significant number of images specifically depicting early pneumonia, fibrosis and other thorax-based diseases affecting chest post covid -19 specifically. This curated dataset serves as the primary training and validation data for the deep learning model. This dataset felicitates to train the model on the chest x- ray images. The dataset comprises of two parts one is more chest x-ray image and other is for reports. In the dataset is possible that more than one image can be there for a single report. So, for that we will process the dataset to contain uniformity.

B. Data Preprocessing

Proper data preprocessing is essential for preparing the collected data for analysis. In this phase, raw data underwent several preprocessing steps to enhance its quality and com- partibility for machine learning. These steps included data cleaning, where missing chest x-ray images were handled through imputation techniques, and outliers were identified and treated appropriately. Feature engineering was performed to extract relevant information and create new informative variables. Data

normalization, augmentation, segmentation was applied to ensure that all features were on a consistent scale, preventing any undue influence of magnitude differences on machine learning algorithms.
Finally, the dataset was split into training, validation, and test sets to enable model development, tuning, and evaluation.

The rigorous data collection and preprocessing steps were pivotal in establishing a robust foundation for our predictive models, ensuring that they were trained on high-quality data and poised for accurate chest x-ray report generation.

III. FEATURE SELECTION AND ENGINEERING

A. Image Feature Extraction and Selection:

In our chest X-ray report generation project, leveraging image extraction features is paramount. We employ a pretrained CheXNet model fine-tuned specifically for extracting features from chest X-ray images, focusing on common chest diseases. These features are represented as global 300dimensional vectors, containing commonly used words. By utilizing these vectors, we ensure that the generated reports are easily understandable in layman's terms. We further enhance our model's efficiency by employing various feature selection techniques. These techniques aim to identify the most informative variables within our extensive dataset, optimizing predictive accuracy and interpretability while mitigating the risk of overfitting.

B. Domain-Specific Features

In the realm of chest X-ray report generation, disease specific features play a pivotal role in refining our dataset. Beyond the extracted image features, we incorporate additional clinical indicators relevant to chest diseases. These include variables related to specific pathologies, anatomical abnormalities, and radiological findings. By integrating disease-specific features, our models can capture the nuances of various chest conditions, enhancing both accuracy and clinical relevance in the generated reports.

C. Data Preprocessing and Standardization:

Ensuring consistency in feature scales is imperative to avoid bias in our machine learning algorithms. In our project, we prioritize data preprocessing and standardization techniques to achieve this goal. Alongside image feature extraction, we apply methods such as Min-Max scaling and z-score normalization to bring all variables to a common range. This step is crucial for integrating features from diverse sources,
such as clinical measurements and image-derived vectors.

IV. METHODOLOGY

A. Machine Learning Algorithms

In the development of predictive models for Polycystic Ovary Syndrome (PCOS), the selection of appropriate ma- chine learning algorithms is a critical decision. To address the complex and multifaceted nature of PCOS, a range of machine learning algorithms was considered. These included traditional techniques like logistic regression, support vector machines, and k-nearest neighbours, which are known for their interpretability. Additionally, we explored the capabilities of more advanced models, such as random forests, gradient boosting, and deep neural networks. Each algorithm was evaluated in terms of its ability to handle the dataset’s size and dimensionality, as well as its capacity to capture non-linear relationships and interactions within the data. The selection process was guided by a balance between predictive performance, interpretability, and the clinical relevance of the resulting models.

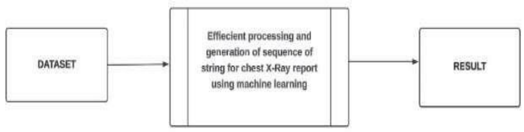

Fig. 1: A sample figure.

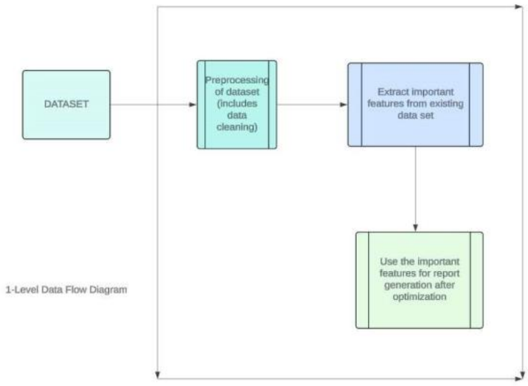

Fig. 2: A Sample Figure.

B. Model Development

The development of chest x-ray report generation models involved a systematic process, starting with data preprocessing, feature selection, and feature engineering. Machine learning algorithms were then trained on the refined dataset, leveraging the selected features. Hyperparameter tuning was performed to optimize the algorithms’ performance, ensuring that models achieved their maximum predictive capabilities. The process encompassed iterative steps, such as cross-validation to prevent overfitting and assembling techniques to enhance predictive accuracy. Model development also integrated domain-specific knowledge, aligning the models with medical imaging criteria for report generation. The resulting model were designed to provide a clear understanding of the relationships between input features, facilitating their interpretation by healthcare professionals.

C. Model Evaluation

The evaluation of models was carried out meticulously to assess their efficacy and reliability. Performance metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) were employed to quantify the models' predictive power. We conducted cross-validation to validate model gen- realizability and mitigate overfitting concerns. Additionally, the clinical implications of model predictions were examined.

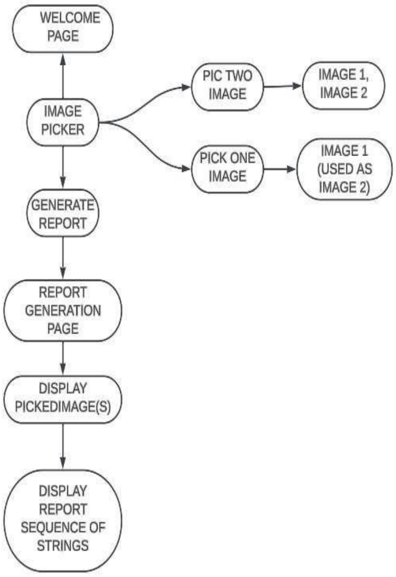

Fig 3: Sample figure.

emphasizing the potential impact and patient care. Comparative analyses were conducted to benchmark the models against each other and against existing clinical methods. Model evaluation aimed not only to measure predictive accuracy but also to demonstrate the clinical utility and interpretability of the models in the context of chest x-ray report generation.

The methodology adopted in this study reflects a comprehensive and systematic approach to the development and evaluation of machine learning models for report generation. It underscores the fusion of computational techniques with clinical knowledge, ultimately contributing to improved healthcare outcomes.

V. EXPERIMENTAL SETUP

A. Data Splitting and Cross-Validation

The design of the experimental setup is crucial to ensure the robustness and reliability of the results. To this end, we employed a data splitting and cross-validation strategy. A division of the dataset into three subsets—comprising a training set, a validation set, and a test set—was performed. The training set, constituting most of the data, was used for model training. The validation set was employed for hyperparameter tuning and model selection to prevent overfitting. The test set, distinct from the validation set, was reserved for final model evaluation. Cross-validation, specifically k-fold cross validation, was applied to enhance the generalizability of the models. This approach involved repeatedly splitting the data into k subsets, using k-1 for training and one for validation, to assess model performance across multiple iterations and mitigate variability.

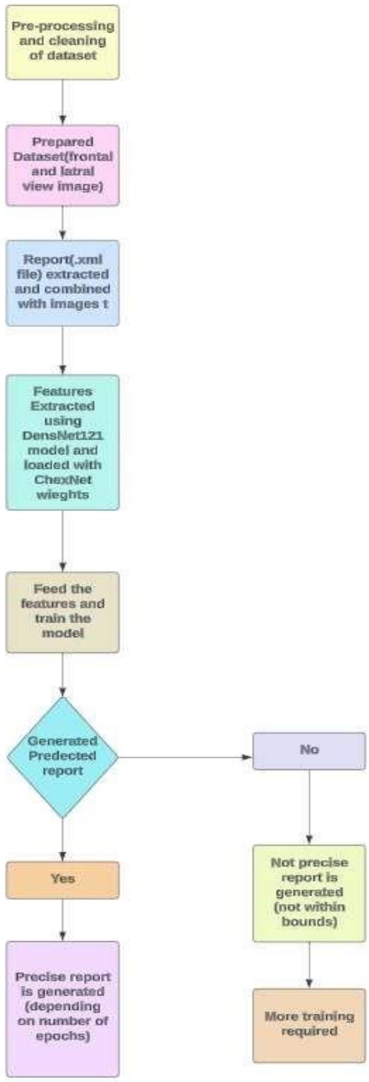

Fig 4: A sample figure

B. Hyperparameter Tuning

Hyperparameter tuning is a critical aspect of model deelopement to optimize the performance of machine learning algorithms. We systematically varied hyperparameters for each algorithm, such as learning rates, regularization terms, and tree depths, to identify the settings that yielded the best results. Grid search and random search techniques were employed to explore hyperparameter combinations efficiently. The hyperparameter tuning process was guided by performance on the validation set, ensuring that the models achieved their maximum predictive capabilities while avoiding overfitting.

C. Performance Metrics

The selection of appropriate performance metrics is es- essential to quantitatively evaluate the predictive power of the models. In this study, we considered a comprehensive set of metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC- ROC). Accuracy provides an overall measure of correct pre- dictions, while precision and recall assess the balance between true positives and false positives and the ability to cases accurately. F1 score considers the harmonic mean of precision and recall, offering a balanced metric. AUCROC, on the other hand, assesses the model's ability to discriminate between PCOS and non-PCOS cases across various probability thresholds. These metrics were systematically calculated and reported to provide a comprehensive evaluation of model performance.

The experimental setup outlined in this section was meticulously designed to ensure the reliability and generalizability of our findings. It encompassed data splitting, cross-validation, hyperparameter tuning, and the use of an array of performance metrics to rigorously evaluate the PCOS prediction models.

VI. RESULTS

A. Model Performance

The heart of our research lies in the evaluation of the report generation using chest x-ray images. In this section, we present a comprehensive analysis of the performance of these models. Performance metrics, including accuracy, precision, recall, F1score, and AUC-ROC, provide quantitative insights into the models' predictive capabilities. We discuss the trade-offs between these metrics, highlighting the models' strengths and limitations in PCOS prediction. Detailed results are presented, including confusion matrices and receiver operating characteristic (ROC) curves, to facilitate a nuanced understanding of the models' classification performance.

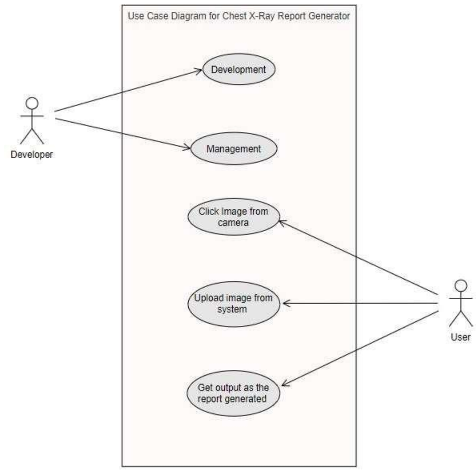

Fig.4: A sample figure.

B. Feature Importance Analysis

An integral aspect of our research involves understanding the key factors. Feature importance analysis elucidates the significance of individual features in driving model decisions. We employ techniques such as feature ranking and permutation importance to identify the most influential clinical, physiological, and lifestyle variables. The results of this analysis not only provide valuable insights into the critical factors contributing to report generation but also offer a foundation for the development of interpretable models. Feature importance analysis plays a vital role in bridging the gap between machine learning predictions and clinical interpretation.

C. Comparative Analysis of Algorithms

To determine the most effective machine learning approach, a comparative analysis of algorithms is conducted. We evaluate the performance of various algorithms, both traditional and advanced, in the context of report generation. This analysis highlights the strengths and weaknesses of each algorithm, considering factors such as accuracy, interpretability, and computational efficiency. The findings from the comparative analysis offer valuable guidance for selecting the most suitable machine learning techniques for report generation, diagnosis and early intervention.

The results section of this research paper provides a com- prehensive overview of the models' performance, feature im- portance, and a comparative analysis of machine learning algorithms. These insights are crucial in assessing the effectiveness of the predictive models and in guiding future research efforts in the field of AI medical imaging and management.

VII. CONCLUSION

The study presented in this research paper signifies a significant stride in the pursuit of effective extracting the images features and accurately generate the report through the application of machine learning techniques. The multifaceted nature of chest x-rays, with its diverse clinical, physiological, and lifestyle components, presents a formidable challenge. Nevertheless, our research demonstrates the potential of machine learning in addressing this complexity and improving the accuracy of report generation.

The results of our investigation highlight the promising performance of machine learning models in report generation. These models, after meticulous feature selection, engineering, and preprocessing, exhibit notable predictive power, as indicated by metrics such as accuracy, precision, and recall. The feature importance analysis not only underscores the critical role of specific attributes in our model but also enhances the interpretability of the models, aligning them with clinical criteria. Our comparative analysis of machine learning algorithms offers valuable insights into the strengths and weaknesses of different techniques, guiding future research directions.

Moreover, the clinical implications of our research extend beyond predictive accuracy. By enabling early intervention, these models have the potential to improve the quality of healthcare provided to individuals affected by chest-based diseases. The integration of domain-specific features ensures that our models capture the clinical nuances of chest x-ray, enhancing their relevance in a healthcare setting.

In conclusion, this research advances the field of medical images through chest x-ray report generation by leveraging machine learning, data-driven insights, and clinical knowledge. The models developed herein offer a valuable tool for healthcare professionals in the early identification of any irregularities in the chest. However, we acknowledge that there is room for further refinement, including the integration of additional data sources, larger sample sizes, and the continuous improvement of model interpretability. Our work sets the stage for ongoing research, with the goal of creating model which can handle and predict all cases accurately.

VIII. FUTURE WORK

While our research has made substantial progress in the application of machine learning for the chest x-ray report generation, there are several promising avenues for future research that can expand upon and enhance the contributions made in this study.

A. Incorporating Genetic Data:

Genetic factors play a significant role in chest x-ray report development. Future research can explore the integration of genetic data, specialised for doctor The fusion of clinical, physiological, lifestyle, and genetic data can lead to more comprehensive and precise prediction models.

B. Clinical Decision Support Systems:

The translation of predictive models into practical clinical tools, such as decision support systems, holds immense po- tertial. Future research can focus on the development of user- friendly applications that provide real-time predictions and recommendations to healthcare professionals, facilitating early diagnosis and personalized treatment plans.

C. Validation in Diverse Populations:

The generalizability of models across diverse populations is a critical consideration. Further research should aim to validate these models in different ethnic and geographical groups to ensure their robustness and applicability worldwide.

D. Interpretable Models:

The interpretability of machine learning models is essential for their acceptance in clinical practice. Future work can explore advanced techniques in model interpretability, such as LIME (Local Interpretable Model-agnostic model) E. Patient-Centric Approaches:

Personalized healthcare is a key focus for the future. Re- search efforts can investigate the incorporation of patient preferences, values, and lifestyle choices into chest x-ray report generation models to create patient-centric treatment recommendations.

F. Real-Time Monitoring:

The development of real-time monitoring systems that continuously assess chest risk a provide immediate feedback to patients can be a transformative direction. Such systems can empower individuals to make timely lifestyle changes and seek early medical intervention.

In conclusion, the future of chest x-ray report generation and management lies in the fusion of advanced data science techniques, clinical insights, and patient engagement. The ongoing exploration of these avenues promises to further advance the field and improve healthcare outcomes for individuals affected by chest-based diseases.

REFERENCES

1. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, Pranav Rajpurkar ,Jeremy Irvin ,Kaylie Zhu, BrandonYang , Hershel Mehta ,Tony Duan ,Daisy Ding ,Aarti Bagul ,Robyn L. Ball ,Curtis Langlotz ,Katie Shpanskaya ,Matthew P Lungren ,Andrew Y. Ng.
2. Attention-guided Chained Context Aggregation for Semantic Segmentation Quan Tang, Fagui Liu, Member, IEEE, Tong Zhang, Member, IEEE, Jun Jiang, and Yu Zhang
3. Deep Visual-Semantic Alignments for Generating Image Descriptions Andrej Karpathy Li Fei-Fei

Department of Computer Science, Stanford University

4. Image in an Image Caption Generator Marc Tanti Albert Gatt Institute of Linguistics and Language Technology University of Malta marc.tanti.06@um.edu.mt albert.gatt@um.edu.mt Kenneth P. Camilleri Department of Systems and Control Engineering
University of Malta kenneth.camilleri@um.edu.mt 12-Mar-2018

5. On the Automatic Generation of Medical Imaging Reports Baoyu Jing Pengtao Xie Eric P. Xing Petuum
Inc, USA School of Computer Science, Carnegie Mellon University, USA. PMID: 11701829.

6. Radiology reporting: a general practitioner's perspective
F M Grieve, MBChB, BSc, MRCS, A A Plumb, MBChB, MRCP, and S H Khan, MBBS, FRCR

7. Automated radiology report generation using conditioned transformers Omar Alfarghaly, Rana Khaled, Abeer Elkorany , Maha Helal , Aly Fahmy

8. Smoothing Convolutional Factorizes Inception V3 Labels and Transformers for Image Feature Extraction into Text Segmentation DOI:10.1109/ICSGTEIS60500.2023.10424317 Komang Ayu Triana Indah,Made Sudarma Udayana University,Rukmi Hartati Udayan University

9. Feature Extraction Model Based on Inception V3 to Distinguish Normal Heart Sound from Systolic Murmur Jinhee Bae; Minwoo Kim; Joon S. Lim

10. Deep Learning for Cardiac Image Segmentation: A Review Chen Qin Huaqi Qiu Giacomo Tarroni Jinming Duan Wenjia Bai Daniel Rueckert

11. Hossain, S.; Lee, D.-J. Autonomous-Driving Vehicle Learning Environments using Unity Real-time Engine and End-to-End CNN Approach. J. Korea Robot. Soc.
2019, 14, 122–130.

12. Koci´c, J.; Jovi˘ci´c, N.; Drndarevi´c, V. An End-to-End Deep Neural Network for Autonomous Driving Designedfor Embedded Automotive Platforms. Sensors 2019, 19, 2064.

13. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.

14. Corrado, G.; Thrun, S.;Dean, J. A guide to deep learning in healthcare. Nat. Med. 2019, 25, 24. [14] Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: